



**THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL**

Préparée à l'École Normale Supérieure

**Statistical modeling of the polarized emission of
interstellar dust**

Soutenue par

Bruno Régaldo-Saint Blancard

Le 2 novembre 2021

École doctorale n°127

**Astronomie et Astrophysique d'Île
de France**

Spécialité

Astrophysique

Composition du jury :

Benoît SEMELIN
Sorbonne Université

Président

Blakesley BURKHART
Rutgers University & Flatiron Institute

Rapportrice

Brice MÉNARD
Johns Hopkins University

Rapporteur

Susan CLARK
Stanford University

Examinateuse

Stéphane MALLAT
Collège de France

Examinateur

Jérôme PETY
Institut de Radioastronomie Millimétrique

Examinateur

François LEVRIER
École Normale Supérieure

Directeur de thèse

François BOULANGER
École Normale Supérieure

Invité, co-directeur

Remerciements

Une thèse est une longue aventure pleine de rebondissements, à la fois passionnante et éprouvante, stimulante et décourageante, et en fin de compte infiniment enrichissante. Je souhaite adresser mes profonds remerciements à celles et ceux qui m'ont accompagné tout au long de ce périple.

Mes premiers remerciements vont à mes directeurs de thèse, François Levrier et François Boulanger, qui m'ont proposé de travailler sur cette thématique fascinante à la croisée de la science du milieu interstellaire et de l'étude du fond diffus cosmologique. Ils m'ont tous les deux progressivement formé au métier de chercheur, et ce de façon toujours très complémentaire. En plus de m'avoir permis de visiter la Corse, la Bretagne, et même New York, je les remercie pour leurs aiguillages à chaque nouvelle étape de ce travail, leurs précieux conseils et nombreuses relectures pointilleuses en tout genre.

Un grand merci aussi à Erwan Allys, troisième membre essentiel du groupe AstroStat, et qui a aussi été pour moi un troisième directeur de thèse : toujours disponible pour échanger, conseiller et relire mon travail avec le plus grand soin.

Je tiens aussi à remercier toute l'équipe Astrophysique du LPENS pour leur accueil, et plus particulièrement les membres de MIST pour m'avoir invité à leurs multiples événements à l'OHP, Cargèse, Roscoff, Goutelas... Merci également à l'équipe BxB qui m'a également accueilli lors de plusieurs workshops à Toulouse ou Paris. Tous ces moments privilégiés, et toujours très chaleureux, m'ont largement permis d'enrichir ma compréhension de la physique du milieu interstellaire.

Je remercie également Stéphane Mallat, qui au delà d'être à l'origine de la démarche statistique adoptée pour ce travail de thèse, a toujours été d'un précieux conseil, ainsi que d'une aide cruciale pour ma poursuite en post-doctorat au Flatiron Institute. Merci également aux membres de l'équipe Data du département d'informatique de l'ENS, et tout particulièrement à Sixin Zhang, pour nos multiples échanges.

Merci aux nombreux collègues et collaborateurs avec qui j'ai eu le plaisir d'échanger et qui m'ont permis d'enrichir ce travail de thèse par divers aspects : Elena Bellomi, Niall Jeffrey, Pierre Lesaffre, Jean-François Cardoso, Jean-Marc Delouis, Jean-Baptiste Durrière, Jonathan Aumont, Léo Vacher, Tanguy Marchand. Merci à Sandrine Codis et Dick Bond pour m'avoir aussi donné l'occasion d'élargir mon champ de recherche en m'aidant à mener à bout un projet de longue date sur l'étude statistique des grandes structures de l'Univers.

Cette thèse a aussi été grandement enrichie par mes expériences diverses d'enseignement. Je remercie Laurette Tuckerman et Kris Van Houcke avec qui j'ai eu le plaisir d'enseigner la physique numérique pendant trois ans, et qui m'ont fait confiance pour mener mon enseignement en grande autonomie. Merci aussi à mes amis de l'équipe de « Physique pour tous » pour cette expérience mémorable d'enseignement : Tristan, Alexis, Thibaud, Clément, Brice.

Finalement, mille mercis à mes amis et ma famille qui m'ont toujours soutenu et encouragé tout au long de cette aventure. Merci à mes cobureaux Thibaud, Pierre, et Clément pour cette fraternité de doctorants et ces innombrables discussions à refaire le monde dans les bars et restaurants les plus raffinés du quartier. Merci aux amis du couloir, des bureaux voisins, ou des étages d'en dessous qui ont largement égayé mon temps passé à l'ENS autour de l'inmanquable Capu-16 ou d'une (et souvent plusieurs..) bière(s) à la Montagne : Alexis, Michaël, Elena, Alice, Adrian, Gaurav, Tristan, Alexandre, Clément, Marlone, Christophe, Stéphane, François. Merci bien sûr aux colocs Tristan (encore lui !) et Lucas, aux côtés de qui faire une thèse de physique semblait n'être qu'une banalité. Je n'oublie pas de remercier mes fameux compères du Torontrio : Tristan (promis c'est fini) et Brice, pour le chemin parcouru depuis Fido jusqu'à nos soutenances respectives. Une pensée toute particulière pour Marine, à qui je dois beaucoup, et qui n'a cessé de m'encourager pendant deux années de cette thèse. Finalement, je souhaite remercier mes parents et mes deux soeurs, pour leur soutien indéfectible et essentiel.

Contents

Remerciements	i
Introduction	vii
Publications related to this thesis	xii
Index of main abbreviations and notations	xiii
I From the observation of the cosmic microwave background to the physics of the interstellar medium	1
I.1 The first light of the Universe	2
I.1.1 An expanding Universe	2
I.1.2 The discovery of the cosmic microwave background	5
I.1.3 From inhomogeneities to anisotropies	6
I.2 Observing the CMB	8
I.2.1 The observables: Stokes parameters	8
I.2.2 The <i>Planck</i> mission	11
I.2.3 A component separation problem	12
a) Foregrounds	13
b) Component separation methods	14
I.3 The Milky Way, its interstellar medium, and the dust foreground	17
I.3.1 The Milky Way and the ISM	17
I.3.2 Interstellar dust grains	19
I.3.3 The dust foreground: sky maps and main statistical properties . .	20
I.3.4 Simulating the dust emission	24
a) Solving MHD equations	24
b) Choice of a physical model of dust	25
c) Resolution of the radiative transfer	26
II Modeling the dust emission: motivations and state-of-the-art	29
II.1 Motivations	29
II.1.1 Why a statistical approach?	30
II.1.2 What is a good model?	31
II.1.3 Standard statistics	32
II.2 Phenomenological models	35
II.2.1 Deterministic approaches	36

a)	Modified blackbody radiation-based models	36
b)	Physical models of dust	37
II.2.2	Statistical approaches	39
II.3	Agnostic models	41
II.3.1	Convolutional neural networks	42
II.3.2	Deep generative modeling and the dust foreground	44
II.4	The best of both worlds?	45
III Statistical characterization with the wavelet scattering transform		49
III.1	The wavelet scattering transform	50
III.1.1	Motivations	50
III.1.2	Definition	51
a)	The wavelet transform	51
b)	The WST coefficients	52
c)	Normalization and interpretation	54
III.1.3	Properties	55
III.2	Methodology illustrated on simulated Stokes maps	56
III.2.1	Building simulated Stokes maps	56
a)	Presentation of the simulation	57
b)	Stokes maps derivation	57
c)	Statistical isotropy/anisotropy of the maps	58
III.2.2	Presentation of the data sets	58
III.2.3	Regularity in the WST coefficients	61
III.2.4	Simplification of the WST description with the reduced WST	63
a)	The RWST model	63
b)	Fourier perspective	64
c)	Fit of the model	65
d)	Extension of the model	66
III.3	RWST data analysis and interpretation	68
III.3.1	Isotropic fluctuations in first order coefficients \hat{S}_1^{iso}	68
III.3.2	Anisotropic fluctuations in first order coefficients \hat{S}_1^{aniso}	70
III.3.3	Second order coefficients and non-Gaussianity of the data	71
III.3.4	Lattice terms	73
III.4	Conclusion and perspectives	73
IV Generative models from scattering-like descriptions		77
IV.1	Microcanonical models	78
IV.1.1	Maximum entropy microcanonical models	78
IV.1.2	Microcanonical gradient descent models	79
IV.1.3	Numerical considerations	79
a)	Notions of optimization	79
b)	Automatic differentiation	80
IV.2	Application with the RWST	81
IV.2.1	Generation of synthetic polarization maps	82
a)	Initial and reference maps	82
b)	Loss function	83

c)	Synthesis from the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set	83
IV.2.2	One-point and two-point statistics of synthetic maps	85
IV.3	Application with the WPH	88
IV.3.1	The wavelet phase harmonics	88
a)	Bump-steerable wavelets	89
b)	WPH moments	90
c)	Choice of a subset of WPH moments	92
d)	Scaling moments	94
e)	Normalized estimates	95
IV.3.2	Generative models of synthetic polarization maps	95
IV.3.3	Statistical assessment and comparison with RWST-based models .	96
IV.4	Discussion	98
V	Statistical denoising and enhanced component separation methods	99
V.1	The impact of noise	100
V.1.1	Presentation of the data	100
V.1.2	Impact on power spectrum statistics	101
V.1.3	Impact on RWST statistics	103
a)	Impact on first-order coefficients	103
b)	Impact on second-order coefficients	104
V.1.4	Impact on WPH statistics	105
a)	Derivation of the statistics	105
b)	Impact on the $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\hat{S}^{(0,1)}$ coefficients	105
c)	Impact on the $\hat{C}^{(0,1)}$ and \hat{C}^{phase} coefficients	107
V.2	A statistical denoising algorithm with the WPH	108
V.2.1	Method and validation	109
a)	Description of the method	109
b)	Validation on a simulation	110
V.2.2	Application to <i>Planck</i> polarization data	115
a)	Presentation of the data	115
b)	Denoising results	115
V.2.3	Comparison to other methods	117
a)	Wiener filtering and posterior sampling	118
b)	GNILC	120
V.2.4	Discussion	120
V.3	Toward enhanced component separation methods	121
V.3.1	Extension of the statistical denoising algorithm	121
V.3.2	Application to the BICEP2 field	122
a)	Presentation of the data	122
b)	Results	123
V.3.3	Perspectives	124
Conclusion and perspectives		127

A Random fields: definitions and elementary properties	129
A.1 Basic concepts	129
A.1.1 Random processes and random fields: definitions	129
A.1.2 Elementary analysis tools	130
A.1.3 Common properties	132
A.2 Spectral analysis	133
A.2.1 Power spectrum: definition and properties	133
A.2.2 Power spectrum: estimation	135
A.2.3 Bispectrum: definition	136
A.3 Gaussian random fields	137
A.3.1 Definition	137
A.3.2 Characterization	138
A.3.3 Examples	139
A.3.4 Simulation	139
A.3.5 Gaussianization	141
a) Rigorous method	141
b) Approximate method: phase randomization	142
B Elements of Fourier and wavelet analysis	145
B.1 The Fourier transform	145
B.1.1 Definition on $L^2(\mathbb{R}^n)$ and elementary properties	145
a) Definition	145
b) Elementary properties	146
B.1.2 Finite discrete signals and the discrete Fourier transform	147
a) Definition	147
b) Properties	148
c) The fast Fourier transform	148
B.2 Complements on wavelets	149
B.2.1 Wavelets used for this thesis	149
a) Morlet wavelets	149
b) Bump-steerable wavelets	150
B.2.2 Convolution with a wavelet	150
C Softwares	153
C.1 PyWST	153
C.2 PyWPH	154
Bibliography	155

Introduction

In March 2014, the BICEP collaboration claimed to have detected an imprint of primordial gravitational waves in the polarization signal of the cosmic microwave background (CMB). The announcement was sensational. Indeed, the detection of these so-called *primordial B-modes* would have confirmed spectacularly the legitimacy of inflation models, that describe the first instants of the Universe. The consequences on our understanding of the origins of the Universe would have been profound, putting an end to a quest that had been occupying cosmologists for decades. If only the observed signal was not misinterpreted... The BICEP2 instrument, a polarimeter based in the South Pole, had observed a region of almost 400 square degrees at 150 GHz during three seasons from 2010 to 2012 [1]. The collaboration knew that, at this frequency, they would have to deal with contaminants on top of the CMB. However, although they were convinced they had treated them well, soon after the official announcement, the *Planck* collaboration raised doubts on the detection. Results obtained with the *Planck* satellite, which were to be published shortly after, were instead attributing the detected signal to something much less dizzying than gravitational waves from the very early Universe, namely dust from our own Galaxy. The *Planck* collaboration's dispute prompted the BICEP team to temper their claims. In June, BICEP conceded that their models of dust were "not sufficiently constrained by external public data to exclude the possibility of dust emission bright enough to explain the entire excess signal." [1]. A joint analysis by the *Planck* and BICEP collaborations, published in March 2015, finally confirmed these doubts [2].

This eventful episode in the history of cosmological research, not only provided a stark reminder that prudence is warranted in any scientific analysis, but also shed light on a significant hurdle on the path to the detection of the primordial *B*-mode signal, that is the accurate characterization of the dust emission from our own Galaxy. This dust is part of the interstellar medium (ISM). It is coupled to the Galactic magnetic field, and emits a polarized thermal radiation in a frequency range overlapping that of the CMB (see Fig. 0.1 for an illustration). The spatial distribution of this emission is complex, namely highly non-Gaussian, while the CMB is very well described by a Gaussian random field. This thesis focuses on the statistical modeling of the non-Gaussian structure of dust polarized emission. This non-Gaussianity relates to the very nonlinear physics of the ISM. Accurate statistical models of the dust emission are thus not only necessary to ensure a robust detection of primordial *B*-modes, but would also contribute to improve our understanding of the physics of the ISM.

Statistical models consist in approximating the statistical properties of a given signal is a self-consistent way. A spatial statistical model is a random field, and when realizations can be efficiently drawn from this random field, the model is said to be generative. In

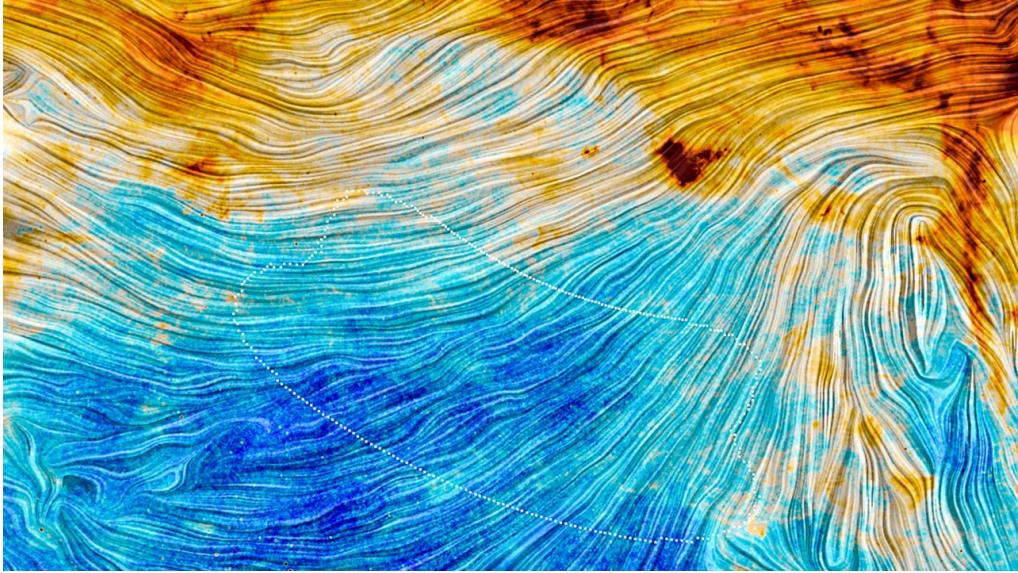


Fig. 0.1 The BICEP2 field, delimited by dashed white lines, is contaminated by the thermal emission of interstellar dust at 150 GHz. This emission (in colors: brown when intense, and blue when weak) is correlated to the structure of the Galactic magnetic field, which is represented by the drapery patterns (see Chapter I for further explanations). Credits: ESA/Planck collaboration.

this work, I explore a new avenue for defining generative models of dust emission, made possible by recent advances in data science. These are based on the wavelet scattering transform (WST) [3, 4] and wavelet phase harmonics (WPH) [5, 6], that provide multiscale representations of signals faithfully encoding interactions between scales. These are key to account for the non-Gaussian properties of the emission. Although the WST and the WPH are inspired by the structure of convolutional neural networks (CNNs), they do not involve any training stage, so that generative models based on WST and WPH can be built from a single map. Moreover, the WST and WPH descriptions allow to analyze dust emission maps from a new perspective, which can be related to the physical properties of the ISM to some extent.

This thesis lies at the crossroads of very different fields: cosmology, focusing on the origin and formation of the Universe; astrophysics, focusing on the physics of the objects of the Universe; and data science, addressing the general problem of extracting knowledge from data. My work aims to bring these different fields into a dialogue. Its main outcome is a new statistical view on the magnetized ISM.

This manuscript is organized as follows. In Chapter I, I introduce the scientific context of the thesis: the quest motivating CMB polarization measurements and the modeling of the foreground ISM. In Chapter II, I review the motivations for defining a statistical model of the dust emission, and describe the state of the art in this field. In Chapter III, I introduce the WST and employ it to describe dust polarization maps from simulated data. In Chapter IV, I introduce the WPH, and define generative statistical models of the same data based on their WST and WPH statistics. In Chapter V, I move from the "perfect" simulated data and address the problem of defining relevant models from

noisy observational data. This leads to the introduction of a statistical denoising algorithm allowing to retrieve the statistical properties of the noise-free emission as deeply as possible. The manuscript also includes three appendices. In Appendix A, I introduce mathematical notions pertaining to random fields and their elementary properties. In Appendix B, I give elements of Fourier and wavelet analysis that are relevant for this thesis. In Appendix C, I introduce PyWST and PyWPH, two public Python packages I have been developing for the purposes of this work.

Publications related to this thesis

I have been involved in the publication or submission of five papers that are directly related to this thesis. I am the first author of two of these, and the results of these two papers (or updated results) are all included and discussed in this manuscript. The references are given below.

- [i] E. ALLYS, F. LEVRIER, S. ZHANG, C. COLLING, B. REGALDO-SAINT BLANCARD, F. BOULANGER, P. HENNEBELLE & S. MALLAT; “The RWST, a comprehensive statistical description of the non-Gaussian structures in the ISM”; *Astronomy & Astrophysics* **629**, A115 (2019). [1905.01372](#).
- [ii] B. REGALDO-SAINT BLANCARD, F. LEVRIER, E. ALLYS, E. BELLOMI & F. BOULANGER; “Statistical description of dust polarized emission from the diffuse interstellar medium - A RWST approach”; *Astronomy & Astrophysics* **642**, A217 (2020). [2007.08242](#).
- [iii] J.-B. DURRIVE, P. LESAFFRE, T. GHOSH & B. REGALDO-SAINT BLANCARD; “A method to statistically characterize turbulent data with physically motivated parameters, illustrated on a centroid velocity map”; (submitted). [2101.07205](#).
- [iv] B. REGALDO-SAINT BLANCARD, E. ALLYS, F. BOULANGER, F. LEVRIER & N. JEFFREY; “A new approach for the statistical denoising of *Planck* interstellar dust polarization data”; *Astronomy & Astrophysics* **649**, L18 (2021). [2102.03160](#).
- [v] N. JEFFREY, F. BOULANGER, B. D. WANDELT, B. REGALDO-SAINT BLANCARD, E. ALLYS & F. LEVRIER; “Single frequency CMB B-mode inference with realistic foregrounds from a single training image”; *Monthly Notices of the Royal Astronomical Society: Letters* (2021). [2111.01138](#).

Index of main abbreviations and notations

CIB	Cosmic infrared background
CMB	Cosmic microwave background
CNN	Convolutional neural network
DFT	Discrete Fourier transform
FWHM	Full-width at half-maximum
ISM	Interstellar medium
GMF	Galactic magnetic field
GRF	Gaussian random field
MBB	Modified blackbody
MHD	Magnetohydrodynamics or magnetohydrodynamical
PDF	Probability density function
RMS	Root mean square
RWST	Reduced wavelet scattering transform
SED	Spectral energy distribution
SNR	Signal-to-noise ratio
WPH	Wavelet phase harmonics
WST	Wavelet scattering transform

$\llbracket a, b \rrbracket$	For $a < b \in \mathbb{Z}$, the set of integer numbers $\{a, a + 1, \dots, b\}$
$f \star g$	Convolution of f and g
\bar{z} or z^*	Complex conjugate of z
$\mathcal{F}[f]$	Fourier transform of the function f
$\hat{f}(\mathbf{k})$	Fourier mode of f
$E[X]$	Expected value of the random variable X
$\text{Cov}[X, Y]$	Covariance of the random variables X and Y

We will refer to random variables/vectors/fields with capital letters (e.g. X), and to their realizations with lowercase letters (e.g. x). We will also use bold letters to refer to vectors or matrices of finite-dimensional vector spaces (e.g.: $\mathbf{r} \in \mathbb{R}^2$, $\boldsymbol{\Sigma} = (\Sigma_{i,j})_{1 \leq i,j \leq N}$ matrix of size $N \times N$). Statistical estimators or estimates of given quantities may be denoted by a hat symbol (e.g.: \hat{x} is an estimate of x). The context will, hopefully, prevent any confusion in case of conflicting notations.

Chapter I

From the observation of the cosmic microwave background to the physics of the interstellar medium

Since its discovery in 1965 by Penzias & Wilson, the cosmic microwave background (CMB) has been one of the main focuses of modern cosmology. This primordial signal, which provides an outstanding window on the early Universe, has been crucial to develop and constrain models of the early Universe and of the subsequent formation of the large-scale structure of the Universe. Its observation and interpretation, from the monopole to the anisotropies, keeps challenging astrophysicists, statisticians, and engineers. If the fluctuations of temperature are now measured up to scales of a few arcminutes with a precision only limited by the unavoidable cosmic variance, the polarization signal, on the other hand, keeps hiding secrets.

Models of inflation, which describe a phase of accelerated expansion at the very beginning of the Universe, explain the origin of the anisotropies of the CMB, and predict the existence of primordial gravitational waves which would have imprinted on the polarization signal of the CMB a certain type of patterns called *B-modes*. While remaining highly speculative, these models would be spectacularly confirmed and constrained if one could find these primordial *B*-modes in the polarization signal of the CMB. For this reason, these *B*-modes are eagerly sought by cosmologists and motivate a variety of upcoming CMB polarization experiments.

Since the *Planck* mission, the search for primordial *B*-modes has become closely entwined with the physics of the interstellar medium (ISM). Indeed, the thermal emission of interstellar dust from our own Galaxy severely contaminates the CMB signal, acting as a foreground to the primordial signal. Component separation methods still fail to properly disentangle the finest details of the polarization signal of the CMB and the filamentary structures of the dust foreground. Polarization properties of the dust emission are the consequence of the alignment of nonspherical dust grains with the Galactic magnetic field. A detailed characterization of the properties of these nonspherical grains and of the pervasive Galactic magnetic field is needed to hopefully, one day, detect cosmological *B*-modes.

This chapter introduces this broad context. We first give an overview of how fas-

cinating CMB science can be for cosmologists who aim at building a consistent picture describing almost 14 billion years of history of our Universe with essentially a single model involving a few parameters only. Then, we introduce the usual observables characterizing the CMB, namely Stokes parameters I , Q , and U , before presenting how the *Planck* satellite has measured them on the whole sky, thus providing invaluable data for cosmologists and astrophysicists. Finally, we describe the beautifully complex physics of the ISM and of its interstellar dust grains, which thermally radiate on top of the CMB and mask its finest details.

I.1 The first light of the Universe

Modern cosmology started with the discovery of the expansion of the Universe. We review this incredible discovery and show how it suggested that photons from the early Universe could be observable at present time. We give some keys to understand the stakes in the observation of this first light of the Universe, called in an esoteric way *cosmic microwave background* (CMB).

I.1.1 An expanding Universe

The discovery of the expansion of the Universe in the 1920s was made possible thanks to the conjunction of remarkable observations and the emergence of a theory that has profoundly changed our understanding of the world.

The astronomer Vesto Slipher, at the end of the 1910s, had already remarked that most of the observable spiral nebulae were *redshifted*, i.e. that their electromagnetic spectrum of emission was almost always shifted toward lower frequencies. Of course he did not know that these nebulae were in fact what we now call galaxies. Indeed, the nature of galaxies was still mysterious at that time, and they were simply called *nebulae* because of their blurry aspect. Nevertheless, it did not prevent him from trying to explain their observed redshifts. The common explanation for that, at that time, and even today, is that this frequency shift must be attributed to a relative motion between the observer (in that case, us), and the emitter (galaxies) in the radial direction. This phenomenon, called *Doppler effect*, is well described mathematically and relates frequency shifts to the radial relative velocity between the observer and the emitter. Thanks to the measurements of these redshifts, defined by $z = \lambda_o/\lambda_e - 1$, where λ_e is the wavelength at the emission (non-distorted) and λ_o is that measured by the observer (distorted), Slipher determined the apparent radial velocity of these objects and found that it was as if galaxies were moving away from us... [7, 8]

In parallel to Slipher's observations, Einstein and his theories of relativity (special and general) had already started a revolution in science: time and space cannot be treated in an independent way [9], and even more shocking, the matter, and more generally, the energy, deforms the geometry of this so-called *spacetime* [10]. After such conceptual breakthroughs, it seemed almost natural for theoreticians to apply this new general relativity theory to describe the Universe as a whole. Assuming that the large scale Universe can be described as homogeneous and isotropic (what is called the *cosmological principle*), the geometry of the four-dimensional spacetime, defined by its metric, formally allows the idea of an expansion of space. Friedmann [11], then in an independent

way, Lemaître [12], established and interpreted the expression of what we call today the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which, assuming no curvature of space¹, reads in cartesian coordinates (t, x, y, z) :

$$ds^2 = -c^2 dt^2 + a(t)^2 (dx^2 + dy^2 + dz^2), \quad (\text{I.1})$$

with c the speed of light, and $a(t)$ a function of time called the *scale factor*. This metric defines distances in spacetime, and what should draw the attention of the reader is the fact that this scale factor function $a(t)$, which depends on the time variable t , acts as a scaling to the spatial part of this metric. Therefore, within this formalism, spatial distances evolve in time because of this scale factor. This is precisely what the expansion of the Universe is about.

Lemaître in 1927 [12], and Hubble in 1929 [14], soon realized that this scale factor, a degree of freedom in the geometry of the Universe, may explain the observed redshift of distant galaxies. In the framework of general relativity, the measured redshift z of galaxies, becomes related to the scale factor by:

$$z(t_e) = \frac{a(t_0)}{a(t_e)} - 1, \quad (\text{I.2})$$

where t_e is the time of light emission, and t_0 is that of the observation, i.e. "today's time". By convention, we usually define $a(t)$ such that $a(t_0) = 1$, so that $z = 0$ refers to today's time and strictly positive values correspond to the past. Within this framework, the observed redshift in the spectrum of galaxies is no longer the signature of a relative velocity between galaxies, a Doppler shift, but rather interpreted as the consequence of the expansion of space itself. Therefore, to explain the tendency for galaxies to be redshifted, the scale factor $a(t)$ must be an increasing function of time.

General relativity tells us that the geometry of the homogeneous Universe, described by the FLRW metric, is governed by its content in energy (or, in an equivalent way, in matter). This statement naturally raises the question of what the Universe is made of. This has been one of the main focuses of cosmology since the discovery of the expansion of the Universe, and even if we are far from having solved it, a simple phenomenological model explaining most of our cosmological observations has progressively emerged. This model, called the *standard cosmological model*, or ΛCDM model, distinguishes different forms of energy in the Universe:

- dark energy (denoted by Λ), a mysterious yet dominant form of energy thought to be responsible of the acceleration of the expansion of the Universe today [15, 16],
- dark matter, an unknown form of matter, thought to be non-baryonic, cold (i.e. with typical velocities far smaller than the speed of light), and interacting gravitationally with matter,
- baryonic matter, meaning ordinary matter, primarily constituted of baryons and notably referring to all the visible Universe,
- radiation, made of photons and neutrinos.

¹Current measurements are consistent with a null spatial curvature [13].

We refer to general textbooks for a detailed presentation of the Λ CDM model as well as its limits (e.g. [17, 18]). However, we choose to discuss here one central equation of this model which formally shows the links between the geometry and the content in energy of the Universe, and illustrates the challenges of modern cosmology. Introducing the Hubble parameter $H(t) = \dot{a}(t)/a(t)$, with \dot{a} the time derivative of a , as well as the cosmological density parameters evaluated at present time $\Omega_{\Lambda,0}$, $\Omega_{c,0}$, $\Omega_{b,0}$, and $\Omega_{r,0}$, corresponding to the density parameters of the dark energy, cold dark matter, baryonic matter, and radiation, respectively,² the first Friedmann equation in the simpler case where we assume a flat Universe (i.e. with no spatial curvature) reads:

$$\left(\frac{H(t)}{H_0}\right)^2 = \frac{\Omega_{r,0}}{a(t)^4} + \frac{\Omega_{c,0} + \Omega_{b,0}}{a(t)^3} + \Omega_{\Lambda,0}, \quad (\text{I.3})$$

where $H_0 = H(t_0)$ is the Hubble constant. On the right hand side, essentially stands the sum of the energy densities of each of the components of our Universe. We first see that these densities scale differently with the expansion factor depending on their nature. The expansion of the Universe thus suggests that in the past, when $a(t)$ was much smaller, radiation must have been the dominant form of energy (what is called *radiation era*), before subsequently giving way to the dark and baryonic matter (*matter era*), while today the dark energy seems to be the dominant form of energy (*dark energy dominated era*). The behavior of the scale factor now clearly appears as depending on a few parameters measurable at present time: the Hubble constant, and the cosmological density parameters. The values of these parameters are well measured today, and we report those determined by the *Planck* mission in Table I.1. In particular, these allow to give an age to the Universe, which is identified to the time t_U at which the scale factor reached zero. From Eq. (I.3), one gets the following expression for t_U :

$$t_U = \frac{1}{H_0} \int_0^1 \frac{1}{a \sqrt{\Omega_{r,0} a^{-4} + (\Omega_{c,0} + \Omega_{b,0}) a^{-3} + \Omega_{\Lambda,0}}} da, \quad (\text{I.4})$$

so that, according to *Planck* measurements, the Universe would be approximately 13.79 billion years old [13]. Note that the moment at which the scale factor vanishes corresponds to a singularity in the metric of the spacetime, which is often called *Big Bang*, but this is no more than a mathematical curiosity as far as we know since our physical theories are no longer valid to describe this moment. It is also interesting to notice that dark energy and dark matter constitute almost 95% of the content in energy of the Universe today, meaning that in the framework of this standard model, the vast majority of our Universe seems to be governed by some unknown form of energy and matter. For these reasons, the nature of dark energy and dark matter is without any doubt, one of the greatest puzzles of modern cosmology.

In the 1940's, the expansion of the Universe introduced the idea of a hot Big Bang model, describing the early Universe as much more dense and hot than what it is today. In this framework, in 1948, Alpher, Bethe and Gamow attempted to explain the observable relative abundances of chemical elements in the Universe, paving the way to what we

²The cosmological density parameter Ω_X is the ratio of the average density of matter/energy of the component X to what is called the *critical density*, which corresponds to the density at which the Universe would stop expanding after an infinite time [17].

$\Omega_{c,0}$	$\Omega_{b,0}$	$\Omega_{r,0}$	$\Omega_{\Lambda,0}$	H_0 [km.s $^{-1}$.Mpc $^{-1}$]
0.27	0.049	5.4×10^{-5}	0.68	67

Table I.1 – Measured values of the Hubble constant and the cosmological density parameters at present time, as determined by the *Planck* mission [13], except for $\Omega_{r,0}$ which is estimated based on the temperature of the CMB [19, 17]. These values are given on an indicative basis with two significant digits only. The accurate measurements, as well as their error bars, depend on the combination of data used to constrain these parameters (see [20]).

now call the *big bang nucleosynthesis theory* (BBN) and which predicts the abundance of light elements (mostly hydrogen, helium, and their isotopes) in the Universe remarkably well [21, 20]. A few months later, within the same picture, Gamow, Alpher, and Herman, predicted that a relic radiation of the early Universe should be observable today, with an estimated temperature of 5 K [22, 23]...

I.1.2 The discovery of the cosmic microwave background

Within the hot Big Bang model, the picture of the early Universe is that of a soup of particles in a thermodynamical equilibrium which has progressively cooled down as the Universe expanded. Particles, which were all coupled at the beginning of the Universe, successively decoupled from each other with the decrease of the temperature and of the energy densities. Right after the BBN, the Universe was primarily made of a fully ionized plasma composed of nuclei of hydrogen and helium, electrons and photons in strong interaction. The temperature had to be high enough to prevent a recombination of electrons with nuclei. Photons were continuously scattered by electrons, mostly through Compton scattering, making the plasma completely opaque to radiation. The spectral radiance $B_\nu(T)$ of the radiation at that time is thought to correspond to that of a blackbody radiation:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1}, \quad (\text{I.5})$$

where k is the Boltzmann constant, and h is the Planck constant. Due to the expansion of the Universe, it can be shown that the temperature of this blackbody radiation as a function of time, or in an equivalent way as a function of the cosmological redshift z , must vary as $T(z) \propto 1+z$ (note that this linear law is supported by observations, see e.g. [24]). Therefore, one expects that with the expansion, the temperature of the radiation could sufficiently decrease to allow the recombination of electrons with nuclei. This recombination happened at a cosmological redshift $z_{\text{rec}} \approx 1100$ [13], which corresponds to a time situated approximately 380,000 years after the Big Bang, and when the temperature of the Universe was approximately 3000 K. Right after this recombination, photons were finally able to freely propagate through the Universe, defining what we call the last scattering surface, and giving rise to the cosmic microwave background (CMB), that is the first light of the Universe.

At the beginning of the 1960s, Robert Dicke and his colleagues were precisely looking for this cosmic radiation. They built a radiometer designed to find it, but were scooped

by two researchers of Bell Telephone Laboratories, Arno Penzias and Robert Wilson, who found out, by coincidence, an unexpected excess of antenna temperature³ while testing their own radiometer in 1964 [25]. They reported an excess of 3.5 ± 1 K of unknown origin at 4080 MHz. It did not take long for Dicke and his colleagues Peebles, Roll and Wilkinson, to interpret this result as a signature of a cosmic radiation coming straight from the very beginning of the Universe: the CMB [26]. Since this discovery, it has been shown that the observed CMB corresponds to an almost perfect blackbody radiation at temperature $T_0 = 2.725$ K [19], confirming in a spectacular way the picture of the early Universe painted by the hot Big Bang model.

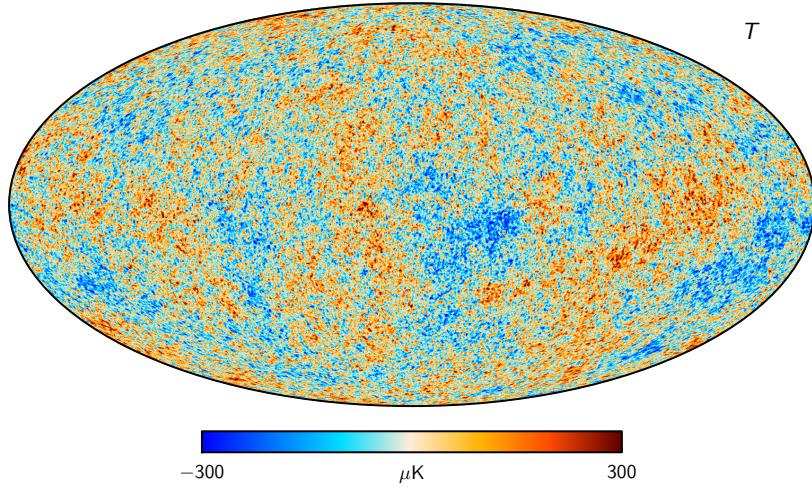
We have seen that the Λ CDM model, and the picture of the hot Big Bang model, explains very well the homogeneous Universe. It explains the expansion of the Universe, the relative abundance of light elements in the Universe, and the CMB. However we know that the Universe cannot be assumed to be homogeneous at all scales, since the observed distribution of galaxies and dark matter is not homogeneous. The Universe exhibits a structure which is that of the *cosmic web*, and which requires to include in this model an explanation for these inhomogeneities [27].

I.1.3 From inhomogeneities to anisotropies

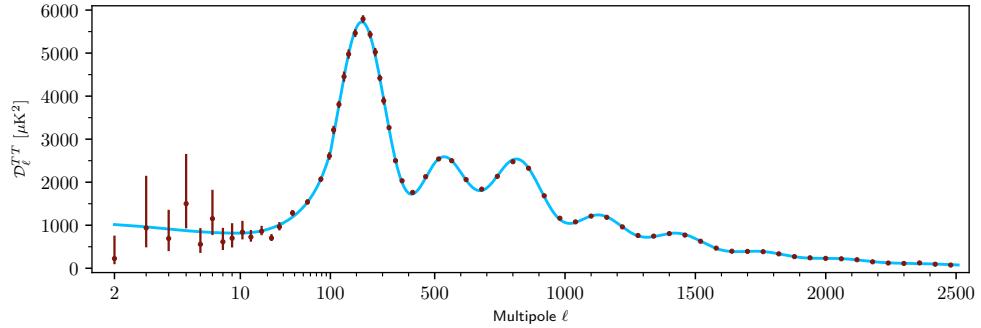
The large-scale structure of the Universe, that is the distribution of dark matter and galaxies at scales typically larger than ~ 1 Mpc, exhibits a clear inhomogeneity. Patterns of galaxies form the so-called *cosmic web*, made of filaments, sheets, walls, and voids. These are thought to arise from the nonlinear evolution of primordial density fluctuations through gravitational instability in an expanding Universe [27, 28]. Therefore, a signature of these primordial density fluctuations is expected in the CMB, taking the form of anisotropies in the signal.

These primordial anisotropies in the CMB are so small that it took almost 30 years after the discovery of the CMB to find them. The first detection was made with the COBE satellite in 1992 [31], and ever since, cosmologists have been trying to improve the measurements, driving the transformation of cosmology into the high-precision science it is today. Thanks to the WMAP mission in the 2000s [32], and the *Planck* mission in the 2010s [13], we now have a detailed view of the fluctuations of temperature of the CMB on the whole sky. We show in Fig. I.1a the map of the fluctuations of the CMB temperature on the sky as determined by the *Planck* mission. This map shows that typical fluctuations of temperature are approximately 100 μK , which gives relative fluctuations in temperature of the CMB of $\sim 10^{-5}$, exhibiting more quantitatively the need for highly accurate measurements for the proper detection of these anisotropies. A statistical characterization of the map shows that the fluctuations of the CMB temperature seem to have Gaussian statistics and to be statistically isotropic [33]. In other terms, this means that the CMB is well modeled from a statistical point of view by a realization of an isotropic Gaussian random field on the sphere (see Appendix A for a presentation of Gaussian random fields and their elementary properties). Consequently, all the statistical information included in these fluctuations of temperature across the sky, simply called *temperature anisotropies*, rests in the angular power spectrum of this map, which is shown in Fig. I.1b. On this

³The antenna temperature T_A describes how much noise an antenna produces in a given environment. It is related to the noise power per unit bandwidth P_N by $P_N = kT_A$.



(a) Map of the fluctuations of temperature of the CMB (SMICA map) in Mollweide projection. Figure is taken from [29].



(b) Angular power spectrum of the temperature map of the CMB. The solid line corresponds to the best-fit of the ΛCDM model [30]. Figure is taken from [13].

Fig. I.1 Temperature map of the CMB and its angular power spectrum as determined by the *Planck* mission [13].

figure, we see that this angular power spectrum is very well fitted by the ΛCDM model, thus giving extremely accurate estimates of its parameters (some of which being shown in Table I.1). Note that the peaks in the angular power spectrum show the imprint of acoustic oscillations in the primordial plasma due to the tight coupling between photons and baryonic matter.

We only show here the anisotropies of the CMB temperature, but anisotropies in the CMB linear polarization signal are also expected due to the effect of Thomson scattering⁴ in the last scattering surface [34, 17]. These anisotropies in the polarization signal have been meticulously observed, but the measurements are not sufficiently accurate to answer current cosmological questions, such as the existence of primordial gravitational waves [35]. Indeed the signal is weaker, so the instrumental noise quickly becomes problematic, and the so-called foregrounds, mainly due to interstellar dust and Galactic

⁴Thomson scattering refers to the low-energy limit of Compton scattering.

synchrotron emission are prominent contaminants. We will give more details on these contaminants in the next section. In this thesis, the focus is the modeling of the dust foreground, which is motivated by this need to improve our measurements of the CMB polarization signal.

To account for the observed inhomogeneities in the last scattering surface, the FLRW metric must be perturbed, giving rise to scalar, vectorial, and tensorial degrees of freedom [18, 17]. Scalar perturbations generate temperature anisotropies and a certain type of polarization signal, called *E*-modes, while vectorial and tensorial perturbations, in addition to generating temperature anisotropies and *E*-modes, must generate complementary patterns in the polarization signal, called *B*-modes [36, 37, 38]. The detection of tensorial perturbations, corresponding to primordial gravitational waves, is one of the main goals of modern cosmology. While CMB *E*-modes have already been well measured, primordial *B*-modes still remain undetected, even if we were able to set upper limits to their amplitude [2, 30, 39]. Note that the current constraints on the *B*-modes tensor-to-scalar ratio $r < 0.06$ (95% confidence level [40]) are expected to be significantly improved by the next generation of CMB experiments, such as ACT [41], SPIDER [42], LiteBIRD [43], PICO [44], the Simons Observatory [45], and CMB-S4 [46], with a detection limit goal of $r \simeq 10^{-3}$. Nevertheless, any claim to the detection of cosmological *B*-modes will have to be critically assessed against alternative explanations involving Galactic foregrounds as we will discuss it in the next section.

Another question remains: what is the origin of these perturbations in the FLRW metric? So-called *inflation models*, which were imagined to solve the defects of the hot Big Bang model (including the horizon, smoothness, flatness, and monopole problems [18, 17]), could also give a satisfactory answer [47, 48]. These models, which describe the very first instants of the Universe postulating a phase of accelerated expansion, explain the origin of these perturbations as arising from a quantum noise whose fluctuations would have been frozen out as the Universe was inflating. Still highly speculative, models of this inflationary Universe demand to be confirmed by observations. As most inflation models predict tensorial modes in the perturbed FLRW metric, the detection of their imprint on the CMB *B*-modes would help to confirm the validity of these models and constrain their parameters [39].

I.2 Observing the CMB

I.2.1 The observables: Stokes parameters

The CMB radiation is usually described in terms of Stokes parameters. These are convenient to describe the intensity and polarization of a given beam of light as they directly correspond to sum or differences of measurable intensities.

Let us consider a plane wave propagating along the direction (Oz), and let us call E_x and E_y the complex components of its corresponding electric field $\mathbf{E}(\mathbf{r}, t)$ in a fixed orthonormal basis $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$. We thus have:

$$\mathbf{E}(\mathbf{r}, t) = E_x(\mathbf{r}, t)\hat{\mathbf{e}}_x + E_y(\mathbf{r}, t)\hat{\mathbf{e}}_y. \quad (\text{I.6})$$

The Stokes parameters are a set of 4 parameters, called I , Q , U , and V , defined as follows:

$$I = \langle |E_x|^2 \rangle + \langle |E_y|^2 \rangle, \quad (\text{I.7})$$

$$Q = \langle |E_x|^2 \rangle - \langle |E_y|^2 \rangle, \quad (\text{I.8})$$

$$U = 2 \operatorname{Re} (\langle E_x E_y^* \rangle), \quad (\text{I.9})$$

$$V = -2 \operatorname{Im} (\langle E_x E_y^* \rangle), \quad (\text{I.10})$$

where $*$ denotes the complex conjugate, and $\langle \cdot \rangle$ corresponds to a time average related to the instrument. The parameter I is simply the total intensity, Q and U measure the intensity of linear polarization, and V measures the intensity of circular polarization. By construction, the Stokes parameters are constrained by the following inequality:

$$Q^2 + U^2 + V^2 \leq I^2, \quad (\text{I.11})$$

which becomes an equality in the case of a monochromatic plane wave.

To better interpret these parameters, let us rewrite them in a more "operational" form. We define two additional bases $(\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'})$ and $(\hat{\mathbf{e}}_l, \hat{\mathbf{e}}_r)$ from transformations of the initial one $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$ with:

$$\hat{\mathbf{e}}_{x'} = \frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_x + \hat{\mathbf{e}}_y), \quad \hat{\mathbf{e}}_{y'} = \frac{1}{\sqrt{2}} ((-\hat{\mathbf{e}}_x + \hat{\mathbf{e}}_y)), \quad (\text{I.12})$$

$$\hat{\mathbf{e}}_l = \frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_x + i\hat{\mathbf{e}}_y), \quad \hat{\mathbf{e}}_r = \frac{1}{\sqrt{2}} (\hat{\mathbf{e}}_x - i\hat{\mathbf{e}}_y). \quad (\text{I.13})$$

$(\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'})$ is simply a 45° rotation of $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$, while $(\hat{\mathbf{e}}_l, \hat{\mathbf{e}}_r)$ defines a circular basis. With these new definitions, we can rewrite the expressions of the Stokes parameters as follows:

$$I = \langle |E_x|^2 \rangle + \langle |E_y|^2 \rangle, \quad (\text{I.14})$$

$$Q = \langle |E_x|^2 \rangle - \langle |E_y|^2 \rangle, \quad (\text{I.15})$$

$$U = \langle |E_{x'}|^2 \rangle - \langle |E_{y'}|^2 \rangle, \quad (\text{I.16})$$

$$V = \langle |E_r|^2 \rangle - \langle |E_l|^2 \rangle, \quad (\text{I.17})$$

where $(E_{x'}, E_{y'})$ and (E_l, E_r) are the corresponding transformed coordinates of \mathbf{E} . Now the Stokes parameters effectively appear as sums or differences of measurable intensities. We show in Fig. I.2 a few elementary examples of correspondences between the polarization of a monochromatic plane wave and the values of its Stokes parameters.

An important property of the Stokes parameters stems from the way they are transformed when rotating the initial basis $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$. For a rotation of this frame by a direct angle θ , they are transformed as follows:

$$I \longrightarrow I' = I, \quad (\text{I.18})$$

$$Q + iU \longrightarrow Q' + iU' = e^{-2i\theta} (Q + iU), \quad (\text{I.19})$$

$$V \longrightarrow V' = V, \quad (\text{I.20})$$

where we introduce the complex variable $Q + iU$ as a convenience to jointly denote Q and U . Therefore, we see that I and V are invariant under the rotation of the initial basis, while $Q + iU$ transforms as a spin-2 quantity.

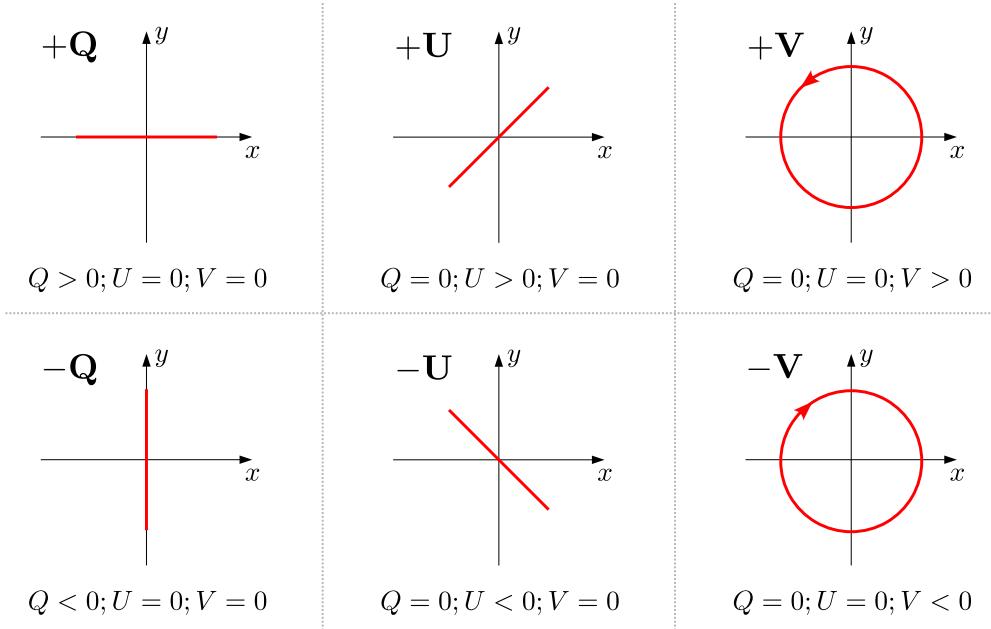


Fig. I.2 Correspondence between polarization and values of the Stokes parameters on a few common examples in the case of a monochromatic plane wave.

In a cosmological context, linear polarization is not often studied straight from the raw Q and U observables, but rather in terms of E and B -modes [49]. E and B -modes are nonlocal linear transforms of Q and U defined as follows in Fourier space (see e.g. [35]):

$$\begin{pmatrix} \hat{E}(\mathbf{k}) \\ \hat{B}(\mathbf{k}) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos 2\varphi_{\mathbf{k}} & \sin 2\varphi_{\mathbf{k}} \\ -\sin 2\varphi_{\mathbf{k}} & \cos 2\varphi_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} \hat{Q}(\mathbf{k}) \\ \hat{U}(\mathbf{k}) \end{pmatrix}, \quad (\text{I.21})$$

where $\varphi_{\mathbf{k}}$ is the angle that \mathbf{k} makes with the x -axis. This definition is given here in a flat-sky approximation, although, in the literature, it is common to find definitions of E and B -modes on the sphere. We refer the reader to [37, 35] for such definitions. The motivation for this transformation is twofold. First, this decomposition of the $Q + iU$ variable is expected to disentangle scalar primordial fluctuations from vectorial and tensorial ones [36]. As explained in Sect. I.1.3, vectorial and tensorial fluctuations in the primordial plasma before the decoupling would give rise to a non-zero B -mode signal in the CMB, while scalar fluctuations cannot source any B -mode signal. Second, contrary to the $Q + iU$ variable which is a spin-2 variable, E and B variables are scalar and pseudo-scalar⁵ variables, respectively, which means in particular that these variables are invariant under any rotation of the initial basis chosen for the measurement of the polarization signal. In this thesis, we will work with the $Q + iU$ variable only to characterize dust polarization, but we will see in Sect. I.3.3 that important statistical properties of the dust signal are formulated in terms of these E and B variables.

⁵Contrary to scalar quantities, the sign of pseudo-scalar quantities is inverted under a parity inversion.

Note that we will not make any further mention of the Stokes parameter V since it is expected to be null for the CMB. Indeed, Thomson scattering in the last scattering surface, which is at the source of linear polarization in the CMB as explained above, cannot generate any circular polarization (see e.g. [35]).

I.2.2 The *Planck* mission

The ESA *Planck* mission is the third space mission dedicated to the measurement of the anisotropies of the CMB [13, 50] (after COBE in the 90s [31, 51] and WMAP in the 2000s and early 2010s [32, 52]). The *Planck* satellite was launched in 2009, after almost 20 years of preparation, from Kourou in French Guiana. The scientific programme of the mission was clear: "*Planck* is designed to extract essentially all information available in the temperature anisotropies and to measure CMB polarization to high accuracy." [53]. Initially planned to scan the sky during 18 months, the *Planck* satellite was finally able to do it during more than 4 years, going from 2009 to 2013. It was situated at the L₂ point, meaning it was continuously along the Sun-Earth axis, situated at approximately 1.5 million kilometers away from Earth and in the direction opposite to the Sun.

The *Planck* satellite was made of two instruments called LFI and HFI (for Low Frequency Instrument and High Frequency Instrument, respectively), with a 1.5-meter-diameter primary mirror, which were designed to observe the microwave and submillimeter⁶ sky in 9 frequency bands ranging from 30 to 857 GHz (corresponding to wavelengths ranging from 1 cm to 350 μ m, respectively), and at resolutions ranging from 32' to 4', respectively [13]. The satellite needed to be actively cooled down to reach temperatures down to 0.1 K for the bolometers of the HFI instrument.

Thanks to a sophisticated data processing [54, 55], *Planck* produced 9 all-sky frequency maps in temperature (i.e. in total intensity I), and linear polarization Q and U . We show in Fig. I.3 the 9 temperature frequency maps as obtained by the *Planck* mission, and once removed the dipole component corresponding to the Doppler distortion due to the relative velocity of the *Planck* satellite with respect to the CMB reference frame. These maps are shown in Galactic coordinates with the Galactic center in the middle of this projection. We clearly see that they significantly differ from the CMB map displayed in Fig. I.1b. What looks like the CMB appears to be superimposed to various other emissions that are spatially correlated with the structure of our own galaxy, the Milky Way. This illustrates the problem faced in retrieving the CMB signal from these observations. Retrieving the CMB demands accurate statistical techniques to disentangle "foregrounds", of astrophysical nature, from the primordial cosmological signal. These techniques, called *component separation methods*, are discussed in the next subsection.

The scientific outcome of the *Planck* mission is impressive. The mission produced the most accurate sky maps, in temperature and in linear polarization, of the CMB anisotropies, thus giving the most detailed snapshot of the early Universe when it was approximately 380,000 years old. It also provided outstanding sky maps of the astrophysical foregrounds to the CMB allowing to probe the physics of the interstellar medium

⁶The microwave domain usually refers to the radio frequencies ranging from 300 MHz to 300 GHz (corresponding to wavelengths ranging from 1 m to 1 mm, respectively), whereas the submillimeter domain typically refers to the range from 300 GHz to 1 THz (corresponding to the range from 1 mm to 300 μ m, respectively). Nevertheless, there is no consensus on these definitions, and these may slightly vary from one scientific community to another.

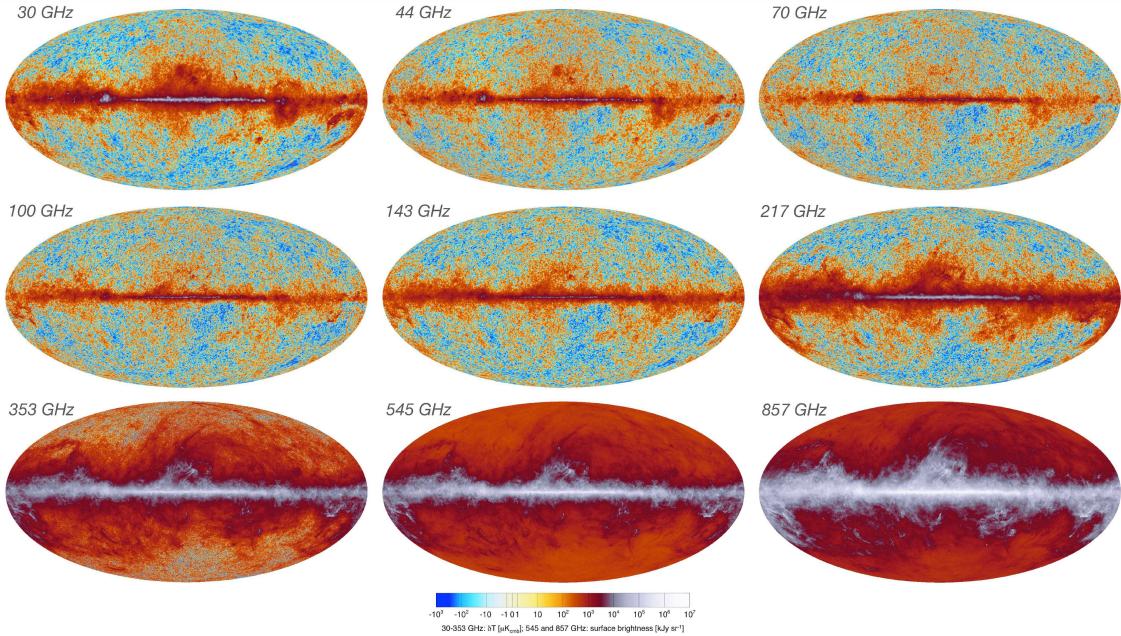


Fig. I.3 All-sky temperature fluctuations maps as obtained by the *Planck* mission for the 9 frequency bands of the satellite. The Doppler dipole component has been removed. Sky maps correspond to Mollweide projections and are displayed in Galactic coordinates with the Galactic center in the middle and longitude increasing to the left. Figure is taken from [13].

and that of the large-scale structure of the Universe [29, 56, 57, 58]. It extracted the temperature angular power spectrum of the CMB with a precision only constrained by the cosmic variance, an unavoidable variance due to the fact that we can only observe one CMB, down to angular scales of $\sim 0.1^\circ$. It also provided one of the most accurate measurements of the angular power spectra of the polarization signal, even if it clearly showed that improvements are needed in that direction to answer fundamental questions of primordial cosmology [30, 2]. From these power spectra, the *Planck* mission, complemented by other probes, was able to constrain with an outstanding accuracy the 6 parameters of the Λ CDM model. This model is able to fit these power spectra with such a precision that its legitimacy to describe the dynamical evolution of the Universe is now undeniable [20, 33]. These results also give strong constraints on the extensions of this model, including inflation models [59, 39].

I.2.3 A component separation problem

As illustrated in Fig. I.3 with *Planck* data, measuring the anisotropies of the CMB is not straightforward since the CMB emission is contaminated by various other emissions regardless of the frequency. These so-called foregrounds correspond to diffuse emissions and compact sources, originating both from our Galaxy and beyond, and for most of them, with an inhomogeneous distribution on the sky as well as complex frequency dependencies. Therefore, component separation methods are essential to disentangle the CMB from these

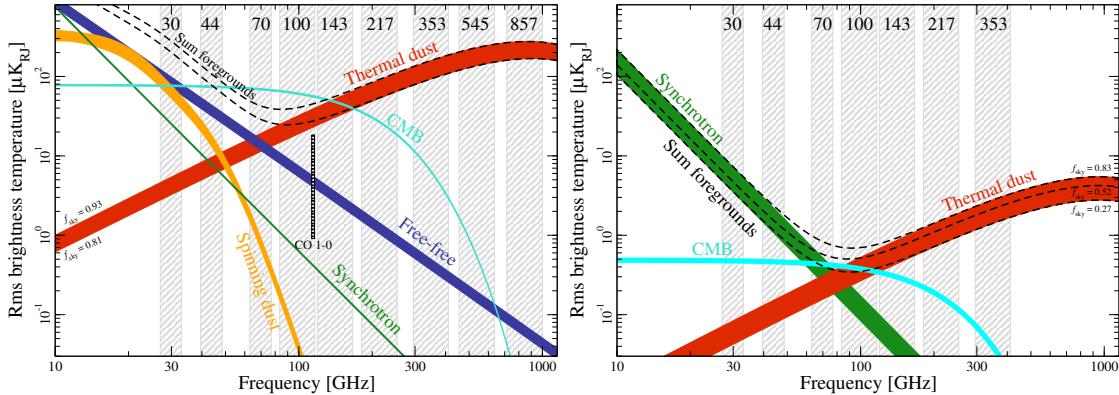


Fig. I.4 Frequency dependence of the RMS brightness temperature for the components of the microwave sky in intensity I (left), and in linear polarization intensity $\sqrt{Q^2 + U^2}$ (right). The grey bands show the bandpasses of the frequency channels of the *Planck* satellite. Figure is taken from [13].

various foregrounds. For the *Planck* mission, several statistical methods have been used, each of them having its own advantages and disadvantages. Almost all of these methods focused on retrieving the CMB emission, and some of them also produced maps of the foregrounds, which are highly valuable for astrophysicists [60, 61, 29]. We briefly review the variety of foregrounds to the CMB at the frequencies of observation of the *Planck* satellite, before discussing some of the statistical methods that have been used for *Planck* data to perform these component separations, focusing on those that produced maps of the thermal dust emission. We refer to [29] and references therein for a detailed presentation of the methods used for the *Planck* mission, and to [62] for a (pre-*Planck*) review on classical statistical approaches of diffuse source separation for CMB observations.

a) Foregrounds

The various foregrounds that have been identified as significant contaminants for the observation of the CMB temperature for the *Planck* mission are (see e.g. [61]):

- the synchrotron emission: diffuse emission due to the spiraling of relativistic electrons around Galactic magnetic fields,
- the free-free emission (or bremsstrahlung emission): of Galactic origin, corresponding to the radiation produced by collisions between electrons and ions,
- the thermal dust emission, simply called *dust emission* or *dust foreground* in the following: it is the main focus of this thesis, due to the thermal emission of interstellar dust from our own Galaxy,
- the spinning dust emission, also known as *anomalous microwave emission*: thought to be the emission of small and fast spinning dust grains with a non-zero electric dipole moment,
- rotational spectral lines of the CO molecule: line emission detected in *Planck* 100, 217, and 353 GHz frequency channels,

- the thermal Sunyaev-Zeldovich emission: of extragalactic origin, corresponding to the distortion of CMB photons scattering off electrons from hot galaxy clusters (inverse Compton scattering),
- the cosmic infrared background (CIB) emission: of extragalactic origin, corresponding to the diffuse redshifted thermal emission of dust in distant galaxies,
- various extragalactic point sources: usually removed in preprocessing.

We should also add to this list the instrumental noise and systematics, which are not actual "foregrounds" although they act as such. This non-exhaustive list shows how complex the microwave sky already is. Fortunately, these contaminants are not significant at all scales and frequencies, and most of them are not expected to exhibit linear polarization. We show in Fig. I.4 the frequency dependence of the root mean square (RMS) brightness temperature for the components of the microwave sky in temperature and polarization. We notably see that in polarization, for the *Planck* sensitivity, the only foregrounds are the synchrotron emission, and the thermal dust emission. In practice, we will see that the instrumental noise is also a main contaminant of the polarization signal.

For B -mode detection, we naturally want to design experiments focusing on frequencies where the CMB is the most visible relatively to its foregrounds. Based on Fig. I.4, targeting frequencies around 100 GHz seems reasonable. At 100 GHz, the CMB and the dust foreground are the two dominant components, which motivates the need for an accurate modeling of the dust foreground. To illustrate the component separation problem addressed by the search of primordial B -modes, we show in Fig. I.5 an artist view of what a satellite typically sees when looking at the polarized sky at 100 GHz.

b) Component separation methods

Several component separation methods have been employed for the *Planck* mission. These were thought to be representative of the variety of approaches in the literature at that time. We briefly discuss Commander, SMICA, and GNILC methods, which all have produced foregrounds maps, including maps of the dust thermal emission.

Commander The Commander algorithm ([63, 64, 61]) adopts a parametric approach. It relies on parametric phenomenological models for each component of the microwave sky, and fits the corresponding parameters in a Bayesian framework. The observed sky d_i at a frequency channel ν_i is modeled as a sum of components $s_i^{\text{CMB}}, s_i^{\text{dust}}, \dots, n_i$ (instrumental noise), that is:

$$d_i = s_i^{\text{CMB}} + s_i^{\text{dust}} + \dots + n_i. \quad (\text{I.22})$$

In particular, the frequency dependence of the dust component s_i^{dust} , corresponding to the thermal emission of interstellar dust grains, is modeled with that of a modified blackbody emission (MBB):

$$s_i^{\text{dust}}(\mathbf{r}) \propto \nu_i^{\beta(\mathbf{r})} B_{\nu_i}(T(\mathbf{r})), \quad (\text{I.23})$$

thus introducing two parameters depending of the position \mathbf{r} : the spectral index $\beta(\mathbf{r})$ and the temperature $T(\mathbf{r})$.⁷ With ν_0 a reference frequency, the model of the dust component

⁷Here, s_i^{dust} corresponds to a specific intensity, contrary to [61] where components are described in terms of brightness temperature, thus explaining slight differences between the formulae.

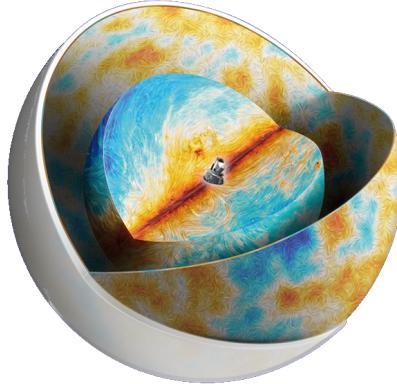


Fig. I.5 Artist view of the component separation problem addressed by the search for B -modes in the CMB. The satellite instrument, here the *Planck* satellite, shown at the center of the image, observes the sum of the polarized emission of the Galactic dust (foreground) and of the CMB (background). Credits: *Planck* collaboration, Canopée.

becomes:

$$s_i^{\text{dust}}(\mathbf{r}) = A(\mathbf{r}) \left(\frac{\nu_i}{\nu_0} \right)^{\beta(\mathbf{r})} \frac{B_{\nu_i}(T(\mathbf{r}))}{B_{\nu_0}(T(\mathbf{r}))}, \quad (\text{I.24})$$

where $A(\mathbf{r})$ corresponds to the amplitude of the dust emission at the reference frequency ν_0 and the position \mathbf{r} . The set of parameters describing the dust component thus includes an amplitude map A , a spectral index map β , and a temperature map T . Calling θ the complete set of parameters of the full sky model (including A , β and T), the Commander algorithm aims at finding the maximum of the so-called *posterior distribution* $p(\theta|\mathbf{d})$ which reads:

$$p(\theta|\mathbf{d}) = \frac{p(\mathbf{d}|\theta)p(\theta)}{p(\mathbf{d})} \propto \mathcal{L}(\theta)p(\theta), \quad (\text{I.25})$$

where $\mathcal{L}(\theta) = p(\mathbf{d}|\theta)$ is called the *likelihood* and $p(\theta)$ is called the *prior distribution* of the parameters. Assuming a Gaussian distribution for the noise as well as its independence across channels, the likelihood becomes a multivariate Gaussian distribution only depending on the noise covariance. After a relevant choice for the prior distribution, the fit finally consists in finding the parameters that maximize this posterior distribution. This is done using a method called Gibbs sampling [65], which avoids a numerically intractable direct sampling of the posterior distribution (indeed the number of parameters can go up to several hundred million).

GNILC The generalized needlet internal linear combination (GNILC) algorithm ([66]) has been originally applied to *Planck* temperature data to disentangle CIB anisotropies from dust emission [67]. In [29], it was also applied to polarization data focusing on

the extraction of a dust foreground freed from noise and CMB contaminations. This algorithm works in a spherical wavelet space, called *needlet space* [68, 69], thus allowing a separation of the data both spatially and spectrally. More precisely, calling $d_i^{(j)}$ the observed map at a frequency ν_i that has been filtered with a needlet at a scale j , for example, in the context of polarization data (Q or U data), we have:

$$d_i^{(j)} = s_i^{\text{CMB},(j)} + s_i^{\text{Gal},(j)} + n_i^{(j)}, \quad (\text{I.26})$$

where $s_i^{\text{CMB},(j)}$, $s_i^{\text{Gal},(j)}$, and $n_i^{(j)}$ correspond to the CMB, Galactic, and noise components filtered at scale j , respectively. The Galactic component $s_i^{\text{Gal},(j)}$ is estimated with a local linear weighting of the $\{d_i^{(j)}\}_i$ maps:

$$\hat{s}_i^{\text{Gal},(j)}(\mathbf{r}) = \sum_{i'} W_{ii'}^{(j)}(\mathbf{r}) d_{i'}^{(j)}(\mathbf{r}). \quad (\text{I.27})$$

The weighting matrix $\mathbf{W}^{(j)}$ of size $N_{\text{chan}} \times N_{\text{chan}}$, which thus depends on the position \mathbf{r} , is built to minimize the variance of the estimator while adapting to local contamination levels of the target signal (see [29] for a definition of $\mathbf{W}^{(j)}$). This latter constraint necessitates to estimate locally the frequency-frequency covariance matrices of the contaminants (here, the CMB and the noise), which may require prior information on the power spectrum of these contaminants.

SMICA The SMICA (spectral matching independent component analysis) algorithm [70] introduces a blind method which relies on very few assumptions on the data: known number of foregrounds, and statistical independence between the foregrounds and the CMB. This algorithm is performed in spectral space via a spherical harmonic decomposition of the signals. Each mode of the CMB s_{lm}^{CMB} is estimated as follows:

$$\hat{s}_{lm}^{\text{CMB}} = \mathbf{w}_l^\dagger \mathbf{d}_{lm}, \quad (\text{I.28})$$

where \mathbf{d}_{lm} is a vector of N_{chan} components, one for each frequency channel, and \mathbf{w}_l is a vector of N_{chan} weights, which does not depend on m . These weights are determined in order to minimize the variance of the estimator for a given l , which gives:

$$\mathbf{w}_l = \frac{\mathbf{C}_l(\theta)^{-1} \mathbf{a}}{\mathbf{a}^\dagger \mathbf{C}_l(\theta)^{-1} \mathbf{a}}, \quad (\text{I.29})$$

where \mathbf{a} is a vector of N_{chan} components describing the spectrum of the CMB (assumed to be that of a blackbody), and $\mathbf{C}_l(\theta)$ is a frequency-frequency covariance matrix of size $N_{\text{chan}} \times N_{\text{chan}}$ depending on parameters θ that have been fitted so that $\mathbf{C}_l(\theta)$ "matches" the empirical covariance matrix of the observed signal (see e.g. [29] for the definition of the matching criterion). In practice, the model for the covariance matrix $\mathbf{C}_l(\theta)$ is quite generic once the independence between the foregrounds and the CMB is assumed which explains the "blind" aspect of the method (see [29]), however this formalism allows some flexibility to add prior information on the components. In [29], this method has been adapted for the estimation of the dust foreground on polarization data with no specific

assumption on the properties of the foreground (except that the dust polarization foreground is dominant at 353 GHz, cf Fig. I.4).

From this very brief presentation of the Commander, GNILC, and SMICA algorithms, let us already make a few remarks. For Commander, the choice of the model for each of the components is obviously crucial, and such a parametric method motivates the need for a relevant model for the dust foreground. However, a parametric approach requiring a detailed knowledge of the components is not the only way, as the SMICA, and GNILC methods illustrate. SMICA can perform a fully blind component separation, only requiring statistical independence between the CMB and the foregrounds, while GNILC in some sense goes halfway between Commander and SMICA, using prior information on the local power spectrum of the contaminants when available. Finally, let us remark that all of these methods strongly exploit the frequency dimension of the observations, either by assuming spectral models for the various components, or defining estimators from linear weightings of the various frequency maps. On the contrary, the spatial information of the various emissions seems to be poorly exploited when it is not simply ignored. One major development of this thesis is the idea that the dust foreground has very rich spatial properties, namely strong interactions between scales, that deserve to be taken into account for the improvement of component separation methods.

Let us finally mention that the research for the development of component separation methods has not stopped with the *Planck* mission, and various strategies are currently in development for the improvement of these methods in the context of the search of primordial B -modes (see e.g. [71, 72, 73, 74]).

I.3 The Milky Way, its interstellar medium, and the dust foreground

I.3.1 The Milky Way and the ISM

Our galaxy, the Milky Way, is no different from most observable spiral galaxies: it is made of stars, gas, and solid particles, which are all immersed in a pervasive magnetic field, radiation, and cosmic rays. The dynamical properties of stars far from the center of the Galaxy also suggest a gravitational interaction with invisible dark matter which is thought to constitute 90% of the total mass of the Galaxy. The structure of the Galaxy is mainly the consequence of gravitational interactions. Most of the visible part of the Galaxy lies within a thin disc with a radius of approximately 15 kpc, and a width of a few hundred parsecs. In terms of mass, the total mass within this disc is $\sim 10^{11} M_\odot$, including $\sim 5 \times 10^{10} M_\odot$ of stars, $\sim 5 \times 10^{10} M_\odot$ of dark matter, and $\sim 7 \times 10^9 M_\odot$ of interstellar gas [75]. In this picture, our solar system is located 8.5 kpc away from the Galactic center, orbiting around it at approximately 220 km/s.

The interstellar medium (ISM), referring to the medium which stands between the stars, is a turbulent mixture of gas and dust grains, coupled to the Galactic magnetic field. While roughly accounting for $\sim 10\%$ of the baryonic mass of the Galaxy, it plays a fundamental role in the Galactic ecosystem, being at the same time the cradle and the graveyard of stars. Indeed, stars are formed in its molecular clouds, and may die

spectacularly in supernovae, thus giving back their remains to the ISM as well as kinetic energy which (partly) drives interstellar turbulence. The chemical composition of the ISM is dominated by light elements: hydrogen, in its neutral, ionized and molecular forms, representing $\sim 73\%$ of the mass of the ISM, and helium, representing $\sim 26\%$. Heavier elements, called *metals*, only constitute $\sim 1\%$ of the mass of the ISM, but are still responsible for complex chemical processes, while remaining valuable tracers of the properties of the medium. The *metallicity* of the ISM, i.e. its abundance in metals, gives clues about the age of the Galaxy, since it is directly related to the number of generations of stars it has known.

We cannot give a typical temperature or typical density of the gas in the ISM, since both temperature and density may vary over several orders of magnitude. We usually distinguish phases in the ISM to account for this variety of conditions of the gas. According to [75], these include:

- Coronal gas, or hot ionized medium (HIM): gas that has been heated up to $T \gtrsim 10^{5.5}$ K by shocks from supernovae,
- HII gas⁸: gas that has been photoionized by ultraviolet (UV) photons from hot stars,
- Warm HI gas, or warm neutral medium (WNM): in majority atomic gas heated to $T \approx 10^{3.7}$ K, and with hydrogen densities $n_{\text{H}} \approx 0.6 \text{ cm}^{-3}$,
- Cool HI gas, or cold neutral medium (CNM): in majority atomic gas with typically $T \approx 10^2$ K and $n_{\text{H}} \approx 30 \text{ cm}^{-3}$,
- Diffuse molecular gas: similar properties to cool HI gas, but with sufficiently large densities and column densities so that H_2 molecules can form,
- Dense molecular gas: gravitationally bound gas, with $n_{\text{H}} \gtrsim 10^3 \text{ cm}^{-3}$, where star formation takes place.

We provide in Table I.2 the typical temperatures, densities, and volume filling factors for these various phases of the ISM.

Galactic magnetic fields (GMFs) are ubiquitous in the ISM (see e.g. [76, 77] and references therein). Historically, their existence was first suggested by Alfvén, then Fermi, to explain the acceleration and Galactic confinement of cosmic rays [78], before being indirectly revealed by the discovery of starlight polarization in 1949 [79, 80]. Since then, we found out that they are key for the understanding of various other astrophysical processes. As well as being involved in the observed alignment of dust grains, GMFs make relativistic electrons spiraling around them, which leads to the synchrotron emission. They are also strongly coupled to the interstellar gas, and known to influence star formation. In terms of structure, we usually decompose them as a sum of a large-scale component, which is coherent over Galactic scales (as for the large-scale GMF lines which seem to follow Galactic spiral arms), and a much more irregular small-scale component, which is strongly coupled to the ISM, thus inheriting its turbulent characteristics. Measurements of the strength of GMFs give typical values of a few μG (see e.g. [81]).

⁸In spectroscopic notations, HII refers to the ionized atomic hydrogen, while HI refers to the neutral atomic hydrogen.

Phases	T (K)	n_{H} (cm^{-3})	f_{V}
HIM	$\gtrsim 10^{5.5}$	0.004	0.5
HII gas	10^4	$0.2 - 10^4$	0.1
WNM	5000	0.6	0.4
CNM	100	30	0.01
Diffuse molecular gas	50	100	0.001
Dense molecular gas	10 - 50	$10^3 - 10^6$	10^{-4}

Table I.2 – Typical temperatures T , densities n_{H} , and volume filling factors f_{V} of the various phases of the ISM according to [75].

The ISM appears turbulent on a vast range of scales (see [82] for a review). Turbulence is characterized by the chaotic motions of a fluid, at scales where its kinetic energy dominates dissipation. A classical picture of turbulence is that of Richardson cascade [83], in which a turbulent flow is described by eddies of various sizes, with larger eddies tending to break up into smaller eddies down to the scale where dissipation occurs, called *dissipative scale*. Within this picture, the *inertial range* refers to the range of scales where these eddies may break up into smaller ones, that is where the kinetic energy cascades down to the dissipative scale. The Richardson cascade, and the subsequent theory of Kolmogorov (see [84] for a modern account of Kolmogorov theory), emphasize that turbulent flows are expected to be self-similar in the inertial range, so that statistics with power-law scale dependence are common in a turbulent regime (see Appendix A for a definition of self-similar random fields). Although Kolmogorov theory cannot directly apply to the magnetized and compressible ISM as it focuses on the (relatively) simpler case of hydro-dynamical incompressible turbulence, observations of the ISM are found to be self-similar over a wide range of scales, typically going from 10^{-1} to 10^3 pc [82]. The turbulent ISM also appears *intermittent*, with localized bursts of dissipation, which significantly complicates its statistical properties.

I.3.2 Interstellar dust grains

Interstellar dust grains are crucial constituents of the ISM. Their existence has first been noticed at the beginning of the XXth century looking for explanations for the observed interstellar extinction in the optical domain [85, 86]. Indeed, dust grains absorb approximately 1/3 of the visible and UV energy radiated by stars, and reemit this energy in the infrared and submillimeter domains through thermal radiation [87]. They are also involved in important chemical processes, notably catalyzing the formation of H₂ which is fundamental for the development of dense molecular clouds, and thus for star formation [88]. Finally, they can heat the gas by releasing energetic electrons pulled off by UV photons (photoelectric effect), as well as cooling it in dense regions thanks to thermal radiation in the infrared domain [75].

Dust grains are thought to be composed of various heavy elements, mainly C, O,

Mg, Si, S, Fe, forming sub-micronic solid particles down to nanometric sizes and polycyclic aromatic hydrocarbon molecules (PAH). Dust models usually distinguish two categories of grains: amorphous silicate grains, and carbonaceous grains which include PAH molecules (see e.g. [89]). Their typical sizes range from $0.01 \mu\text{m}$ to $0.2 \mu\text{m}$, with large grains dominating the mass budget and small grains dominating the total surface area of the grains. The shape of grains is unknown, but it is clear that some of them must have nonspherical shapes, usually modeled with ellipsoids or simpler spheroids.

The shape of grains is fundamental to explain the observed polarization properties of the ISM. The discovery of starlight polarization in 1949 by Hall and Hiltner [79, 80] was quickly attributed by Davis and Greenstein to the differential extinction of nonspherical dust grains which must be aligned somehow with the Galactic magnetic field [90, 91]. This differential extinction also implies a complementary differential thermal emission thanks to Kirchoff's law, and thus thermal radiation of dust grains in the infrared and submillimeter domains must also be polarized, the direction of polarization being orthogonal to that of starlight [87]. Therefore, observations of the polarization properties of the ISM trace the alignment of dust grains and the topology of the Galactic magnetic field.

Although many questions remain open, the physical processes explaining the alignment of dust grains with the local magnetic field seem quite clear (see [92] for a review). We observe that large non spherical grains (with a typical size $\gtrsim 0.1 \mu\text{m}$) have their short axis, that is the principal axis with largest moment of inertia, preferentially aligned with the local magnetic field. Indeed, on one hand, grains have an angular momentum that is expected to align with the short axis of the grain body so that they can minimize their rotational kinetic energy. This alignment is efficient provided the rotation is suprathermal, in order to overcome disalignment by collisions. On the other hand, grains are expected to have a magnetic moment due to their angular momentum (mainly because of the Barnett effect [93]), thus resulting in a precession of the angular momentum around the local magnetic field. Now, to account for the alignment of the angular momentum of grains with the local magnetic field, there is still an ongoing debate, but the most plausible explanation seems to be that of radiative torque (RAT) alignments theory [92]. In this theory, the RATs exerted onto grains by the absorption and scattering of photons can both spin-up and align grains with the magnetic field. This is the current paradigm used to model grain alignment [94].

I.3.3 The dust foreground: sky maps and main statistical properties

This thesis focuses on the statistical modeling of the dust foreground at the frequencies of observations of the CMB. Figure I.4 shows that, using *Planck* data, the dust foreground is the most prominent signal at 353 GHz for both polarization and intensity data. We show in Fig. I.6 the I , Q , and U GNILC maps at 353 GHz. The resolution of the maps have been reduced to $80'$ from the effective beam full-width at half-maximum (FWHM) of the instrument at this frequency $5'$ to have a good signal to noise ratio over the whole sky. The sky emission is obviously very inhomogeneous, being intense towards the Galactic plane, and much fainter at high latitudes. For the search of primordial B -modes, fields at high latitude, thus much less contaminated by the dust foreground, are prime targets.

From the observed Stokes maps I , Q , and U of the dust emission, one can derive several useful quantities among the gas column density N_{H} (derived from the dust optical

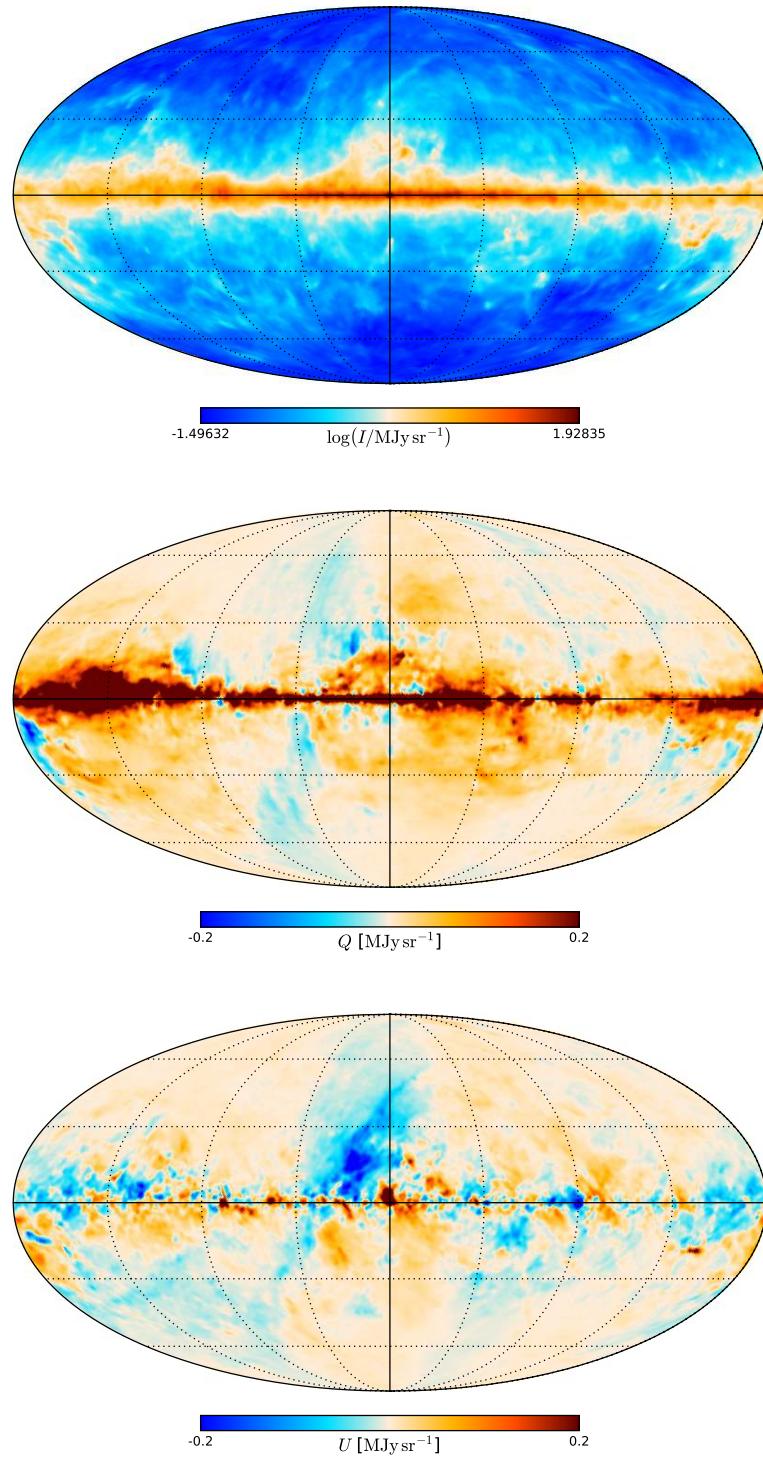


Fig. I.6 I , Q , and U GNILC maps of the dust foreground at 353 GHz derived from Planck data, in Galactic coordinates with the Galactic center in the middle. The resolution of the maps is uniform and equal to $80'$. Figure is taken from [58].

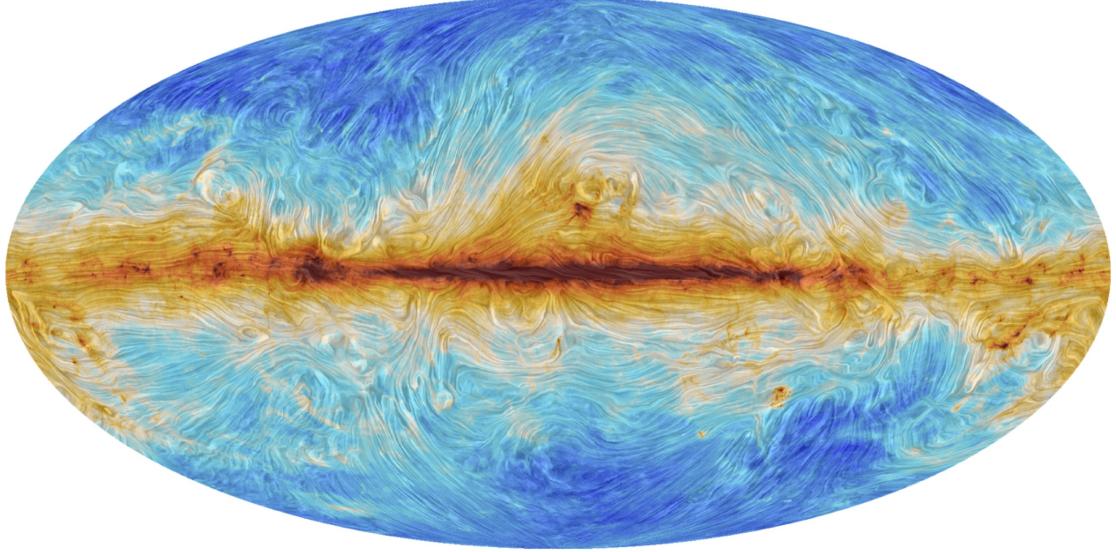


Fig. I.7 Superposition of total intensity at 353 GHz and drapery patterns showing the orientation of the Galactic magnetic field as inferred from the polarization angle at 353 GHz rotated by 90°. Figure is taken from [95].

depth τ_d ⁹, see [96]), the polarization fraction p , and the polarization angle ψ . These two latter quantities are defined as follows:

$$p = \frac{\sqrt{Q^2 + U^2}}{I}, \quad (\text{I.30})$$

$$\psi = \frac{1}{2}\text{atan}2(U, Q), \quad (\text{I.31})$$

where $\text{atan}2(b, a)$ is the function that returns the angle restricted to $(-\pi, \pi]$ between the positive x -axis and the ray to the point (a, b) in the Euclidean plane. Note that when noise becomes prominent, these quantities may suffer from a significant bias due to the noise. Strategies have been developed to tackle this problem, and we refer to [97, 98, 99] for definitions and comparisons of unbiased estimators for p and ψ . The polarization fraction p is a measure of the degree of polarization of the dust emission, and the polarization angle ψ traces the orientation of the projection of the Galactic magnetic field. Indeed, assuming an effective alignment of interstellar grains with the magnetic field, and approximate coherence of the direction of the magnetic field along the line of sight, the polarization angle ψ rotated by 90° indicates the direction of the magnetic field projected on the plane of the sky. We show in Fig. I.7, the sky map of the direction of the projected magnetic field as inferred from the polarization angle at 353 GHz. This illustrates the structure of the Galactic magnetic field, notably its correlation with structures in the intensity map.

⁹The optical depth is the logarithm of the ratio of incident to transmitted radiant power through a medium.

Finally, p and ψ obviously relate to Q and U in the following manner:

$$Q = p \times I \times \cos(2\psi), \quad (\text{I.32})$$

$$U = p \times I \times \sin(2\psi). \quad (\text{I.33})$$

We now give a brief overview of the main properties of the dust thermal emission foreground extracted from *Planck* data. These have been guiding the modeling of the dust emission. We discuss separately spectral and spatial properties of the emission.

One important spectral property is that the dust spectral energy distribution (SED) is in first approximation well modeled by a MBB with a mean temperature $T_d \sim 20$ K and a mean spectral index $\beta_d \sim 1.5$ [67, 57]. This was exploited by the Commander algorithm as seen in Sect. I.2.3 to perform its component separation, and we will discuss the physical interpretation of such a SED in more details in the next chapter. Note however that this approximation could not be sufficiently accurate for the search of primordial B -modes, and a lot of work has been done to go beyond this MBB model (see e.g. [100, 101]).

In terms of spatial properties, dust statistics are very rich, showing, in particular, complex filamentary patterns arising from the highly nonlinear ISM dynamics. In addition to being very inhomogeneous, statistics of the dust foreground are highly non-Gaussian¹⁰ both in intensity and polarization (see e.g. [102]). This non-Gaussianity is also naturally expected with nonlinear processes which strongly couple scales.

At 353 GHz, the polarization fraction p does not exceed ~ 20 %, tending to be maximal where the gas column density N_{H} is low, corresponding to so-called *diffuse* regions. More generally, the maximum polarization fraction tends to decrease when the gas column density N_{H} increases [96, 58]. It has also been noted that the dispersion of the polarization angle ψ , measured by the polarization angle dispersion function \mathcal{S} is anti-correlated to the polarization fraction p . These properties seem to be related to the turbulent structure of the magnetic field [103, 58].

Interestingly, the orientation of the magnetic field probed by the polarization angle ψ has been found to be correlated to the filamentary structures observed in intensity data [104, 105] and HI emission [106]. In the diffuse ISM, we observe that the filaments are preferentially aligned with the orientation of the magnetic field. Conversely, in dense molecular complexes, structures tend to be perpendicular to the orientation of the magnetic field. Observations of neutral hydrogen structures in the diffuse ISM have also given another perspective to this observed correlation of the magnetic field with the structures of the gas, showing a preferential alignment of linear structures with the structures of the magnetic field probed by starlight polarization and dust polarization [106, 107].

In the context of the search of primordial B -modes, Q and U maps are rather studied in terms of their E and B transforms (see Sect. I.2.1). Spatial properties of E and B -modes maps of the dust foreground at 353 GHz away from the Galactic plane have been studied in [108, 57]. Notably, it has been shown that the power spectra of E and B -modes are well fitted by power laws. These works have also exhibited a significant correlation between intensity and E -modes data, called *TE* correlation, as well as an asymmetry of power between E and B -modes, the power spectrum of E being approximately twice that of B . These signatures can be accounted for by the preferential alignment of filaments of

¹⁰We will say that data have non-Gaussian statistics if their statistics differ from those of Gaussian random fields (see Appendix A).

matter with the orientation of the magnetic field [109, 107]. Finally, in [57], a significant correlation between intensity and B -modes, called TB correlation, has also been reported. This was also related to the relative geometry of the magnetic field and the filamentary ISM in [110].

I.3.4 Simulating the dust emission

Observations of the dust emission need to be confronted to realistic numerical simulations to improve our understanding of the various processes involved in the ISM. However, as convenient as simulations may be, as we have depicted it in this section, the physics of the ISM remains very complex. Important approximations, or simplifications must be done at some point to simulate the dust emission. Usually, as dust grains are well mixed with the interstellar gas, simulations of the dust emission go through the resolution of the magnetohydrodynamical (MHD) equations describing the gas dynamics. The lack of a satisfying analytical theory of MHD turbulence in the multi-phase ISM demands direct numerical solving of these equations. This requires heavy computations, which can easily take days or weeks to run on computer clusters. Although the computation time of such simulations depends on many factors (target resolution, number of processors, physical processes involved, etc), to give a more precise order of idea, the MHD simulation I employed for the purposes of this thesis (which will be introduced in Chapter III) took almost 8 days while running on 520 processors. This heavy computational cost is one of the reasons why it is hopeless to solve these equations over the huge range of scales of interstellar turbulence and for parameters or a choice of physical processes that perfectly reflects the conditions of the ISM. Nevertheless, keeping these limits in mind, producing simulated Stokes maps I , Q , and U from a MHD simulation which can be compared to observations can be roughly divided into three steps:

1. Direct numerical solving of the MHD equations for a specific choice of the physical processes, parameters, and initial and boundary conditions.
2. Choice of a physical model of dust: nature of grains, abundances, shapes, sizes, alignment properties, etc.
3. Resolution of the radiative transfer leading to maps of the Stokes parameters (when this step is not coupled to the first one).

We briefly discuss these steps, and give examples of state-of-the-art strategies for each of them.

a) Solving MHD equations

The interstellar gas dynamics is usually described with a single fluid. In a simple case where we assume that the fluid has a perfect conductivity (ideal MHD), and we neglect any gravitational interaction as well as viscous stresses, the MHD equations take a (relatively)

simpler form (see e.g. [111]):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (\text{I.34})$$

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot \left(\rho \mathbf{v} \otimes \mathbf{v} + \left(P + \frac{B^2}{8\pi} \right) \mathbf{I} - \frac{\mathbf{B} \otimes \mathbf{B}}{4\pi} \right) = 0, \quad (\text{I.35})$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \cdot (\mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v}) = 0, \quad (\text{I.36})$$

where ρ , \mathbf{v} , \mathbf{B} , and P , are the density, velocity, magnetic, and pressure fields, respectively.¹¹ Obviously, we need to make sure that the magnetic field remains a solenoidal vector field, that is that $\nabla \cdot \mathbf{B} = 0$, which is a crucial constraint for any numerical solving. These equations need to be complemented by an energy equation, which is usually formulated as follows:

$$\frac{\partial E}{\partial t} + \nabla \cdot \left(\left(E + P + \frac{B^2}{8\pi} \right) \mathbf{v} - \frac{1}{4\pi} (\mathbf{v} \cdot \mathbf{B}) \mathbf{B} \right) = -\rho \mathcal{L}, \quad (\text{I.37})$$

where E is the total energy density (kinetic plus thermal plus magnetic), and \mathcal{L} is a net cooling function. This net cooling function measures the exchange of energy per unit time and unit mass of the gas with its environment due to the various heating and cooling processes in the ISM. Various solvers exist for these equations, such as RAMSES [112, 113], or Athena [114].

Obviously, one can add various processes corresponding to additional terms in these equations. State-of-the-art strategies, such as described in [115], may add for example gravitational interaction, including self-gravity or an external gravitational potential accounting for the influence of the stellar disk and the dark matter halo. This typically adds a term in both the momentum equation (I.35) and in the energy equation (I.37), and requires a Poisson equation connecting the gravitational potential to the density field. One can also take into account the influence of Galactic differential rotation and of the corresponding inertial forces when solving the equations in the rotating frame, also adding complementary terms in the momentum equation [115]. A detailed consideration of dissipative processes (viscous, resistive, ambipolar diffusion) as well as of the various chemical processes would also significantly complicate these equations (see e.g. [116, 117, 118]).

Turbulence also needs to be driven by an injection of kinetic energy at largest scales. In the local ISM, this injection of energy is mainly attributed to supernovae. One can therefore either account for the feedback of supernovae as described in [115], or model this injection of energy by means of a turbulent stochastic forcing term (see e.g. [119]).

b) Choice of a physical model of dust

The choice of the dust model is also very important. Simplest models will assume uniform properties of alignment of the grains with the magnetic field, while more convoluted descriptions will introduce several populations of grains (usually carbonaceous and silicates), with a given distribution of sizes (usually power laws) and shapes (usually spherical or spheroidal) per population, from which absorption, scattering and polarization properties are derived (see e.g. [94]).

¹¹Here \otimes denotes the tensorial product, and \mathbf{I} is the identity tensor of order 2. We adopt Gaussian units for the definition of the electromagnetic field.

c) Resolution of the radiative transfer

The resolution of the radiative transfer may be a very complex and computationally demanding step. Various approaches are possible, such as that of POLARIS [120], which is an example of state-of-the-art three-dimensional Monte-Carlo radiative transfer code. In this code, the transfer is solved for photon packages drawn randomly. The emission and absorption properties of the medium are progressively updated along the pathways of the photons, which may involve several scatterings or absorptions.

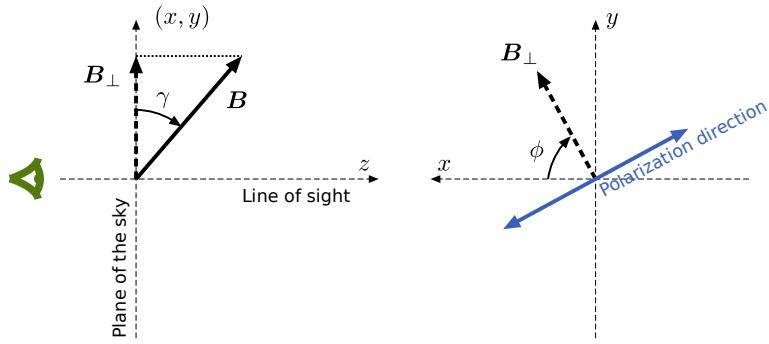


Fig. I.8 *Definition of the angles γ and ϕ related to the orientation of the local magnetic field with respect to the plane of the sky when the z -axis corresponds to the line of sight. The line of sight points away from the observer and ϕ is counted positively clockwise from the x direction in the HEALPix convention. Figure is adapted from [103].*

Another much more straightforward strategy is to resort to simplifying assumptions on the properties of the medium. For a given frequency ν and line of sight, the Stokes parameters I_ν , Q_ν , and U_ν can be written (see e.g. [103]) as follows:

$$I_\nu = \int S_\nu e^{-\tau_\nu} \left[1 - p_0 \left(\cos^2 \gamma - \frac{2}{3} \right) \right] d\tau_\nu, \quad (\text{I.38})$$

$$Q_\nu = \int p_0 S_\nu e^{-\tau_\nu} \cos(2\phi) \cos^2 \gamma d\tau_\nu, \quad (\text{I.39})$$

$$U_\nu = \int p_0 S_\nu e^{-\tau_\nu} \sin(2\phi) \cos^2 \gamma d\tau_\nu, \quad (\text{I.40})$$

where S_ν is the source function associated to the grains, τ_ν is the optical depth, p_0 is an intrinsic polarization fraction parameter, γ is the angle that the local magnetic field makes with the plane of the sky, and ϕ is the angle that the projection of the local magnetic field on the plane of the sky makes with the x direction when z is the line of sight in the HEALPix¹² [121] convention (see Fig. I.8). With simple prior assumptions for S_ν and τ_ν , one may easily compute these integrals, thus avoiding computationally demanding methods. This strategy will be employed in Chapter III.

While simulations of the dust emission are essential to gain understanding of the physics of the ISM, comparing them to actual observations of the ISM is not necessarily

¹²<https://healpix.jpl.nasa.gov>

straightforward. For example, when should we say that a simulation reproduces accurately the properties of the observed ISM? Obviously, we cannot count on reproducing a given observation pixel per pixel, such comparisons often necessitate statistical comparisons (see [122] for a recent review of statistical diagnostics employed for studying interstellar turbulence). The problem of defining a relevant metric to compare simulations and observations is still very open. Most of the time this metric is implicitly defined by comparing simple statistics, such as power spectra statistics or distributions of pixel values as we will see in the next chapter. A relevant metric first requires to identify the salient statistical properties of the dust emission. Some of them have already been presented in Sect. I.3.3, but it seems that this picture remains incomplete. We will see that the statistical modeling of the dust emission addresses related problems. In this thesis, we will delve into what might be the important statistical properties of the dust emission, emphasizing on the importance of statistics quantifying the couplings between scales. In the next chapter, we will start by detailing our motivations, before establishing a state-of-the-art for the modeling of the dust emission.

Chapter II

Modeling the dust emission: motivations and state-of-the-art

This thesis aims at building a comprehensive statistical model of the thermal emission of interstellar dust. We discuss the motivations for building such a model, then we give an overview of the various approaches that have already been employed in the recent literature. We attempt to classify these models in broad categories. We first differentiate phenomenological models, that leverage a physical understanding of the properties of the ISM, from agnostic models, that make no use of this physical understanding. We will also distinguish deterministic models, designed to retrieve the true emission of the sky, from statistical models, focusing on its statistical properties. We will see that all of these existing methods are parametric, with various degrees of interpretability. Finally, we will situate our approach in this context, which is designed to be non-parametric, statistical, and in some sense halfway between phenomenological and agnostic approaches.

II.1 Motivations

The motivation for developing a model of the dust emission is twofold. First, for cosmology, it is clear that the search for primordial B -modes necessitates a significant improvement of component separation methods to better disentangle dust from CMB [2]. Component separation methods might directly (e.g. Commander) or indirectly rely on models of the dust emission, therefore the better the model, the cleaner the component separation. Second, from the perspective of Galactic astrophysics, a comprehensive model of the dust emission would also mark a step forward in our understanding of the ISM. Such a model would crystallize our knowledge of the properties of the dust emission in a consistent way, thus defining a valuable reference for analysis, comparison, or prediction of observations.

A modeling of the dust emission would have to exploit observational data that can be either mono-frequency or multi-frequency, over the whole sky or just focusing on a limited field. It would have to take into account the fact that the emission is usually significantly mixed with other components, may be noisy, or affected by systematics. The resolution of the observations also depends on the instrument, and on the frequency channel.

We must distinguish two kinds of models with complementary goals, deterministic

ones and statistical ones. Deterministic models aim at retrieving the exact sky map of the emission at some observed frequency, free from the potential contamination of other components, noise, or systematics. This deterministic modeling is obviously essential to analyze specific regions of the sky in a deterministic way. It can also help for data cleaning, or calibration. It gives a solid ground for deterministic predictions, such as the prediction of the emission at an unobserved frequency. On the other hand, statistical models do not aim at retrieving the exact emission, but rather focus on finding a comprehensive and self-consistent characterization of its statistical properties. In other words, statistical approaches picture the observed sky as a realization of a random field, and aim at identifying the law of this underlying random field. Note that such statistical models are common in cosmology. As discussed in Sect. I.1, the CMB is very well modeled by a Gaussian random field, and its power spectrum allows to constrain the parameters of the Λ CDM model. A statistical model is usually generative, meaning new realizations of the sky can be drawn based on this model. This is of strong interest to simulate the dust foreground, thus giving an alternative to computationally expensive simulations (e.g. MHD simulations, see Sect. I.3.4). Statistical models can also be predictive. For example, they may predict the statistical properties of the emission at an unobserved frequency, allow for data inpainting (that is the replacement of lost or corrupted data), or they may predict unresolved angular scales based on the statistical properties of larger ones (what is called *super-resolution*).

II.1.1 Why a statistical approach?

We can question the relevance of a statistical approach. Indeed, if we had a perfect deterministic knowledge of the dust emission, at every frequency and angular scales, the component separation problem would be obviously solved. We would just have to subtract the known dust emission from the sky emission. Therefore, a deterministic approach would be a priori more relevant to component separation problems. However, the fact is we will never be able to perfectly measure the dust emission, either because the sky emission is always a mixture of components at some point, or just because our instruments are not perfect, and they will always introduce systematics or a certain amount of noise which will naturally introduce randomness in the measurements. Ignoring this randomness would be missing some potentially valuable information. For instance, this can go from elementary estimations of measurements uncertainties, to correlations in the data that could improve the accuracy of the measurements when fitting a model to the data.

We also stress the importance of statistical simulations for the validation of data analysis pipelines. In particular, for CMB experiments, this was and is still now one of the main applications of the *Planck* sky model [123] (PSM) and Python sky model [124] (PySM). Such simulations are essential to assess the capabilities of upcoming CMB experiments and their component separation methods in finding primordial *B*-modes.

From a physics point of view, interstellar turbulence gives us no choice but to adopt a statistical approach. Indeed, turbulent motions are described statistically, and even if time scales of interstellar motions are way too large for us to observe the time evolution of the ISM, an ergodic assumption is usually relevant to a certain extent for estimating statistics that characterize this turbulence. Also, for CMB observations the target of cosmologists is the statistical information contained in the CMB, meaning its angular

power spectrum, or potential signs of non-Gaussianity [59]. In this cosmological context, the estimation of the power spectrum (or complementary non-Gaussian statistics) of the CMB demand statistical models for the foregrounds.

Nevertheless, we will see that the distinction between deterministic and statistical approaches becomes blurry at some point. A relevant statistical model of the dust emission will necessarily need to incorporate a form of determinism, if only to describe the largest angular scales on the sky from which it is pointless to make a statistical model because of cosmic variance.

II.1.2 What is a good model?

The definition of a "good" statistical model is very arbitrary and depends on the applications. For B -mode detection, a good model is typically a model allowing to estimate relevant uncertainties or bias of a component separation method. Or, when the model is directly included in the component separation method (e.g. as for Commander), a good model will be a model allowing a sufficiently accurate removal of the dust foreground to hopefully claim a detection. For Galactic astrophysics, a good model is roughly a model that tells us something about the physics of the ISM. Either way, we try here to draw common characteristics that will guide the approach of this thesis.

A good statistical model needs to be comprehensive, meaning that it should capture most of the salient statistical properties of the observational data. This is of the utmost importance for the simulation of the dust emission. To give an example, for B -mode detection, a statistical model of the spatial properties of the polarization signal will be all the more convincing if it correctly reproduces the observed power spectra, the TE correlation, and the E - B asymmetry (see Sect. I.3.3). In any case, the comprehensiveness of a model will be defined relatively to a choice of *summary statistics* (on this example power spectra measurements), usually inspired by the literature. Until recently, the main target of the statistical models of the dust emission has been to reproduce the observational power spectra of the emission, making use of homogeneous Gaussian random fields (GRFs). Homogeneous GRFs are entirely characterized by their mean and their power spectrum (see Appendix A), for this reason they are very natural choices for this kind of modeling. However, we know that the dust emission has highly non-Gaussian statistics, due to the nonlinearity of interstellar physics, meaning that a characterization in terms of power spectrum (and mean) only is not satisfactory. A comprehensive model thus needs to go beyond Gaussian models, or in other terms, in order to assess the comprehensiveness of our model, our choice of summary statistics will need to include a wider class of statistics than mere power spectra measurements. From a more physical point of view, this non-Gaussianity relates to the couplings of physical fields across scales, therefore a comprehensive model will need to capture and reproduce these interactions between scales.

To improve our understanding of the properties of the ISM, a good statistical model should be physically interpretable. Ideally, its definition should be guided by physics. This way, reproducing the statistical properties of observations from a physically motivated model would already be satisfactory. Moreover, the parameters of such models (when they exist) have often a straightforward physical interpretation, which facilitates a physical reasoning and a characterization of observational data. The drawback of this approach

is that we project the data into a pre-conceived framework, which might preclude our ability to extend our knowledge. Let us also emphasize that interpretability usually goes hand in hand with simplicity, and a too intricate model, for example a model involving a large number of parameters, will necessarily be less interpretable than a model with a few parameters only.

Finally, let us mention that in the prospect of simulating the dust emission, ideally, it should be computationally efficient to generate statistical realizations from the model.

II.1.3 Standard statistics

We review here standard statistics that are usually employed to study the statistical properties of the ISM, or that are often picked as summary statistics to assess the comprehensiveness of a statistical model. Since our confidence in a model usually depends on its ability to reproduce a given set of summary statistics, our model should include most of the information captured by these statistics. However let us keep in mind that this is not necessarily the only criterion that determines the quality of the model as discussed in the previous subsection.

Let us consider a random field X , from which the observed data x is a realization. We assume that X is defined over a two-dimensional domain, each dimension corresponding to a spatial dimension. This domain can be a sphere in the case of sky observations, or a rectangular domain when making a flat-sky approximation with observational data, or using simulated data that is naturally defined on a plane. We will also assume X to be homogenous so that statistics are invariant under global translations. This assumption is open to serious criticism as it is obvious that the sky emission of interstellar dust is not statistically homogeneous. However, we will assume that at least locally, this assumption makes sense. To simplify, we finally assume X to be a real-valued random field, although we will deal with both real and complex-valued random fields in this thesis.

A natural way to study a random field is to estimate its so-called *n-point statistics* (see Appendix A for additional details). These include n^{th} -order probability distribution functions (PDFs) and n^{th} -order moments. In practice, the estimation of n^{th} -order PDFs for $n \geq 2$ will be numerically intractable, and we will refer more directly to the first-order PDF of a random field X as *the PDF of X* , or simply $p(X)$. The PDF of X does not depend on the position \mathbf{r} since we assumed X to be statistically homogeneous. On the other hand, n^{th} -order moments, which are projections of the n^{th} -order PDFs, will be relatively simpler to estimate provided we keep n relatively small. There are several ways to define these statistics, here we choose definitions that coincide with n -point correlation functions. Therefore the n^{th} -order moment of X , called $C_n(\mathbf{r}_1, \dots, \mathbf{r}_n)$, is defined as:

$$C_n(\mathbf{r}_1, \dots, \mathbf{r}_n) = \mathbb{E}[X(\mathbf{r}_1) \dots X(\mathbf{r}_n)]. \quad (\text{II.1})$$

Statistical homogeneity helps to get rid of one variable, so that C_n can be easily identified to \tilde{C}_n defined as:

$$\tilde{C}_n(\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_{n-1}) = C_n(\mathbf{r}, \mathbf{r} + \boldsymbol{\tau}_1, \dots, \mathbf{r} + \boldsymbol{\tau}_{n-1}), \quad (\text{II.2})$$

for any \mathbf{r} . The first-order moment corresponds to the mean of X , called μ_X , while the second-order moment corresponds to the autocorrelation function of X , called $R_X(\boldsymbol{\tau})$.

Second-order and third-order moments are often studied from a spectral perspective, that is with the power spectrum $P(\mathbf{k})$ and the bispectrum $B(\mathbf{k}_1, \mathbf{k}_2)$, respectively. These are defined as:

$$P(\mathbf{k}) = \mathcal{F}[\tilde{C}_2](\mathbf{k}), \quad (\text{II.3})$$

$$B(\mathbf{k}_1, \mathbf{k}_2) = \mathcal{F}[\tilde{C}_3](\mathbf{k}_1, \mathbf{k}_2), \quad (\text{II.4})$$

where \mathcal{F} denotes both the bidimensional and quadridimensional Fourier transform, respectively (see Appendix B). When X is defined on a discrete grid of size $M \times N$ and assuming periodic boundary conditions, we can show that the power spectrum and the bispectrum relate to moments derived from the Fourier transform of X , denoted $\hat{X}(\mathbf{k}) = \mathcal{F}[X](\mathbf{k})$, as follows:

$$P(\mathbf{k}) = \frac{1}{MN} \mathbb{E} [|\hat{X}(\mathbf{k})|^2], \quad (\text{II.5})$$

$$B(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{MN} \mathbb{E} [\hat{X}(\mathbf{k}_1) \hat{X}(\mathbf{k}_2) \hat{X}(\mathbf{k}_1 + \mathbf{k}_2)^*]. \quad (\text{II.6})$$

We refer to Appendix A for additional mathematical details on the power spectrum and bispectrum. For the ISM, the power spectrum is one of the most immediate statistics that can be derived for any statistical analysis (see e.g. [125]). On the contrary, the bispectrum, while being more difficult to represent and interpret, also suffers from a higher statistical variance compared to the power spectrum, which thus requires larger data sets for accurate estimates. Beyond these limitations, the bispectrum is a useful statistics to probe non-Gaussianity in the data. It naturally quantifies interactions between two given scales. Bispectra analyses were first introduced in the ISM community in [126] to study non-Gaussianity in MHD simulations of interstellar turbulence. Further analyses in [127] have shown the interest of the non-angular averaged three-point correlation function to better discriminate different regimes of interstellar turbulence (see Fig. II.1 for an illustration).

Standard statistics for the analysis of the ISM also include the PDFs of the increments of X . The increments of X for a given lag \mathbf{l} are the fields $\delta X_{\mathbf{l}}$ defined by $\delta X_{\mathbf{l}}(\mathbf{r}) = X(\mathbf{r} + \mathbf{l}) - X(\mathbf{r})$. The PDFs of the increments for a lag \mathbf{l} are simply the distributions of the field differences between values at points separated by the lag \mathbf{l} . For the ISM, these are for instance computed on velocity maps for their tails that characterize the intermittent dissipation of interstellar turbulence [128, 129].

We also mention Minkowski functionals which are common statistics to characterize the morphological aspects of smooth random fields. In cosmology, these have been used in various contexts, as the investigation of potential non-Gaussianity and anisotropy in the CMB (see e.g. [130, 33]), the characterization of the large-scale structure (e.g. [131]) or that of weak lensing data (e.g. [132]). In the ISM community, these are however much less popular statistics, although it has already been applied in the context of dust modeling [73, 133]. In two dimensions, there are three Minkowski functionals V_0 , V_1 , and V_2 , which are defined as follows. Choosing a given threshold α , we define for a particular realization x of X , the excursion set $\Gamma_{\alpha} = \{\mathbf{r} \mid x(\mathbf{r}) \geq \alpha \times \sigma\}$, where σ is the standard deviation of X . This set simply corresponds to the region where the field is greater than a given threshold expressed in units of the standard deviation of the field. The Minkowski

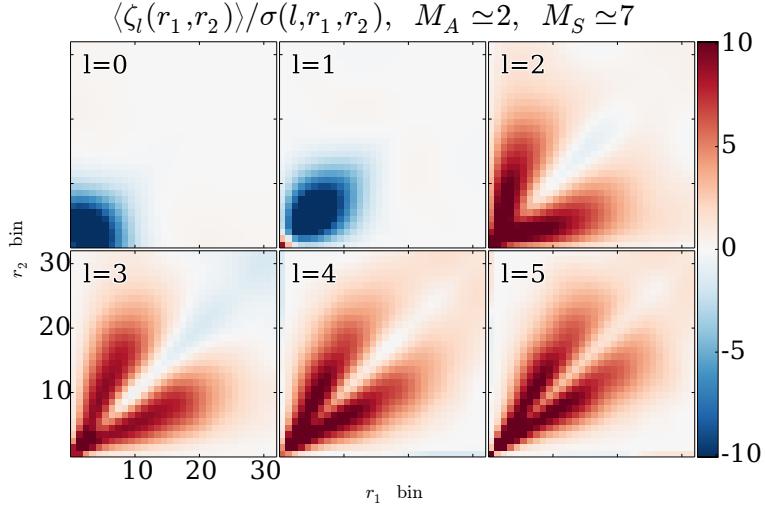


Fig. II.1 Example of representation of a three-point correlation function $\tilde{C}_3(\mathbf{r}_1, \mathbf{r}_2)$ as computed for a MHD simulation in [127] (signal-to-noise coefficients). Statistical isotropy is assumed, and the three-point correlation function $\tilde{C}_3(\mathbf{r}_1, \mathbf{r}_2)$ is expanded on Legendre polynomials P_l , thus defining $\zeta_l(r_1, r_2)$ coefficients. Figure is taken from [127].

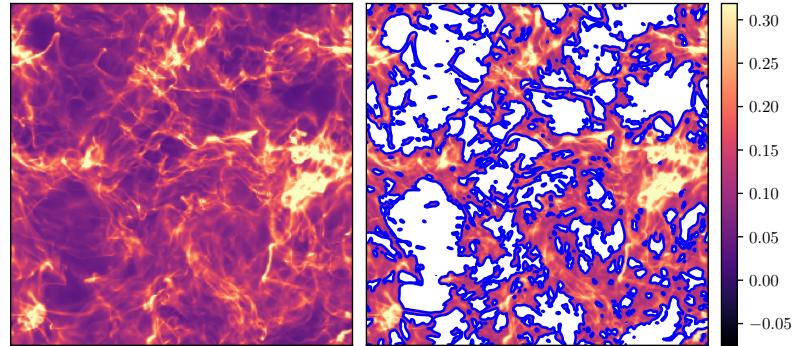


Fig. II.2 Visual interpretation of Minkowski functionals on simulated intensity data. The right panel shows the restriction of an example map, shown in left, to the excursion set Γ_α for $\alpha = 1.5$. The contours of the excursion set are shown in blue. Here, $V_0(\alpha)$, $V_1(\alpha)$, and $V_2(\alpha)$ are proportional to the uncovered area, length of the contours, integrated curvature along the contours, respectively. The example map is a dust emission map computed from a MHD simulation that will be introduced in Chapter III.

functionals associated with x are computed as:

$$V_0(\alpha) = \frac{1}{A} \int_{\Gamma_\alpha} da, \quad (\text{II.7})$$

$$V_1(\alpha) = \frac{1}{4A} \int_{\partial\Gamma_\alpha} dl, \quad (\text{II.8})$$

$$V_2(\alpha) = \frac{1}{2\pi A} \int_{\partial\Gamma_\alpha} \kappa dl, \quad (\text{II.9})$$

where A is the total area of the field, $\partial\Gamma_\alpha$ is the boundary of Γ_α , da and dl are the surface and line elements associated with Γ_α and $\partial\Gamma_\alpha$, respectively, and κ is the curvature of $\partial\Gamma_\alpha$ (see e.g. [132] for more "operational" definitions). The functionals V_0 , V_1 and V_2 are called the area, perimeter, and genus (i.e. number of "holes") characteristics, respectively (see Fig. II.2 for an illustration). For GRFs, or mildly non-Gaussian random fields, analytical expressions of these statistics have been derived, thus giving a way to quantify morphological non-Gaussianity [132].

Let us finally briefly cite statistics derived from topological data analysis. Topological properties of the data, i.e. properties that remain invariant under continuous deformations, can add a complementary insight to the previous statistics. In [134], the authors make use of Betti numbers and persistence diagrams to compare simulations and observations of the turbulent ISM. For cosmology, [135] discusses the interest of such topological statistics for characterizing cosmological GRFs. One spectacular application is the systematic identification and characterization of the skeleton of the cosmic web [136, 137]. We will not make use of these statistics in this thesis, however they are mentioned here mainly to recall that there is an important variety of approaches to describe the properties of images. This variety of approaches shows that the assessment of the quality of a statistical model will therefore be very dependent on the choice of statistics we use as a reference. We emphasize that this choice must be done in light of the subsequent applications of the model.

II.2 Phenomenological models

Physics of the ISM and its interstellar grains is rich and complex, involving a myriad of phenomena, which are nonlinear and coupled to each other. Added to that, the ISM exchanges matter and energy with stars and the extragalactic medium, which makes it a fundamentally open system. Taking into account all this complexity to build a comprehensive model of the dust emission valid at any frequency or resolution would be illusory. We have seen in the first chapter that even MHD simulations of the ISM cannot reasonably tackle the full complexity of the medium. Similarly, models of the dust emission will require important approximations. Most of these models will resort to phenomenological models to approximate this complexity, as far as possible. We discuss such phenomenological models here, distinguishing those that are directly fitted to data in a deterministic way, defining deterministic models, from those introducing a form of randomness, defining statistical models.

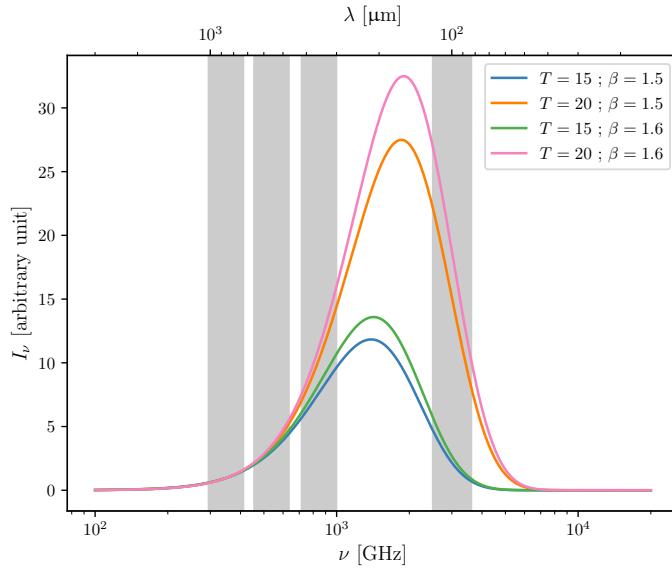


Fig. II.3 Frequency dependence of a MBB emission for a few typical values of β and T . The grey bands show the 353, 545, and 857 GHz bandpasses of the Planck satellite as well as the 100 μm bandpass of IRAS.

II.2.1 Deterministic approaches

Deterministic models often rely on a model of the spectral dependence of the emission for each pixel of the sky. They typically involve a *template*, which corresponds to the map of the sky emission at a given frequency, as well as sky maps of the parameters of the spectral model. The template and the parameters are determined by fitting the observational data. The fitted parameters give insight into the properties of the dust foreground, and they allow to predict what would be the observed sky at unobserved frequencies (provided these remain in the frequency range of the spectral model).

a) Modified blackbody radiation-based models

The most important class of spectral models that are fitted to observational data derive from the modified blackbody radiation (MBB) model from the far-infrared domain to the millimeter domain. Over this range of frequencies, the Galactic emission is dominated by the thermal radiation of large grains. In the optically thin limit, the intensity I_ν at a frequency ν reads:

$$I_\nu = \tau_\nu B_\nu(T), \quad (\text{II.10})$$

where τ_ν is the dust optical depth, and $B_\nu(T)$ is the Planck function describing a blackbody radiation at temperature T . The optical depth τ_ν can be written as $\tau_\nu = \kappa_\nu M_{\text{dust}}$, where κ_ν is the dust emission cross section per unit of mass, and M_{dust} is the dust mass surface density [138]. This parameter κ_ν is usually modeled as a power law $\kappa_\nu = \kappa_0(\nu/\nu_0)^\beta$, introducing the spectral index β and κ_0 the dust emission cross section per unit dust mass

at a reference frequency ν_0 . This modeling leads to the MBB emission:

$$I_\nu = \kappa_0 \left(\frac{\nu}{\nu_0} \right)^\beta M_{\text{dust}} B_\nu(T). \quad (\text{II.11})$$

We show in Fig. II.3 a plot of the frequency dependence of I_ν , choosing a few typical values for the dust emission for β and T . We clearly see from this plot that the maximum of the emission is typically at a few THz, and that towards the microwave domain and beyond the high-frequency limit of the far-infrared domain, the emission appears negligible in comparison (although, in the microwave domain, dust emission remains a significant foreground to the CMB down to ~ 70 GHz, see Sect. I.2.3). This simple model of the spectral dependence of the dust emission has been the cornerstone of various approaches.

In [138], the authors fitted this MBB model for each pixel of the intensity sky using *Planck* data at 353, 545, and 857 GHz and IRAS 100 μm data. They have found that this model reproduces well the data, and they have extracted T and β sky maps which characterize the spectral energy distribution (SED) of the dust foreground. These are shown in Fig. II.4 added to the corresponding map of the optical depth at 353 GHz τ_{353} . Notably, the means and standard deviations on the whole sky for T and β are (19.7 ± 1.4) K and 1.62 ± 0.10 , respectively. The Commander algorithm [61, 29] has performed a component separation also relying on a MBB modeling of the dust foreground using *Planck* and WMAP data, for both intensity and polarization data. For intensity, the correlation of dust model maps with the distribution of galaxies emphasizes contamination by the CIB [139]. The GNILC component separation method was designed to mitigate the CIB contamination [67]. GNILC has thus derived more accurate maps of the intensity dust foreground, and extracted improved estimates of the spectral parameters T and β . On the whole sky, updated means and standard deviations of these parameters have been found to be $T = (19.4 \pm 1.3)$ K and $\beta = 1.6 \pm 0.1$.

Two-component MBB models have also been used to describe the dust emission (see [140]). In [141], the authors have shown that modeling the dust emission in intensity with the sum of two MBB emissions rather than using a single MBB component better fits observational data on a 100-3000 GHz frequency range. The physical interpretation of this result is unclear, but might be interpreted as the sign of the existence of two distinct dust grain species. Note, however, that we expect a non-uniform temperature of grains along the line of sight, whose distribution would obviously be better captured by a model with more parameters.

The dust polarization SED measured by *Planck* for the diffuse ISM is also well fitted by a MBB from 353 to 45 GHz [57]. However, departures from the MBB model and its simple extensions have been anticipated at microwave frequencies, although they were not detected at the *Planck* sensitivity (see e.g. [100]). This could be an additional complication for components separation methods in the context of primordial B -mode detection. Strategies are currently in development to improve these parametric models of the spectral dependence of the dust emission on sky maps [142] and in the analysis of dust power spectra [101].

b) Physical models of dust

Another approach for the modeling of the dust SED is to make use of a physical model of interstellar dust instead of a phenomenological law of dust emission. In [143], the authors

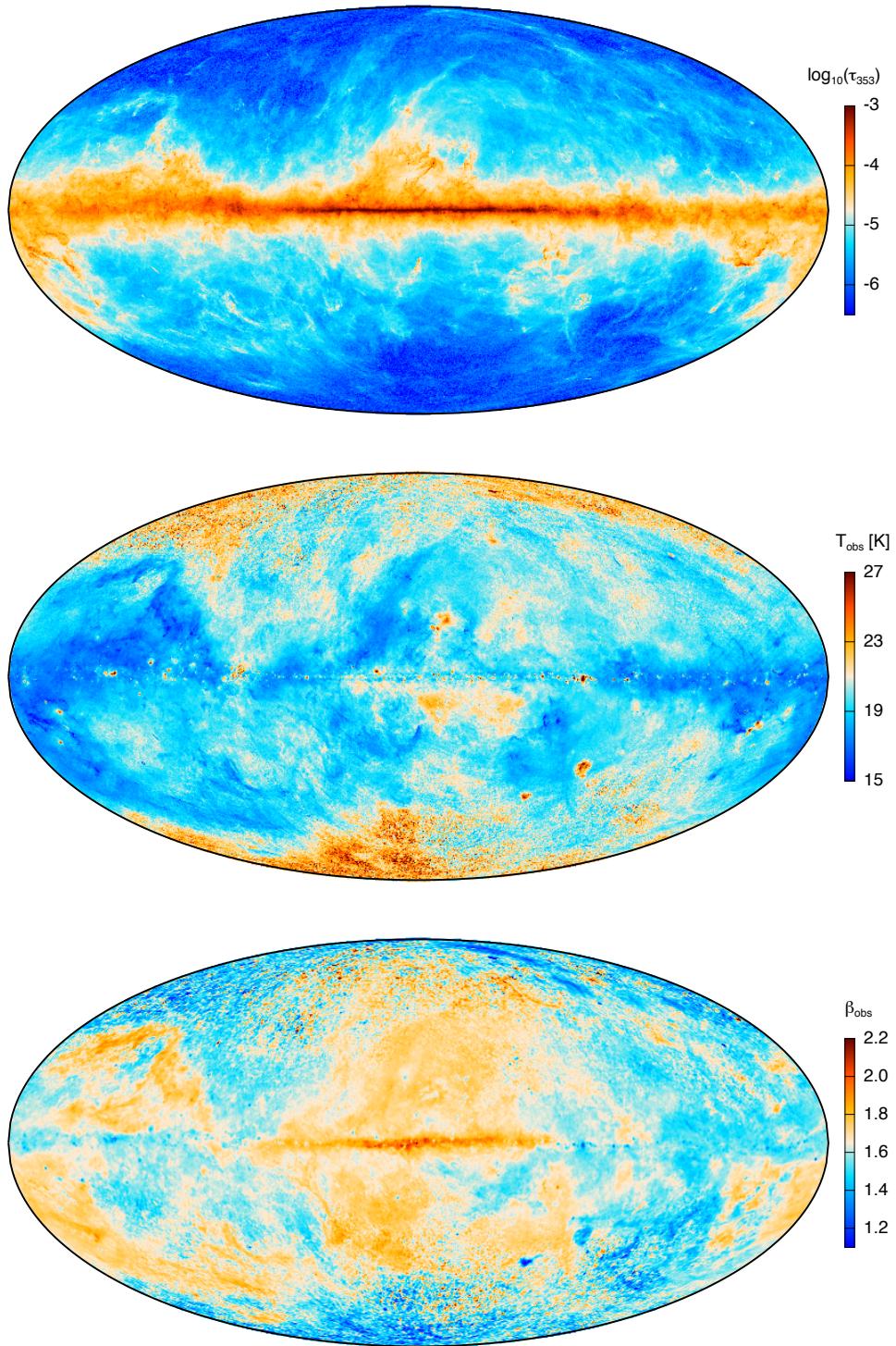


Fig. II.4 All-sky maps of the optical depth at 353 GHz τ_{353} , temperature T , and spectral index β as derived from a MBB fit of Planck 353, 545, and 857 GHz and IRAS 100 μm data. Figure is taken from [138].

made use of the dust model introduced in [144] to describe the emission observed in intensity by *Planck*, IRAS, and WISE infrared observations. This approach is based on the modeling of the dust grains as a mixture of amorphous silicate grains and carbonaceous grains that are heated by a distribution of starlight intensities. This model introduces parameters that characterize the emission, including the dust mass surface density and the intensity of the starlight radiation that is heating grains. This kind of approach, while depending on a fair amount of assumptions and parameters, gives a valuable complementary insight to constrain the properties of interstellar grains. Finally, let us note that, in return, these approaches allow to better constrain the existing physical models of dust (see e.g. [145]).

II.2.2 Statistical approaches

An important family of phenomenological statistical models has been motivated by the search for primordial B -modes. These models have been mainly focusing on the modeling of the Galactic magnetic field (GMF), whose structure is fundamental to explain the statistical properties of Q and U maps. We review the main strategies that have been introduced to tackle this problem.

In Sect. I.3.4, we introduced the expressions of the Stokes parameters I , Q , and U describing the polarized thermal emission of interstellar dust (see Eqs. (I.38)-(I.40)). The construction of simulated maps deriving from a model usually relies on these formulae. A few assumptions are often made to simplify these integrals. First, the source function of the dust may be assumed to be that of a blackbody radiation, that is $S_\nu = B_\nu(T_d)$, where T_d is the dust temperature. Second, the polarization fraction parameter p_0 , which is related to cross section parameters of the grains, and the degree of alignment of the grains with the magnetic field (see [103] for more details), may be assumed uniform (see [94] for a discussion). A typical value for p_0 is that used in [103], i.e. $p_0 = 0.2$. Third, the dust infinitesimal optical depth is written $d\tau_\nu = \sigma_\nu n_H dz$, where σ_ν is the dust cross section per H at frequency ν , assumed to be uniform, n_H is the hydrogen density, and z is the coordinate along the line of sight. Finally, we often assume the medium to be optically thin to the radiation, meaning that $\tau_\nu \ll 1$, which gives $e^{-\tau_\nu} \approx 1$. Note that this assumption is always valid in the diffuse ISM. Within these assumptions, Eqs. (I.38)-(I.40) read:

$$I_\nu = S_\nu \sigma_\nu \left[N_H - p_0 \int n_H \left(\cos^2 \gamma - \frac{2}{3} \right) dz \right], \quad (\text{II.12})$$

$$Q_\nu = p_0 \sigma_\nu S_\nu \int n_H \cos(2\phi) \cos^2 \gamma dz, \quad (\text{II.13})$$

$$U_\nu = p_0 \sigma_\nu S_\nu \int n_H \sin(2\phi) \cos^2 \gamma dz, \quad (\text{II.14})$$

where we have introduced the hydrogen column density $N_H = \int n_H dz$. From these formulae, Stokes parameters I , Q , and U appear to be proportional to an integration along the line of sight of quantities depending on the same three variables: the hydrogen density n_H , and the angles γ and ϕ which describe the orientation of the GMF (see Fig. I.8). The statistical modeling of the polarized emission of interstellar dust reduces to choosing a model for the hydrogen density field and the GMF orientation.

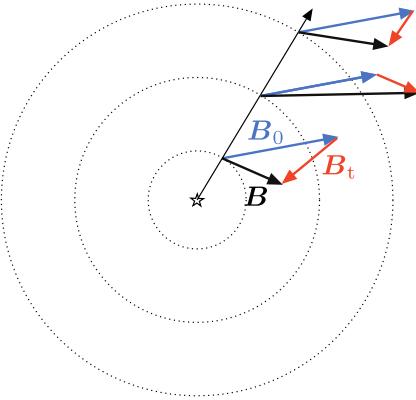


Fig. II.5 Schematic view of the GMF as modeled in [146], showing $N = 3$ layers. The observer is at the center. Figure is adapted from [147].

A series of papers [146, 148, 149, 150] has focused on the modeling of the GMF in the solar neighborhood using *Planck* data at 353 GHz at intermediate and high Galactic latitudes, aiming in particular at reproducing the statistical properties of the polarization fraction p and the polarization angle ψ . These models basically consist in describing the GMF as a piecewise constant field in a space divided into N concentric spherical shells, the observer being at the center. This leads to the following simplified expressions for the Stokes parameters:

$$I_\nu = S_\nu \sigma_\nu \sum_{i=1}^N N_{\mathrm{H},i} \left[1 - p_0 \left(\cos^2 \gamma_i - \frac{2}{3} \right) \right], \quad (\text{II.15})$$

$$Q_\nu = p_0 \sigma_\nu S_\nu \sum_{i=1}^N N_{\mathrm{H},i} \cos(2\phi_i) \cos^2 \gamma_i, \quad (\text{II.16})$$

$$U_\nu = p_0 \sigma_\nu S_\nu \sum_{i=1}^N N_{\mathrm{H},i} \sin(2\phi_i) \cos^2 \gamma_i, \quad (\text{II.17})$$

where $N_{\mathrm{H},i}$, γ_i , and ϕ_i are the column density, and angles describing the orientation of the magnetic field in the shell i , respectively. The angles γ_i and ϕ_i are drawn from a random modeling of the magnetic field \mathbf{B}_i . The magnetic field \mathbf{B}_i is decomposed as $\mathbf{B}_i = \mathbf{B}_0 + \mathbf{B}_{t,i}$, where \mathbf{B}_0 is a mean component, which is assumed to be constant and independent from the shell i , and a turbulent component $\mathbf{B}_{t,i}$, which is a homogeneous GRF with a power-law power spectrum (the exponent of the power-law is an explicit parameter of the model). The turbulent components are assumed to be independent across the shells. We show in Fig. II.5 a representation of the model of the GMF with three layers. This decomposition of the structure of the magnetic field has been employed in [146] making use of seven layers to fit polarization data of the southern Galactic cap¹. This approach has been extended in [148] to build a model that also fits the observed power spectra and cross-

¹Note that, in this model, the number of layers is not a very well constrained parameter, although it has been shown that it must be kept reasonably small to correctly reproduce the PDF of the polarization fraction.

spectra of the dust emission at 353 GHz. This notably incorporates in the model a TE correlation, and an asymmetry of power between E and B -modes. A similar model has been employed in [149] and [150] to constrain the statistical properties of the emission at both southern and northern Galactic caps combining *Planck* and HI data. These models comprise three layers that represent the cold and warm neutral phases of the ISM, and thermally unstable gas at intermediate temperatures. The hydrogen column density in each of these three layers is derived from a Gaussian decomposition of HI spectroscopic data (the GASS and EBHIS surveys). Note that HI observations also allow to constrain the 3D structure of the GMF building on its alignment with the filamentary structure of the atomic ISM [151, 152]. In line with the statistical approaches discussed here, in [152], the authors built from HI4PI data 3D (position-position-velocity) Stokes parameter maps Q and U based on the estimation of local orientation of the HI emission as a function of the velocity dimension of the observations. While purely deterministic, this approach could hopefully help to improve the statistical models of the magnetic field discussed in this paragraph.

Another strategy to model the polarized emission of interstellar dust has been implemented in [153]. This approach defines statistical models for both n_H and \mathbf{B} variables. Contrary to the previous strategy where the turbulent magnetic field was drawn on spheres, here the models are true 3D models (within the flat-sky approximation). The statistical models are derived from 3D fractional Brownian motions (fBm), which are examples of self-similar GRFs (see Appendix A for a definition). The gas density n_H is modeled as the exponential of a 3D fBm, while the magnetic field \mathbf{B} is defined as the rotational of a vector potential whose components are independent 3D fBms. The Hurst exponents associated with these 4 fBms are parameters of the model. This model was applied to the characterization of the Polaris Flare molecular cloud by means of a Markov chain Monte Carlo fitting of its parameters.

Finally, we must recall that the strategy most directly motivated by ISM physics to define a statistical model of the polarized emission of dust remains that relying on MHD simulations (see e.g. [154, 155, 94]).² However, the computational cost of a MHD simulation is usually incomparable with that needed to generate simulated Stokes maps according to the previous strategies. In consequence, one will not be able to fit the parameters of a statistical model based on MHD simulations with the same flexibility as with the above strategies. For this reason, these MHD-based models cannot be put on the same level as phenomenological models. This underlines the need to find a reasonable trade-off between the computational cost needed to generate realizations of the model, and the degree of refinement of the model.

II.3 Agnostic models

Since 2019, a new class of models of the dust foreground has emerged, taking an opposite direction to phenomenological models. These models are built making no use of the physical nature of the data, and because they are applied to dust data in the same way as they would be applied to images of cats and dogs, we call these models *agnostic*. These

²Strictly speaking, models based on MHD simulations are not phenomenological models as they derive directly from fundamental MHD equations.

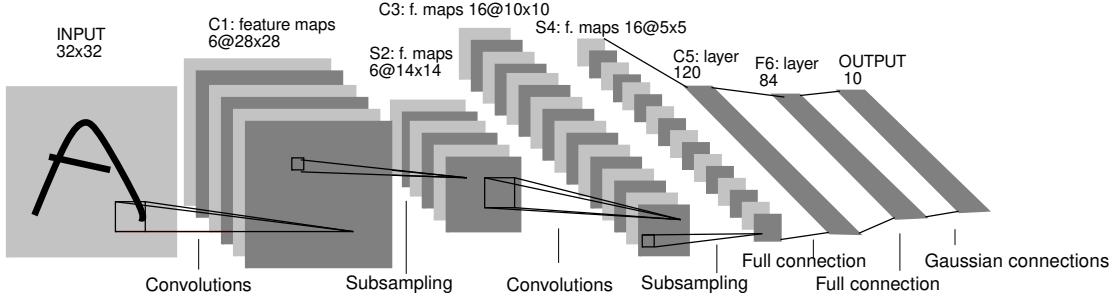


Fig. II.6 Architecture of LeNet-5, which is a deep neural network designed for the classification of images of handwritten digits. The CNN part of the network is made of a succession of convolutional layers (convolutions + pointwise nonlinearities) and pooling layers. The second part of the network is the actual classifier and is made of fully connected layers. Figure is taken from [156].

are derived from recent advances in machine learning, and particularly from an emerging subfield named *deep generative modeling*. Deep generative modeling relies on neural networks architectures to approximate probability distributions in a high-dimensional space. These models are thus statistical parametric models, the parameters being the weights of the network. Contrary to phenomenological models, the interpretability of the parameters of such models remains very remote. After a very short introduction on the structure of convolutional neural networks (CNNs) which are at the core of these agnostic models, we discuss the various deep generative models that have been used in the dust foreground modeling context.

II.3.1 Convolutional neural networks

CNNs are deep neural networks³ that were originally inspired by the workings of the visual cortex [157]. Since their popularization in the 1990s by Le Cun [158], CNNs have brought spectacular applications to various domains among image recognition, video analysis, natural language processing, drug discovery, or financial time series analysis. We briefly introduce these networks here, and refer the reader to general textbooks for a detailed presentation (e.g. [159]).

The specificity of CNNs is their ability to build a representation of some input data that is essentially based on successive convolutions by filters (also called kernels), pointwise nonlinearities, and downsampling operations. The filters are parameters of a CNN and these are learned during a training step. More precisely, in the context of image analysis which is our concern here, with $x : \llbracket 0, M - 1 \rrbracket \times \llbracket 0, N - 1 \rrbracket \rightarrow \mathbb{R}$ an image of size $M \times N$, the convolution of x with a filter $\psi : \llbracket 0, K - 1 \rrbracket^2 \rightarrow \mathbb{R}$ of size $K \times K$, denoted $x \star \psi$ and called a *feature map*, reads:

$$x \star \psi(r_1, r_2) = \sum_{k_1=0}^{K-1} \sum_{k_2=0}^{K-1} x(r_1 - k_1, r_2 - k_2) \psi(k_1, k_2). \quad (\text{II.18})$$

³The adjective "deep" refers to the use of multiple layers in these networks.

Finite dimensions of the image x also demand to choose a padding method to deal with pixels at the boundaries of x (zero-padding, periodic boundary conditions, etc). After a convolution, pointwise nonlinearities, which are defined by an activation function ρ , are usually applied to the resulting feature maps. This activation function is typically a rectified linear unit (ReLU), which simply refers to the function $u \rightarrow \max(0, u)$. Finally, downsampling operations, defining *pooling layers*, are used to progressively lower the dimension of the images and increase the invariance of the network to small translations. These can also help the network to learn an invariance to other kinds of transformations such as rotations. Common downsampling functions are the max pooling function, corresponding to a downsampling of an image made by only keeping the maximum value of pixels in local neighborhoods, and the average pooling function, keeping local averages of pixels.

These operations are building blocks of the network. They are organized in sequential functional layers, and the output of the network may be connected to other networks depending on the application. Typically, for a classification problem, which is a very common problem of supervised learning, the purpose of the CNN is to extract relevant features from data before the actual classification. To give a more concrete example, let us sketch the typical architecture of a CNN that is designed for an image classification problem, such as the classification of handwritten digits which made the CNNs famous in the 1990s (see e.g. LeNet [156] architecture shown in Fig. II.6). Such a network takes as an input an image x , and the first layer would typically correspond to a convolutional layer, computing a set of feature maps $x_i^{(1)} = x \star \psi_i^{(1)}$, with $\{\psi_i^{(1)}\}_i$ a corresponding set of filters. After the application of an activation function ρ , the second layer would be a pooling layer, which would extract downsampled maps $x_i^{(2)}$ from each of the $\rho(x_i^{(1)})$ maps. A third layer would be another convolutional layer computing an additional set of feature maps $\{x_i^{(3)}\}$. It would be connected to all, or part, of the maps of the previous layer, such that each map $x_i^{(3)}$ would be defined as:

$$x_i^{(3)} = \sum_j x_j^{(2)} \star \psi_{i,j}^{(3)}, \quad (\text{II.19})$$

with $\{\psi_{i,j}^{(3)}\}_j$ the set of filters related to $x_i^{(3)}$. Additional layers could be stacked similarly. Finally, the output of this network might be connected to an actual classifier, usually made of fully connected layers (i.e. layers made of neurons that are connected to all the neurons of the previous layer).

In these networks, filters are optimized through a training process. This consists in a minimization of a given loss function, which is made possible thanks to the combination of the backpropagation algorithm and an optimizer (see e.g. [159]). For a classification problem, the loss is typically a mean-square error measuring the error of classification of the network on already-classified data.

The success of CNNs originally comes from their application to supervised learning problems, e.g. classification problems or regression problems. However, for a few years new architectures of CNNs have been designed to tackle more intricate problems of unsupervised learning. Prime examples are density estimation problems for high-dimensional data, which is the concern of deep generative modeling.

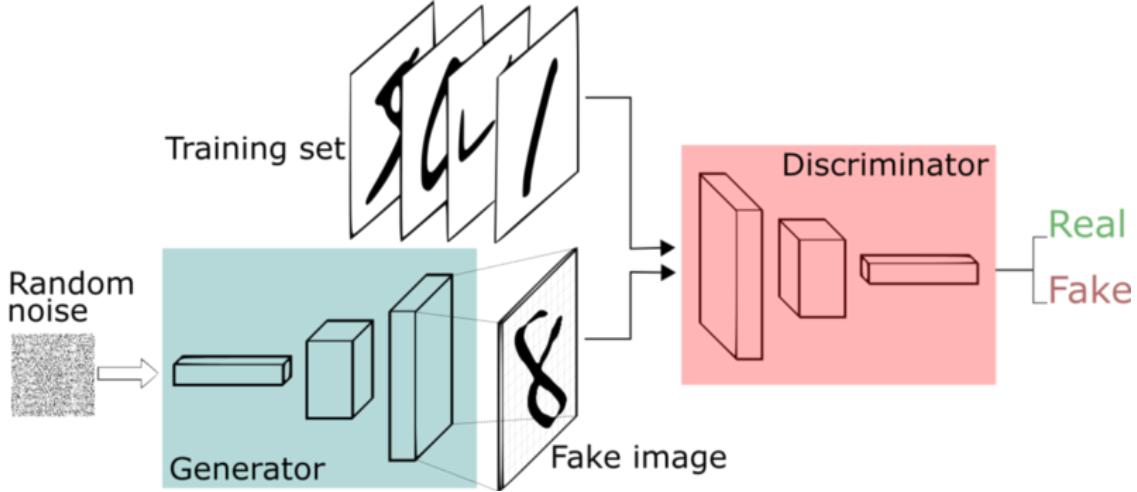


Fig. II.7 Illustration of a GAN. The generative network is trained to fool the discriminative network, while the latter is trained to classify mock and training data. Credits: Thalles Silva.

II.3.2 Deep generative modeling and the dust foreground

The most prominent generative neural network architectures are generative adversarial networks (GANs) [160], and variational autoencoders (VAEs) [161]. These architectures have been recently applied to model the dust foreground in a series of works [73, 133, 162]. We review here some of their main results.

A GAN is a combination of two networks: a generative network (G) and a discriminative network (D). These are trained as if they were two opponents in a game. The goal of D is to estimate the probability for its input to be part of the training set or to be mock data generated by G, while the goal of G is to generate mock data that would fool D with the highest probability (see Fig. II.7 for an illustration). Once the network is trained, G thus defines a generative statistical model of the data. In [73], the authors trained a modified DCGAN (deep convolutional GAN) [163] using approximately a thousand of $20^\circ \times 20^\circ$ patches extracted from the GNILC 353 GHz dust intensity map. Then, they assessed the quality of the mock maps by comparing them to the real maps using the following summary statistics: PDFs of the pixel values, empirical power spectra and their uncertainties, and Minkowski functionals and their uncertainties. The mock maps seem realistic by eye (see Fig. II.8), and show globally good quantitative agreements with the original maps, although significant discrepancies remain (for the tails of the PDFs, and Minkowski functionals notably). In [133], the authors also trained a modified DCGAN using the same intensity data but here focusing on the generation of the smallest scales of the data given their largest scales. They have shown good agreement of the Minkowski functionals, and power spectra of their mock small scales maps with that of the real ones. Then, they applied the same strategy to polarization Q and U maps separately. However, GNILC polarization maps are much less resolved than intensity maps due to the globally lower SNR of polarization measurements. Thus they assumed the smallest scales of intensity maps to be representative of that of Q and U maps and used them to train their

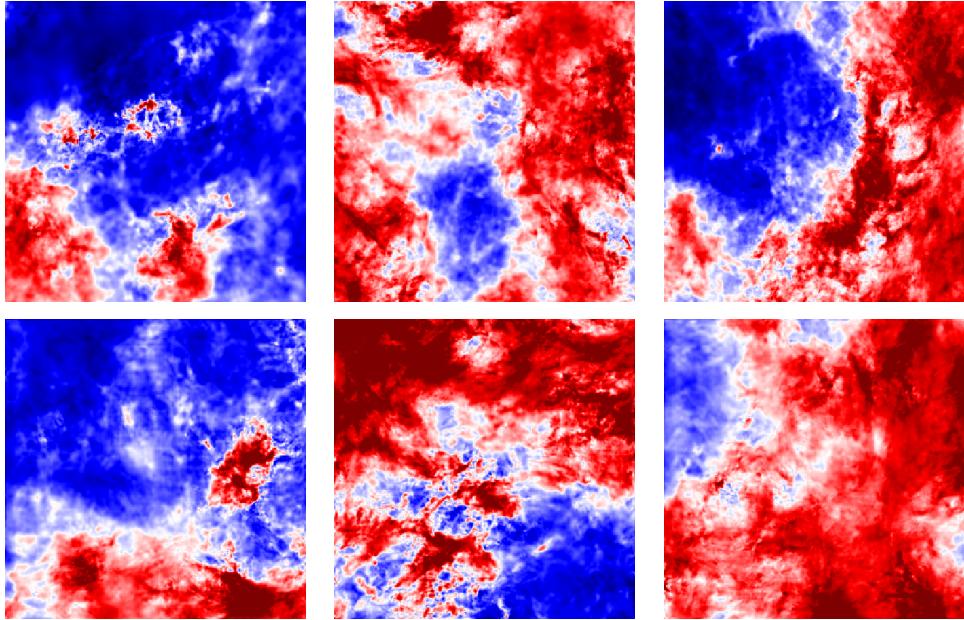


Fig. II.8 Selection of $20^\circ \times 20^\circ$ patches from the GNILC 353 GHz intensity map (top row), and mock maps generated by the GAN introduced in [73] (bottom row). Figure is taken from [73].

network in a similar way. This assumption allowed them to generate a sky map of the polarization signal with mock small scales.

A VAE aims at compressing some input data into a multivariate distribution defined by *latent variables* (*encoder* part), and from which the input data can be reconstructed (*decoder*). In [162], the authors built a convolutional VAE and trained it using $8^\circ \times 8^\circ$ patches of the GNILC 545 GHz intensity map. They used the power spectrum and the PDF of pixel values as summary statistics to assess the quality of their model. They have shown a good agreement of the PDF of pixel values and of the power spectrum at large scales between reconstructed maps and their corresponding original maps (picked from the test set). However an important flaw is that the reconstructed maps have a blurry aspect, which is quantitatively associated with a lack of power at small scales compared to the original data (see Fig. II.9). The authors attributed this flaw to a recurrent issue of these VAE models.

II.4 The best of both worlds?

The dust foreground has two "dimensions": a spectral one, and a spatial one. The spectral dimension is at the core of phenomenological deterministic modeling. It is exploited to fit the parameters of MBB-based models, or more physical models of interstellar dust, to multi-frequency observations, thus providing maps of the dust foreground and of its parameters. This kind of model gives a direct valuable insight into the properties of the ISM. However, they do not leverage the spatial structure of the data yet which has

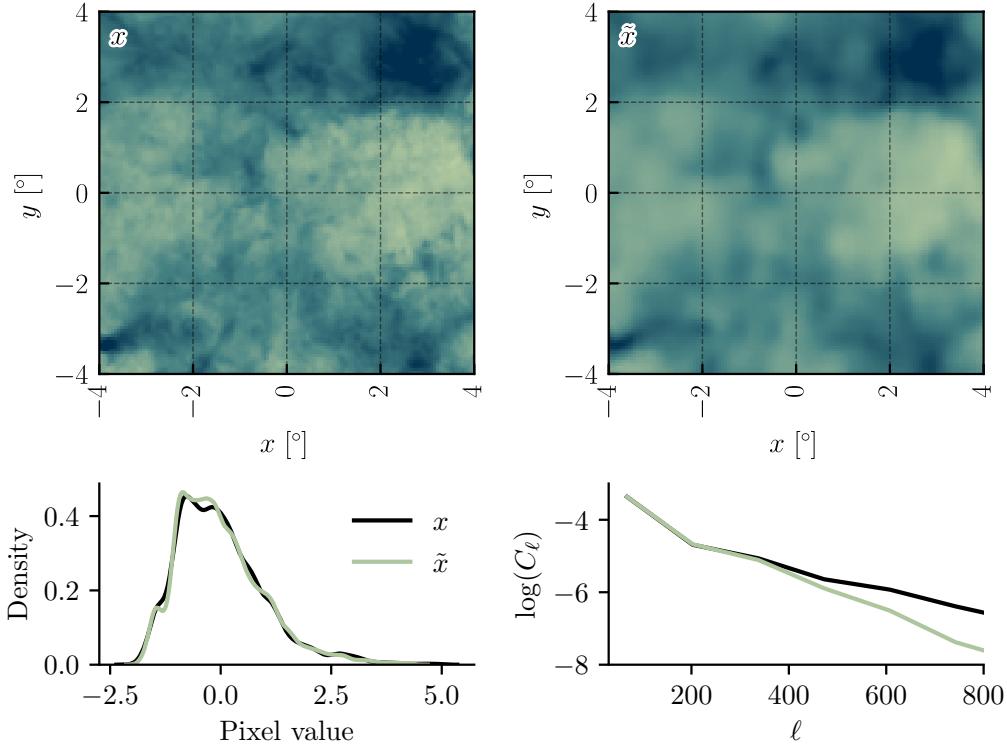


Fig. II.9 *Top panel:* Test intensity map x (left) next to its reconstruction \tilde{x} with a VAE (right). The test map is a $8^\circ \times 8^\circ$ patch of the GNILC 545 GHz intensity map. *Bottom panel:* Distribution of pixel values (left), and log power spectra (right) of the original image and its reconstruction. Figure is taken from [162].

been shown to exhibit important distinctive properties. On the contrary, so far statistical models have focused on the characterization of this spatial structure, with either phenomenological models or agnostic models to capture and simulate the spatial statistical properties of the dust emission. Some phenomenological statistical models make use of GRFs with power-law power spectra, the turbulent component of the magnetic field being a GRF as in [148, 153], the density field being the exponential of a GRF as in [153]. However this approach implicitly ignores the interactions between scales that are inherent to nonlinear physics. Interstellar turbulence, which is thought to be responsible in great part of the structure of the emission, is fundamentally nonlinear, and capturing the interactions between scales is crucial if one wants to fully characterize its properties. On a side note, let us mention that non-Gaussian alternatives to GRFs exist to describe the physical fields of the ISM. For example, in [164, 165], the authors have introduced statistical models to characterize 3D MHD turbulence using multiplicative chaos. This approach gives a practical way to parametrize the intermittency of the random fields, which is an important non-Gaussian property of the ISM. MHD simulations can also bridge this gap, however their computational cost, and their lack of flexibility are their main drawbacks. For example, merely reproducing *Planck* power spectra statistics remains a significant challenge for MHD simulations [154, 155]. Agnostic models, and the CNNs on which they rely, directly tackle the problem of capturing interactions between scales. A CNN

is designed to identify connections between feature maps that are themselves optimized to represent the data in some sense. This is a promising avenue for dust modeling, although the definition of the training set is a difficulty that may restrict the relevance of the model. Though training typically requires a training set that is as large as possible, from a physical point of view we want to restrict the training set to data that is most representative of the target field being modeled. We also point out that the filters, which are learned during the training of the network, remain much less interpretable than the parameters of phenomenological models.

In the following chapters, we will explore a new avenue to model the dust foreground, attempting to be halfway between agnostic models and phenomenological models. We will focus on the statistical modeling of the spatial properties of the polarization signal of the dust foreground, with particular care for the characterization of interactions between scales, which are fundamental for highly nonlinear physics. Note that, as a first step, we will ignore the spectral dimension of the data. We will take advantage of recent advances in data science aiming at understanding the mathematical properties of CNNs, and the origin of their success to build relevant representations of data [166]. We will define models based on non-parametric representations of the data that are directly inspired from the structure of CNNs, namely the *wavelet scattering transform* (WST) [3, 4] and the *wavelet phase harmonics* (WPH) [5, 6]. Note that we will conveniently refer to these two representations as *scattering-like representations*. These take the form of statistics allowing to characterize non-Gaussianity by means of a characterization of interactions between scales. Contrary to deep generative modeling approaches, generative models based on these statistics do not involve any training step (see Chapter IV). Moreover, we will see that they allow for interpretability, which permits to relate the model to the physics of the ISM to a certain extent (see Chapter III). Finally, we will see that, while they may be impacted by noise, this will not prevent us to propose solutions to build satisfying models from noisy observational data (see Chapter V).

Chapter III

Statistical characterization with the wavelet scattering transform

The statistical characterization of the diffuse magnetized ISM poses a major challenge. To account for its non-Gaussian statistics, we need a data analysis approach capable of efficiently quantifying couplings across scales. This information is encoded in the data, but most of it is lost when using conventional tools, such as one-point statistics and power spectra. The wavelet scattering transform (WST), a low-variance statistical descriptor of non-Gaussian processes, opens a path towards this goal.

After a presentation of the WST, we apply this transform to noise-free maps of dust polarized thermal emission computed from a numerical simulation of MHD turbulence in the diffuse ISM. We analyze normalized complex Stokes maps and maps of the polarization fraction and polarization angle. The WST yields a few thousand coefficients; some of them measure the amplitude of the signal at a given scale, and the others characterize the couplings between scales and orientations. The dependence on orientation can be fitted with the reduced wavelet scattering transform (RWST), an angular model introduced in [167] for total intensity maps. The RWST provides a statistical description of the polarization maps, quantifying their multiscale properties in terms of isotropic and anisotropic contributions. It allows us to exhibit the dependence of the data on the orientation of the mean magnetic field and to quantify non-Gaussianity.

In this chapter, we establish a methodology, showing how we can extract a few hundred interpretable statistical coefficients from polarization maps of the dust emission that characterize its non-Gaussian properties. This constitutes a first step towards a statistical modeling of the spatial properties of dust. Actual models based on the WST are discussed in the next chapter.

This work has also motivated the development of a Python package, called PyWST, designed to perform WST and RWST analyses of two-dimensional data.¹ We refer to Appendix C for a detailed presentation of this package.

Note that an important part of this chapter is directly extracted from [168].

¹PyWST is available at <https://github.com/bregaldo/pywst>.

III.1 The wavelet scattering transform

III.1.1 Motivations

The wavelet scattering transform (WST) was first introduced by Mallat in [169, 3]. One of the main motivations from the perspective of data science was to provide a mathematical framework to understand the properties of convolutional neural networks (CNNs, see Sect. II.3.1 for an introduction) and the origin of their success for classification problems [166]. CNNs manage to build relevant representations of data that get rid of its uninformative variability, thus allowing spectacular applications especially for classification problems (see e.g. [170]). A core notion of these CNNs is that of *invariance*. CNNs somehow learn to be invariant to the transformations that do not change the nature of the data, thus projecting the high-dimensional data in a space of lower dimension where the uninformative variability disappears. A concrete example is that of handwritten digits classification as explained in [4]. For this type of problem, the classification of handwritten digits should not be affected by a global translation or rotation of the input image. Ideally, one thus needs a representation that remains invariant to these transformations. Similarly, different handwritings, which can be assimilated to small deformations of the digits, must not perturb the classification. However, a representation of the data for handwritten digits classification cannot be completely invariant to deformations, as it would for example involve misclassifications between the digit 7 and the digit 1 which can be continuously deformed one into the other. Instead of an invariance to deformations, [4] thus requires a linearization of deformations, called *Lipschitz continuity to deformations*, so that the changes induced in the representation by a deformation remain controlled by the size of this deformation.

The WST allows us to define such representations of data. These can be made invariant to a given class of transformations (typically translations, rotations, or scalings), and are Lipschitz-continuous to deformations by construction [3]. Contrary to CNNs, whose filters are learned during a training step, the WST does not necessitate any training as the filters are fixed a priori; these correspond to *wavelets*, as we will see. The WST has been applied to various kinds of signals such as audio signals [171], images [4, 172], 3D distributions of molecular electron density [173]. The WST also provides representations of stationary non-Gaussian processes [3, 174], by means of so-called *scattering moments*. The estimators of these moments benefit from a low-variance compared to usual estimators of high-order moments (e.g. bispectrum). On a practical level, this means that accurate estimates of the scattering moments do not necessarily require a large amount of samples.

From the perspective of astrophysics, the interest of the WST mainly relies in that it provides an interpretable low-variance characterization of non-Gaussian processes, with a quantification of couplings between scales. Low-variance estimators are advantageous for any astrophysical extended emission as, contrary to digits classification problems, we are far from being overwhelmed by the amount of available data. Moreover, we will see that the interpretability of the WST coefficients stems from the multiscale approach introduced by the wavelet transform as well as the invariance and stability properties of the transform. Let us finally mention that the WST can be applied locally, which is of strong interest for statistically inhomogeneous data, which are common in all observations.

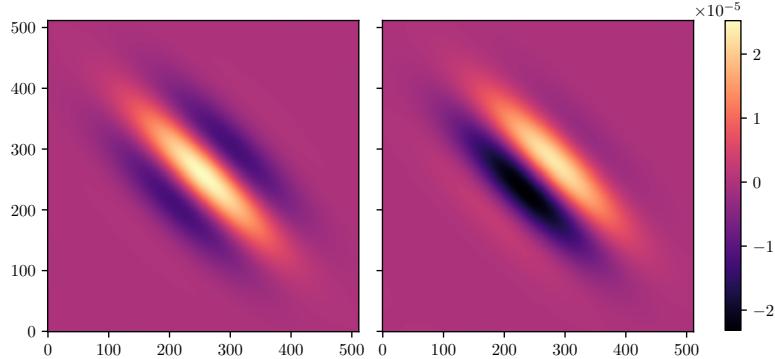


Fig. III.1 *Real part (left) and imaginary part (right) of a 512×512 two-dimensional Morlet wavelet $\psi_{6,\pi/4}$. The wavelet is centered on the middle of the map for a better visualization.*

[167] provided the first astrophysical application of the WST, by applying it to column density maps inferred from MHD simulations and to an *Herschel* observation of the thermal emission from Galactic dust [125]. The physical regularity of the maps led them to introduce the reduced wavelet scattering transform (RWST), a statistical description of reduced dimensionality, obtained through a fit of the angular dependence of the WST coefficients. I extended this work to the analysis of simulated maps of the polarized thermal emission of dust in [168]. We discuss in this chapter the common methodology of these two previous works and present in details the results of [168]. Note that the WST has since been applied to cosmological inference problems in [175, 176] in the context of weak lensing, and more recently, [177] applied it to the classification of MHD simulations in terms of sonic and Alfvénic Mach numbers.

We emphasize that although this approach relies on the wavelet transform, it differs from existing wavelet analysis strategies already discussed in the ISM literature (see e.g. [178, 179]). Similarly to these previous strategies, the WST uses wavelets as building blocks for a multiscale analysis, however, it goes beyond classical analyses by also quantifying interactions between scales.

III.1.2 Definition

We introduce here the WST for its application to image analysis in the same way as in [4, 167]. We consider an image x of size $M \times N$ pixels which may be real or complex-valued.

a) The wavelet transform

The WST relies on the wavelet transform. We briefly introduce the wavelet transform here, and refer the reader to Appendix B for additional details and illustrations.

A wavelet is a localized waveform with a zero mean which acts as a bandpass filter. In this chapter we will make use of Morlet wavelets, which are essentially plane waves modulated by Gaussian envelopes (see Fig. III.1). A Morlet wavelet is a quite general kind of complex-valued wavelet to study physical fields and it provides a good angular

selectivity [180]. It is defined in 2D as:

$$\psi(\mathbf{r}) = \alpha \left(e^{i\mathbf{k}_0 \cdot \mathbf{r}} - \beta \right) e^{-|\mathbf{r}|^2/2\sigma^2}, \quad (\text{III.1})$$

with α and β two constants that are adjusted to ensure a zero mean and a unit L^1 norm, $\mathbf{k}_0 = k_0 \mathbf{e}_x$ the wave vector of the plane wave factor, and σ the standard deviation of the Gaussian envelope^{2,3}. A bank of wavelets $\{\psi_{j,\theta}\}_{j,\theta}$ is built from dilations and rotations of the initial wavelet ψ , called the *mother wavelet*, where j is an index of dilation and θ is an angle of rotation. Formally, we have:

$$\psi_{j,\theta}(\mathbf{r}) = 2^{-2j} \psi(2^{-j} R_\theta^{-1} \mathbf{r}). \quad (\text{III.2})$$

The index j thus refers to a dilation by a factor 2^j , and the scaling factor 2^{-2j} in this equation ensures the conservation of the L^1 norm (see Appendix B for a definition) under these dilations. The number of wavelets of our bank will be parametrized by two integers J and L , where J is the number of dilations so that $0 \leq j \leq J-1$, and L is the number of discrete angles that evenly divide the interval $[0, \pi)$ so that $\theta \in \{k\pi/L, 0 \leq k \leq L-1\}$. For real-valued images we only need to consider $\theta \in [0, \pi)$, while for complex-valued images θ will take its values in $[0, 2\pi)$. The bank of wavelets will thus be made of either $J \times L$ or $2 \times J \times L$ wavelets for the analysis of real or complex-valued images, respectively.

The wavelet transform of x is finally defined as the set of bandpass filtered maps $\{x \star \psi_{j,\theta}\}_{j,\theta}$, where \star denotes the circular convolution operation (see Appendix B for a definition). These convolutions correspond to local bandpass filterings of x at spatial frequencies centered on modes of the form $\mathbf{k}_{j,\theta} = k_0 2^{-j} (\cos(\theta) \hat{\mathbf{e}}_x + \sin(\theta) \hat{\mathbf{e}}_y)$.

b) The WST coefficients

Inspired by the structure of CNNs which cascades convolutions, pointwise nonlinearities, and pooling operations, the WST of x is defined as a cascade of wavelet transforms, pointwise moduli, and average pooling operations. The average pooling operations can be done locally or globally, thus introducing respective local and global WST statistics.

Global WST. The WST coefficients S_0 , S_1 , S_2 , corresponding to the output of the zeroth, first, and second layer of operations, respectively, are defined as follows:

$$S_0 = \begin{cases} \langle x \rangle & \text{if } x \text{ is real-valued,} \\ \langle |x| \rangle & \text{if } x \text{ is complex-valued,} \end{cases} \quad (\text{III.3})$$

$$S_1(j_1, \theta_1) = \langle |x \star \psi_{j_1, \theta_1}| \rangle, \quad (\text{III.4})$$

$$S_2(j_1, \theta_1, j_2, \theta_2) = \langle ||x \star \psi_{j_1, \theta_1}| \star \psi_{j_2, \theta_2}| \rangle, \quad (\text{III.5})$$

where the averaging operator $\langle \cdot \rangle$ here is simply an average over the definition domain T , meaning $\langle x \rangle = \frac{1}{|T|} \int_T x(\mathbf{r}) d^2\mathbf{r}$ with $|T|$ the area of T . These coefficients depend on the

²In practice we choose $k_0 = 3\pi/4 \text{ pixel}^{-1}$ and $\sigma = 0.8 \text{ pixel}$ as in [4].

³To simplify, we omit in Eq. (III.1) the fact that, in practice, the Gaussian envelope has an elliptical shape. This elliptical shape enhances the angular selectivity of the wavelets. We refer to Appendix B for additional details.

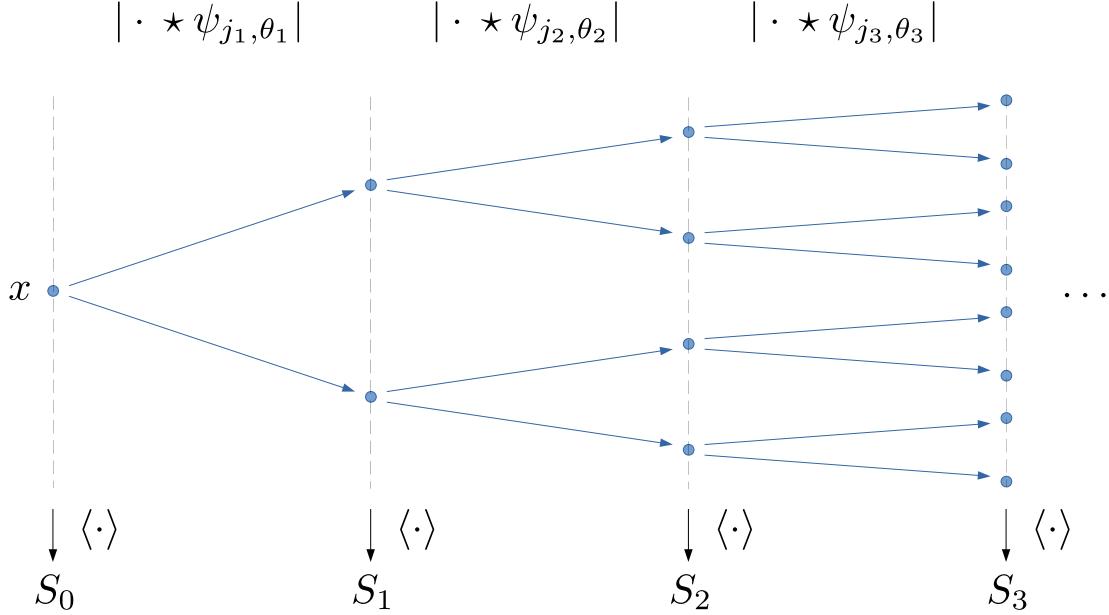


Fig. III.2 Network structure of the WST. We show the cascade of convolutions, pointwise moduli, and average operations leading to the definition of the WST coefficients. For simplicity, we represent two outgoing edges per node only.

indices of the wavelets involved for each convolution. We can define coefficients for deeper layers in a similar fashion, however, in practice, we only make use of the coefficients of these first three layers. This will be justified below. We show in Fig. III.2 a schematic view of the construction of WST coefficients, which illustrates the structural similarities of the WST with a typical CNN (see Sect. II.3.1). Note that the WST coefficients are all real coefficients.

Local WST. The WST coefficients can be computed locally by replacing the global averaging operator $\langle \cdot \rangle$ by local averages. The local WST may be relevant when the statistical properties of x are not expected to be homogeneous, as for astrophysical observations. To compute local averages, we usually introduce a low-pass filter ϕ_J which depends on the number of scales J we consider in our analysis. We typically choose a Gaussian window with size 2^J . Finally, the local WST coefficients simply correspond to local samples of S_0 , S_1 , and S_2 , which are now functions of the position \mathbf{r} and defined as:

$$S_0(\mathbf{r}) = \begin{cases} x \star \phi_J(\mathbf{r}) & \text{if } x \text{ is real-valued,} \\ |x| \star \phi_J(\mathbf{r}) & \text{if } x \text{ is complex-valued,} \end{cases} \quad (\text{III.6})$$

$$S_1(j_1, \theta_1, \mathbf{r}) = |x \star \psi_{j_1, \theta_1}| \star \phi_J(\mathbf{r}), \quad (\text{III.7})$$

$$S_2(j_1, \theta_1, j_2, \theta_2, \mathbf{r}) = ||x \star \psi_{j_1, \theta_1}| \star \psi_{j_2, \theta_2}| \star \phi_J(\mathbf{r}). \quad (\text{III.8})$$

Scattering moments. When x is viewed as the realization of a statistically homogeneous random field X , the WST coefficients may be considered as estimates of the

scattering moments of X . These are defined similarly to the WST coefficients but replacing spatial averages $\langle \cdot \rangle$ by expected value operators $E[\cdot]$ (and involving X instead of x obviously). Natural estimators of the scattering moments are unbiased and have a reduced variance compared to estimators of high-order moments as discussed in [4]. We will conveniently conflate these two views in the context of this thesis.

c) Normalization and interpretation

In practice, we will make use of normalized coefficients \bar{S}_0 , \bar{S}_1 , and \bar{S}_2 . These are defined in a similar way as in [174], that is, in the case of global WST coefficients:

$$\bar{S}_0 = S_0, \quad (\text{III.9})$$

$$\bar{S}_1(j_1, \theta_1) = \frac{\langle |x \star \psi_{j_1, \theta_1}| \rangle}{S_0}, \quad (\text{III.10})$$

$$\bar{S}_2(j_1, \theta_1, j_2, \theta_2) = \frac{\langle ||x \star \psi_{j_1, \theta_1}| \star \psi_{j_2, \theta_2}| \rangle}{S_1(j_1, \theta_1)}. \quad (\text{III.11})$$

Note that \bar{S}_0 and S_0 are identical. This normalization allows us to increase the invariance of the coefficients of the first and second layers. It makes both first and second-layer coefficients invariant to a multiplication of x by a global factor. Added to that, \bar{S}_2 coefficients at fixed (j_1, θ_1) are invariant to linear filters with a uniform action on the bandpass of the wavelet ψ_{j_1, θ_1} .

These normalized coefficients can be interpreted in the following manner: \bar{S}_0 is simply the mean of the field (or of its modulus when x is complex-valued), $\bar{S}_1(j_1, \theta_1)$ is a measure of the amplitude of modes of the normalized field $x/\langle x \rangle$ (or $x/\langle |x| \rangle$) that are centered on $\mathbf{k}_{j_1, \theta_1}$, and finally $\bar{S}_2(j_1, \theta_1, j_2, \theta_2)$ characterizes how the normalized amplitude of modes centered on $\mathbf{k}_{j_1, \theta_1}$ is modulated by modes centered on $\mathbf{k}_{j_2, \theta_2}$. In other words, the $\bar{S}_1(j_1, \theta_1)$ coefficient characterizes the amplitude at a single oriented scale (j_1, θ_1) , while the $\bar{S}_2(j_1, \theta_1, j_2, \theta_2)$ coefficient measures a form of coupling between two oriented scales (j_1, θ_1) and (j_2, θ_2) .

We note that even if the \bar{S}_1 coefficients characterize the Fourier amplitude of the field under study in the spectral band of the wavelets, they differ in practice from the power spectrum. While the power spectrum relates to L^2 norms of the wavelet transform, the \bar{S}_1 coefficients involve L^1 norms. One can however recover the power spectrum with a good approximation from a quadratic sum of \bar{S}_1 and \bar{S}_2 coefficients⁴. Notably, we expect a non-Gaussian field to have higher \bar{S}_2 coefficients compared to those of a Gaussian field with identical power spectrum, and this should be counterbalanced by smaller \bar{S}_1 coefficients. This is related to the sparsity of the wavelet representation of the data (see [167] and [4] for further discussions).

The WST coefficients of x also relate to some of its normalized high-order moments up to the order 4 [3]. However, it is not clear how these may relate to the bispectrum mentioned in Sect. II.1.3. Let us also point out that if the phases of the $\{x \star \psi_{j, \theta}\}$ signals naturally disappear once the modulus is applied, this does not mean that this phase information is completely lost as it can be recovered from the $\{|x \star \psi_{j, \theta}|\}$ signals for particular wavelets [4, 181].

⁴See Eq. (9) in [167].

III.1.3 Properties

We now discuss the properties of the WST. Let us call $S[x] = \{S_0, S_1, S_2\}$ the set of WST coefficients associated to an image x , and $\|S[x]\|$ the Euclidean norm of the corresponding vector of coefficients. The energy of an image x is defined as its quadratic Euclidean norm $\|x\|^2$.

Inspired by the invariance properties of CNNs, the global WST is naturally invariant to translations, due to the covariance of convolutions with respect to translations⁵ and the action of the map average operator. For local descriptions however, the local WST coefficients are only invariant to local translations. No rotational invariance is enforced here. While it is not a limitation of the approach (see [3, 172] for alternative definitions of the WST), a rotational invariance would be irrelevant for the characterization of the dust foreground as we will see. However, note that the WST as defined here is covariant to the finite group of rotations including rotations by angles that are multiple of π/L , which is the discretization step of the angular variable θ chosen for the construction of the wavelet bank.

The WST is a nonexpansive transform, meaning that for two images x and y , we have:

$$\|S[x] - S[y]\| \leq \|x - y\|. \quad (\text{III.12})$$

This property makes the WST stable to additive noise, and the estimators of the scattering moments of reduced variance [4]. The WST is also able to linearize small deformations, being Lipschitz continuous to deformations [3, 4]. This makes the WST stable to deformations, which is of prime importance for classification purposes as discussed previously.

For appropriate wavelets which would make the wavelet transform unitary (see Appendix B), the WST preserves the signal energy [3, 4], allowing us to write:

$$\begin{aligned} \|x\|^2 &= \|S[x]\|^2 + \epsilon \\ &= S_0^2 + \sum_{j_1, \theta_1} S_1(j_1, \theta_1)^2 + \sum_{j_1, \theta_1, j_2, \theta_2} S_2(j_1, \theta_1, j_2, \theta_2)^2 + \epsilon, \end{aligned} \quad (\text{III.13})$$

where ϵ is a complementary minor term corresponding to the energy contained in higher-order WST coefficients. In practice, the energy contained in the coefficients of higher order rapidly decreases as the order increases, which is one of the reasons why we restrain the WST to S_0 , S_1 , and S_2 coefficients for image analysis [4]. Moreover, this relation shows that sparse wavelet coefficients, characterized by low S_1 values, necessarily involve a propagation of energy towards higher order coefficients. However, we must say that for Morlet wavelets, the wavelet transform is not unitary, and this conservation of energy is not exact even if it is reasonably well approximated.

Nevertheless, a conservation of energy does not imply a preservation of the information contained in the signal. Generally, the WST cannot be inverted, so that reconstructions based on WST coefficients only cannot be exact [4]. However, as we shall see it in the next chapter, the WST coefficients may define relevant statistical models, showing that statistical information may be correctly preserved to a certain extent.

⁵An operator A acting on functions $f : \mathbb{R}^n \rightarrow \mathbb{C}$ is *covariant* (or *equivariant*) to translations if for any translation $L_\tau f$ of a given function f , that is $L_\tau f(\mathbf{r}) = f(\mathbf{r} + \boldsymbol{\tau})$, we have $A[L_\tau f] = L_\tau A[f]$. This notion may be extended to other symmetry groups.

	S_0	S_1	S_2	Total
real field	1	56	1344	1401
complex field	1	112	2688	2801

Table III.1 – Number of WST coefficients per layer for $J = 7$ and $L = 8$ for both real and complex-valued data cases.

Not all the coefficients of the WST are equally informative. First, the modulus tends to push the energy of wavelet coefficients to lower frequencies, so that in practice S_2 coefficients with $j_2 \leq j_1$ are negligible, thus allowing to restrain the computation of S_2 coefficients with $j_2 > j_1$ only. Second, because for a Morlet wavelet $\psi_{j,\theta}$ we have $\psi_{j,\theta}^* = \psi_{j,\theta+\pi}$, provided x is a real-valued image, the WST coefficients verify:

$$S_1(j_1, \theta_1 + \pi) = S_1(j_1, \theta_1), \quad (\text{III.14})$$

$$S_2(j_1, \theta_1 + \pi, j_2, \theta_2) = S_2(j_1, \theta_1, j_2, \theta_2), \quad (\text{III.15})$$

$$S_2(j_1, \theta_1, j_2, \theta_2 + \pi) = S_2(j_1, \theta_1, j_2, \theta_2). \quad (\text{III.16})$$

For this reason, for real-valued images we only consider $\theta_1, \theta_2 \in [0, \pi)$. However, in the case of complex-valued images, only Eq. (III.16) holds in general, so that we will consider $\theta_1 \in [0, 2\pi)$ and $\theta_2 \in [0, \pi)$. Consequently, taking into account these simplifications, the effective number of global coefficients will be for a real-valued map: a single S_0 coefficient, $J \times L S_1$ coefficients, and $L^2 \times J \times (J - 1)/2 S_2$ coefficients. And, for a complex-valued map, there will be twice as many S_1 and S_2 coefficients. We give in Table III.1 the corresponding number of coefficients per layer for $J = 7$ and $L = 8$, which are the parameters chosen in the following.

III.2 Methodology illustrated on simulated Stokes maps

We now employ the WST to describe simulated data of the polarized dust emission. The main focus of this section is to introduce the methodology of our analysis, showing how immediate properties of regularity in the WST coefficients may be leveraged to define reduced WST (RWST) coefficients, which constitute a simple and interpretable description of the non-Gaussian statistical properties of the dust emission. We will focus on the statistics of polarization maps as it is the main target for B -mode detection, although this methodology remains general and applies similarly to intensity maps.

III.2.1 Building simulated Stokes maps

We work with Stokes maps I , Q , and U which are directly computed through integrations along a given direction of 3D cubes of data extracted from a numerical simulation of magnetized interstellar turbulence (this direction of integration is the analog of the line of sight for observations). This procedure follows what was already outlined in Sect. I.3.4. Working with simulated data avoids the usual complications of observational data, namely noise, non-periodic boundary conditions, and potential statistical inhomogeneity. As a first step, this will facilitate the physical interpretation of our statistical descriptions.

a) Presentation of the simulation

We use a MHD simulation designed to study the biphasic nature of the diffuse ISM within the context of [117]. The simulation employs the adaptive mesh refinement code RAMSES [112, 113] to solve the equations of ideal MHD as described in [182], neglecting self-gravity and taking into account heating and cooling processes of the gas. A turbulent forcing is applied to inject kinetic energy and balance numerical dissipation, thus allowing the simulation to reach a statistical steady state. In practice this turbulent forcing consists in a large-scale stochastic force field (a three component Ornstein-Uhlenbeck process defined in spectral space). Details on the turbulent forcing model can be found in [119].

The simulation volume is a $(50 \text{ pc})^3$ box divided into 512^3 cells with periodic boundary conditions. At $t = 0$ the gas has uniform properties with a density $n_{\text{H}} = 1.5 \text{ cm}^{-3}$ and a temperature $T = 8000 \text{ K}$, and the magnetic field is also uniform, with $\mathbf{B}_0 = B_0 \mathbf{e}_x$ and $B_0 \sim 3.8 \mu\text{G}$. In steady state, the turbulent forcing leads to an approximate velocity dispersion $\sigma_v \sim 2.6 \text{ km/s}$. At the scale of the box, that is 50 pc, the turnover time is thus $\tau_L \approx 18.8 \text{ Myr}$. Finally, an isotropic Habing radiation field is applied at the boundaries of the box. Its intensity is scaled by a factor $G_0 = 1$ (see [75] for a definition).

Once the simulation has reached a statistical steady state, we extract 14 snapshots that are approximately statistically independent using an approach similar to [183]. In practice we make sure that two consecutive snapshots correspond to a minimal time evolution of $\delta\tau = 1.25 \text{ Myr}$ which is roughly ten percent of τ_L . The phenomenology of turbulence in the sense of Kolmogorov shows that the turnover time τ_l at a given scale l scales as $l^{2/3}$ [84]. This scaling holds for incompressible hydrodynamical turbulence only but we assume for the sake of simplicity that it extends reasonably well to ideal compressible MHD. At the range of scales considered in the following, we find that $\delta\tau$ is about five times smaller than the corresponding largest value of τ_l . Although this value is not completely satisfactory, we assume in the following that the snapshots are statistically independent.

b) Stokes maps derivation

We compute a set of Stokes maps I , Q , and U for each of these snapshots following Eqs. (II.12)-(II.14) and choosing both the z -axis and the x -axis of the cubes as the directions of integration. This integration procedure is identical to the one used in [103]. We choose a typical value for the polarization fraction parameter $p_0 = 0.2$ and arbitrary values for σ_ν and T_d as these only determine the global amplitude of I , Q and U maps, which does not impact the following analysis. These maps are relevant for any frequency ν provided that the dust emission remains optically thin.

We will not work directly with these Stokes maps, but rather focus on the statistical properties of complementary maps derived from them. These are chosen to ease the physical interpretation of the data and hopefully simplify the statistics. We compute three kinds of dimensionless maps: a normalized version of the polarization maps $Q + iU$ denoted $\tilde{Q} + i\tilde{U}$, polarization fraction maps p and complex polarization angle maps $\exp(2i\psi)$ (see Sect. I.3.3 for a definition of p and ψ). The normalized variable $\tilde{Q} + i\tilde{U}$ is defined as

follows:

$$\tilde{Q} + i\tilde{U} = \frac{Q + iU}{I + P}, \quad (\text{III.17})$$

with $P = |Q + iU|$ the polarized intensity. This definition allows to roughly disentangle the structure of the magnetic field and the structure of dust density, at least at the lowest order. Indeed, in the limit where the orientation of the magnetic field is constant along the line of sight, one gets using Eqs. (II.12)-(II.14):

$$\tilde{Q} + i\tilde{U} = \frac{3p_0}{3 + 2p_0} \cos^2 \gamma \exp(2i\phi), \quad (\text{III.18})$$

which is independent of the density field n_H . Also, we choose to study $\exp(2i\psi)$ maps rather than ψ maps to avoid an unnecessary complication of the statistics due to the nonlinear action of the atan2 function as well as non-physical spatial discontinuities in ψ maps where the polarization angle approaches $\pm\pi/2$. Moreover, we expect the statistical properties of $\exp(2i\psi)$ to be easier to compare to those of $\tilde{Q} + i\tilde{U}$ since $\tilde{Q} + i\tilde{U} \approx p \exp(2i\psi)$ when $P \ll I$ (which is typically the case).

We emphasize that using the complex variable $Q + iU$ instead of Q and U separately is not just a mathematical convenience. Indeed, we have seen in Sect. I.2.1 that $Q + iU$ is a spin-2 quantity, so that its global phase is directly related to the reference frame in which γ and ϕ are defined. The complex variable $Q + iU$ is thus more apt to represent linear polarization than Q and U separately (see e.g. [184]). Moreover, we will see in the next subsection that it will ease the definition of a statistical description of polarization maps that is independent from this choice of reference frame.

c) Statistical isotropy/anisotropy of the maps

We point out that the initial conditions of the MHD simulation are anisotropic because of the initial direction of the uniform magnetic field. This anisotropy remains once the simulation has reached a steady state, due to magnetic flux conservation, and the value of the mean magnetic field is $\bar{\mathbf{B}} \approx \mathbf{B}_0$. This privileged direction in the simulation has an impact on the properties of isotropy of the derived maps. Indeed, when integrating along the z -axis, the direction of integration is orthogonal to the direction of the mean magnetic field $\bar{\mathbf{B}}$ in the simulation, so that we expect these derived maps to be statistically anisotropic. On the contrary, when integrating along the x -axis, which is the direction of the mean magnetic field, we expect statistically isotropic maps. We refer to the first set of maps with \perp , and to the second one with \parallel . For instance, we call p_\perp *data set* the set of 14 maps of polarization fraction p computed from the 14 maps I , Q , and U for which the z -axis was the axis of integration.

III.2.2 Presentation of the data sets

We make use of eight data sets, each one comprising 14 statistically independent maps. The various maps presented in the last subsection define 6 data sets: $\exp(2i\psi)_\perp$, p_\perp , $\tilde{Q}_\perp + i\tilde{U}_\perp$, $\exp(2i\psi)_\parallel$, p_\parallel , and $\tilde{Q}_\parallel + i\tilde{U}_\parallel$. In addition, we also analyze phase randomized data sets $R[\tilde{Q}_\perp + i\tilde{U}_\perp]$ and $R[\tilde{Q}_\parallel + i\tilde{U}_\parallel]$ built respectively from the $\tilde{Q}_\perp + i\tilde{U}_\perp$ and $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data sets.

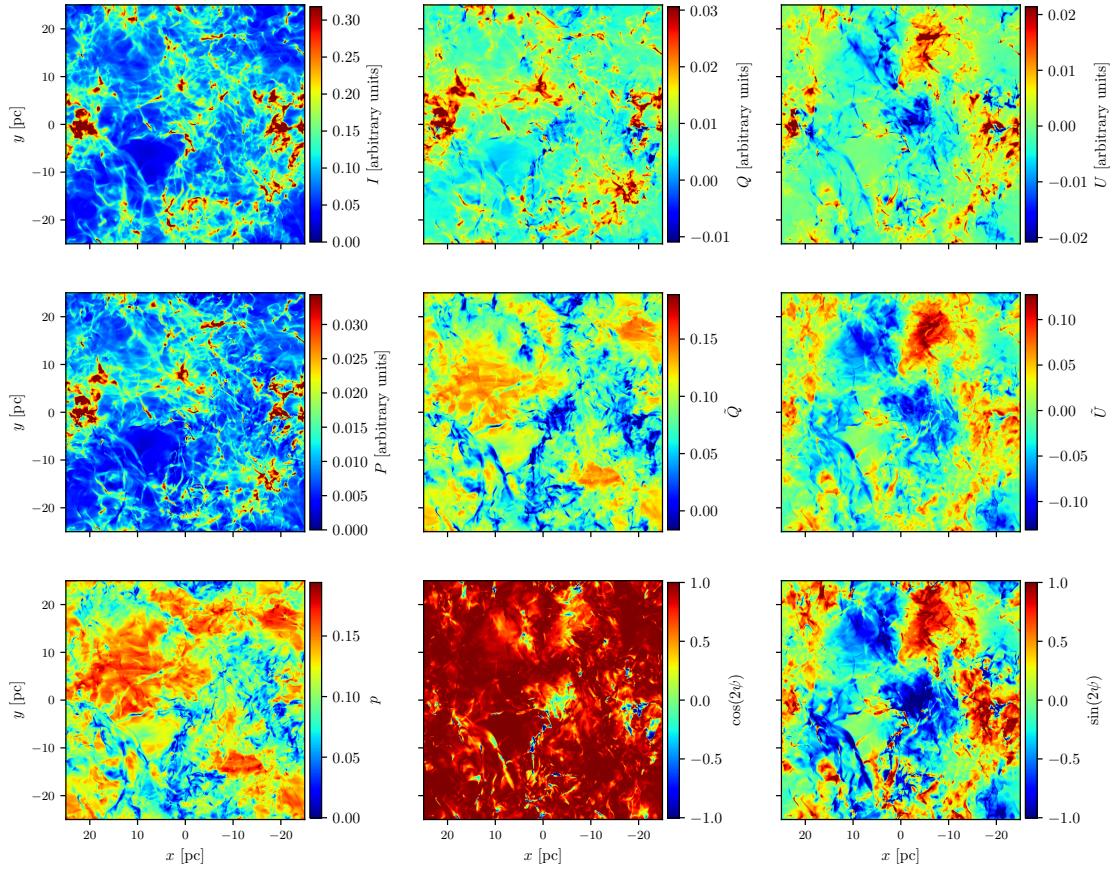


Fig. III.3 Examples of I_{\perp} , Q_{\perp} , U_{\perp} , P_{\perp} , \tilde{Q}_{\perp} , \tilde{U}_{\perp} , p_{\perp} , $\cos(2\psi)_{\perp}$, and $\sin(2\psi)_{\perp}$ maps (from top to bottom and left to right) that are built from a given snapshot of the MHD simulation.

The phase randomization operator $R[\cdot]$ acts on a map by randomizing the phases in Fourier space, meaning that the new phases are drawn independently for each mode from a uniform distribution on $[0, 2\pi)$. On the contrary, the moduli of the Fourier coefficients are retained during this operation so that the power spectrum is unchanged. The phase information of an image is tightly bound to its structure [185, 186] and the main effect of the $R[\cdot]$ operator is to severely damage the structure of the image. We use phase randomization as a simple way to approximate Gaussianization⁶. We have checked a posteriori that a standard Gaussian random field generation (see Appendix A) does not impact the analysis. In practice $R[\tilde{Q}_{\perp} + i\tilde{U}_{\perp}]$ (respectively $R[\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}]$) refers to the set of 14 maps produced by randomizing separately \tilde{Q} and \tilde{U} maps from the $\tilde{Q}_{\perp} + i\tilde{U}_{\perp}$ data set (respectively $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$). Technical details and illustrations about phase randomization can be found in Appendix A.

Figure III.3 shows examples of maps for the I_{\perp} , Q_{\perp} , U_{\perp} , P_{\perp} , \tilde{Q}_{\perp} , \tilde{U}_{\perp} , p_{\perp} , $\cos(2\psi)_{\perp}$,

⁶Stationary Gaussian random fields (GRFs) do have uniformly distributed phases on $[0, 2\pi)$ but this property alone does not define them (see [187] and Appendix A).

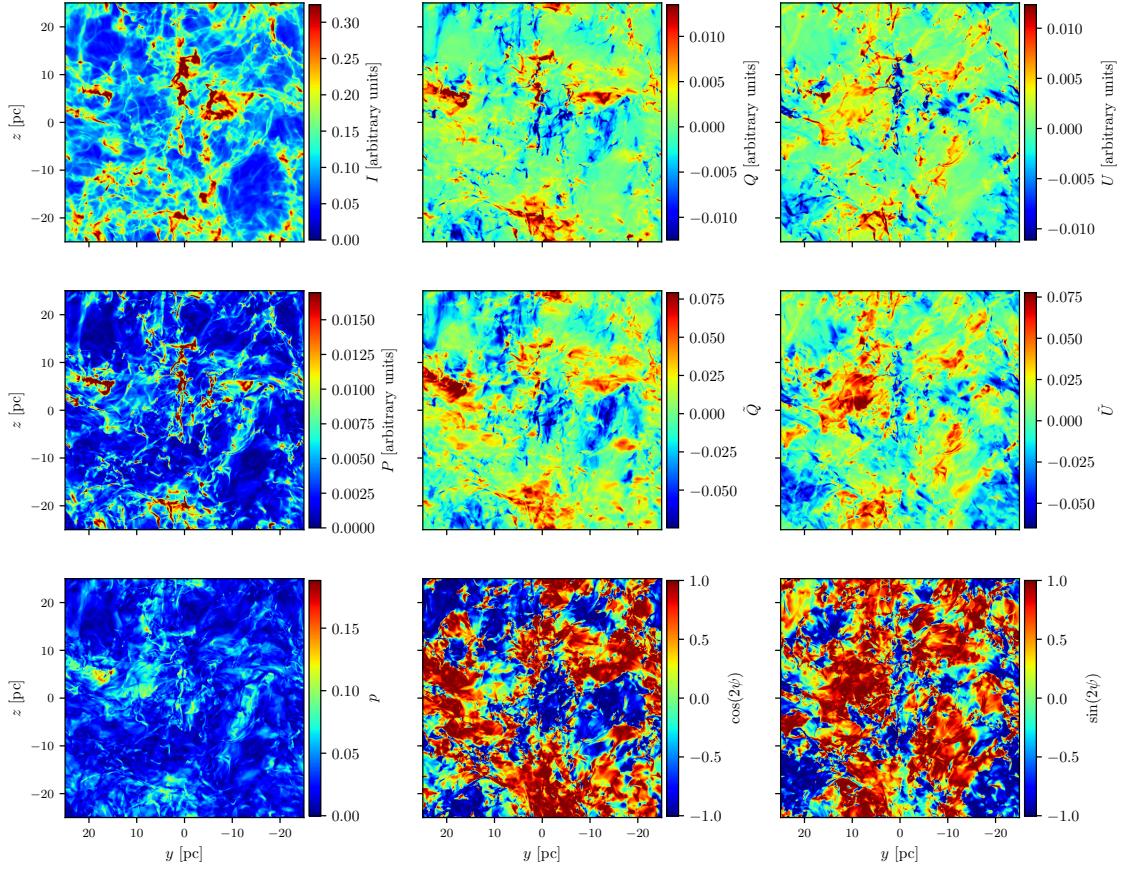


Fig. III.4 Same as Fig. III.3 but for the I_{\parallel} , Q_{\parallel} , U_{\parallel} , P_{\parallel} , \tilde{Q}_{\parallel} , \tilde{U}_{\parallel} , p_{\parallel} , $\cos(2\psi)_{\parallel}$, and $\sin(2\psi)_{\parallel}$ data sets.

and $\sin(2\psi)_{\perp}$ data sets, while Fig. III.4 shows examples of maps for the corresponding \parallel data sets. These figures show maps that all rely on the same snapshot of the MHD simulation. We draw the attention of the reader to a few points. First, we clearly see filamentary patterns on these maps that will demand a statistical description involving higher orders statistics compared to simple power spectra. We also note that P is an order of magnitude lower than the intensity I (this is due to the value of the polarization fraction parameter p_0). Hence we have approximately $\tilde{Q} + i\tilde{U} \approx (Q + iU)/I$ so that p roughly behaves as the modulus of $\tilde{Q} + i\tilde{U}$, and 2ψ roughly behaves as its complex argument. Next, we note that the magnitude of p_{\parallel} is unsurprisingly much lower than that of p_{\perp} as the direction of the projection of the local magnetic field on the plane of the sky is much less coherent along the line of sight when the line of sight corresponds to the direction of the mean magnetic field. Finally, we see on the $\cos(2\psi)_{\perp}$ map that the anisotropy of the magnetic field in the simulation leads to values that are concentrated close to 1.

j	0	1	2	3	4	5	6
$\tilde{\lambda}$ [pixels]	2.67	5.33	10.7	21.3	42.7	85.3	171
λ [pc]	0.26	0.52	1.04	2.08	4.17	8.33	16.7

Table III.2 – Correspondence between scale index j and related wavelengths on simulated maps, both in pixel units and dimensional units. Those wavelengths $\tilde{\lambda}$ come from the definition of dilated Morlet wavelets $\psi_{j,\theta}$ (for more details, see [4]).

III.2.3 Regularity in the WST coefficients

We now focus on the WST coefficients associated with the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set, but the following reasoning remains valid similarly for the other data sets.

For the 14 maps of the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set, we compute the WST coefficients with $J = 7$ and $L = 8$ (see Eqs. (III.9)-(III.11)). J could have been fixed to a higher value for 512×512 maps, but because of our limited number of maps we chose to restrict our statistical description to scales for which we have a sufficient number of modes for reliable estimations. We give in Table III.2 the correspondence between the scale index j and the central wavelength of the related dilated Morlet wavelet both in pixel units and in dimensional units (related to the MHD simulation). These J and L values lead to 2801 coefficients per map (see Table III.1). We define mean \bar{S}_0 , \bar{S}_1 and \bar{S}_2 coefficients as means of the WST coefficients over the 14 maps and we also compute the standard deviation of the mean for each of these coefficients assuming that the maps are statistically independent.

Figures III.5 and III.6 represent (in blue) respectively $\bar{S}_1(j_1, \theta_1)$ coefficients and a representative subsample of $\bar{S}_2(j_1, \theta_1, j_2, \theta_2)$ coefficients (for $j_1 = 1$) on a logarithmic scale, for $\tilde{Q}_\perp + i\tilde{U}_\perp$. In both figures, we represent the multivariate functions \bar{S}_1 and \bar{S}_2 in a lexicographical order for the (j_1, θ_1) and $(j_1, \theta_1, j_2, \theta_2)$ variables, respectively. Vertical gray lines help to mark increments of these variables.

In Fig. III.5, we see that for each scale j_1 a smooth pattern with respect to the angular variable θ_1 emerges. This smoothness is not surprising as it reflects the regularity of the underlying physical processes. Much more surprising is the apparent π -periodicity of these smooth patterns for every j_1 . \bar{S}_1 coefficients are naturally 2π -periodic with respect to the angular variable θ_1 , but here this π -periodicity with respect to the angular variable θ_1 was unexpected. Let us rephrase this result by noticing that:

$$|(\tilde{Q} + i\tilde{U}) \star \psi_{j,\theta}| = |[(\tilde{Q} + i\tilde{U}) \star \psi_{j,\theta}]^*| = |(\tilde{Q} - i\tilde{U}) \star \psi_{j,\theta+\pi}|, \quad (\text{III.19})$$

as Morlet wavelets verify $\psi_{j,\theta}^* = \psi_{j,\theta+\pi}$. Therefore, a π -periodicity of \bar{S}_1 coefficients with respect to θ_1 is equivalent to the fact that $\tilde{Q} + i\tilde{U}$ and $\tilde{Q} - i\tilde{U}$ share the same WST first order statistics.

We can identify the same kind of regularity properties for \bar{S}_2 coefficients in Fig. III.6. However, since \bar{S}_2 coefficients depend on four variables, it is hard to clearly identify the extent of this angular regularity on this figure. In order to better exhibit it, we show in Fig. III.7 \bar{S}_2 coefficients as a surface parametrized by θ_1 and θ_2 for arbitrary fixed values of j_1 and j_2 . Smooth angular patterns are obvious. In particular we see that $\theta_1 - \theta_2 = c$

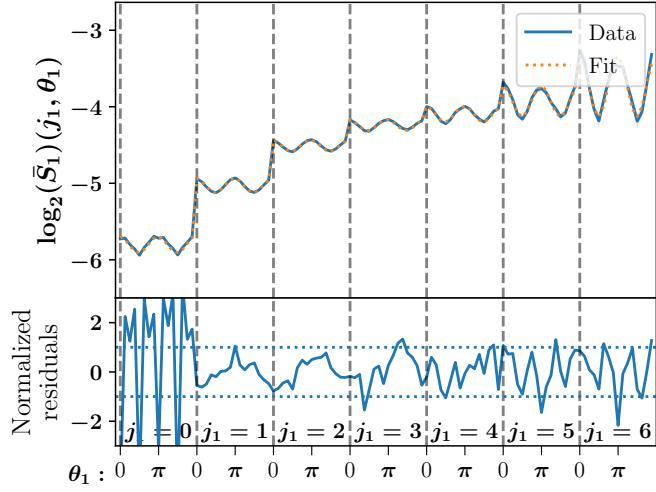


Fig. III.5 WST coefficients $\bar{S}_1(j_1, \theta_1)$ on a logarithmic scale for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set, presented in a lexicographical order on (j_1, θ_1) . Vertical dashed lines delimit distinct j_1 values. The top panel shows the original data (solid lines) and the RWST fit (dotted lines) corresponding to the model of Eq. (III.20), while the bottom panel shows the normalized residuals of the fit.

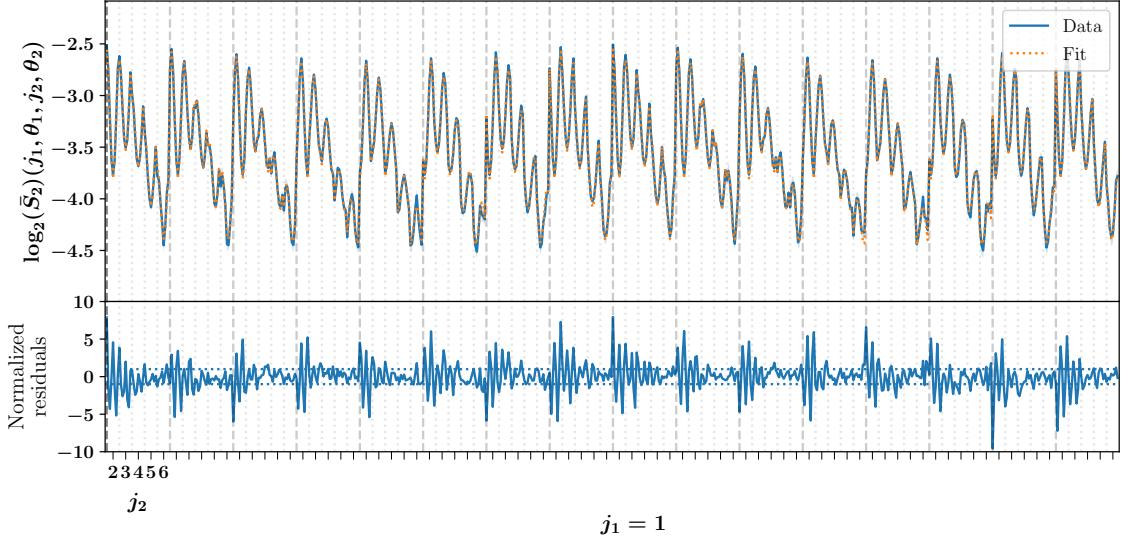


Fig. III.6 $j_1 = 1$ selection of WST coefficients $\bar{S}_2(j_1, \theta_1, j_2, \theta_2)$ on a logarithmic scale for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set presented in a lexicographical order on $(j_1, \theta_1, j_2, \theta_2)$. This specific selection of coefficients is arbitrary, and we find similar results for other scales and the other data sets. Vertical dashed and dotted lines delimit distinct θ_1 and j_2 values, respectively. The top panel shows the original data (solid lines) and the RWST fit (dotted lines) corresponding to the model of Eq. (III.21), while the bottom panel shows the normalized residuals of the fit.

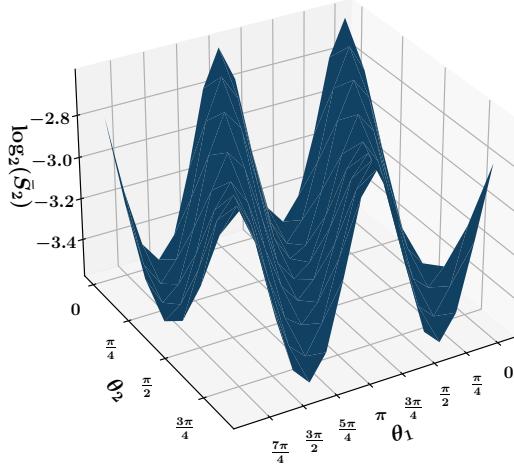


Fig. III.7 Surface representation of WST coefficients $\bar{S}_2(j_1 = 1, \theta_1, j_2 = 3, \theta_2)$ for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set as a function of θ_1 and θ_2 only.

cuts of this surface for arbitrary constants c give roughly constant \bar{S}_2 coefficients. The angular modulation thus seems to depend on the $\theta_2 - \theta_1$ variable only for this data set.

III.2.4 Simplification of the WST description with the reduced WST

The smooth periodic patterns identified in the $\tilde{Q}_\perp + i\tilde{U}_\perp$ WST coefficients suggest that a simplification of the WST statistical description is possible through an adequate modeling of its angular dependence. The purpose of such a modeling of the WST coefficients is twofold: 1) lowering the dimensionality of the statistical description of our data, and 2) providing an interpretable representation of this angular dependence in terms of isotropic and anisotropic properties of the data.

a) The RWST model

We model the regularity of WST coefficients with respect to angular variables θ_1 and θ_2 using the RWST model introduced in [167]. It is remarkable that a model developed for total intensity maps may be applied to $\tilde{Q} + i\tilde{U}$ complex Stokes maps of polarized thermal emission of dust without any modification. This highlights the generality of such an angular modeling of the WST coefficients for astrophysical maps, and this model may surely be extended to other kinds of complex-valued fields in physics.

We now briefly recall the RWST model, and refer the reader to [167] for more details.

In the RWST model, the \bar{S}_1 coefficients are written as:

$$\log_2 [\bar{S}_1 (j_1, \theta_1)] = \hat{S}_1^{\text{iso}}(j_1) + \hat{S}_1^{\text{aniso}}(j_1) \cos \left(2 [\theta_1 - \theta^{\text{ref},1}(j_1)] \right), \quad (\text{III.20})$$

where $\hat{S}_1^{\text{iso}}(j_1)$, $\hat{S}_1^{\text{aniso}}(j_1)$, and $\theta^{\text{ref},1}(j_1)$ are parameters depending on the scale j_1 . This model thus depends on $3 \times J$ parameters. We also enforce $\hat{S}_1^{\text{aniso}}(j_1) \geq 0$ in order to lift a degeneracy between the $\hat{S}_1^{\text{aniso}}(j_1)$ and $\theta^{\text{ref},1}$ parameters. $\hat{S}_1^{\text{iso}}(j_1)$ quantifies the isotropic

component of the data fluctuations at a scale j_1 , while $\hat{S}_1^{\text{aniso}}(j_1)$ corresponds to the degree of angular modulation of the coefficients at scale j_1 , introducing a reference angle $\theta^{\text{ref},1}(j_1)$ that defines a privileged direction in the maps⁷.

Similarly, the RWST model for \bar{S}_2 coefficients reads:

$$\begin{aligned} \log_2 [\bar{S}_2(j_1, \theta_1, j_2, \theta_2)] &= \hat{S}_2^{\text{iso},1}(j_1, j_2) \\ &+ \hat{S}_2^{\text{iso},2}(j_1, j_2) \cos(2[\theta_1 - \theta_2]) \\ &+ \hat{S}_2^{\text{aniso},1}(j_1, j_2) \cos(2[\theta_1 - \theta^{\text{ref},2}(j_1, j_2)]) \\ &+ \hat{S}_2^{\text{aniso},2}(j_1, j_2) \cos(2[\theta_2 - \theta^{\text{ref},2}(j_1, j_2)]), \end{aligned} \quad (\text{III.21})$$

where $\hat{S}_2^{\text{iso},1}(j_1, j_2)$, $\hat{S}_2^{\text{iso},2}(j_1, j_2)$, $\hat{S}_2^{\text{aniso},1}(j_1, j_2)$, $\hat{S}_2^{\text{aniso},2}(j_1, j_2)$, and $\theta^{\text{ref},2}(j_1, j_2)$ are the parameters of this angular model for each pair of scales (j_1, j_2) . As we have $j_2 > j_1$, we end up with $5 \times J \times (J - 1)/2$ parameters for this model. Here again we make sure that $\hat{S}_2^{\text{aniso},1}(j_1, j_2) \geq 0$ to avoid any parameter degeneracy. $\hat{S}_2^{\text{iso},1}$ measures the overall amplitude of coupling between the scales j_1 and j_2 . $\hat{S}_2^{\text{iso},2}$ represents the amplitude of the modulation due to the relative orientation of the wavelets ψ_{j_1, θ_1} and ψ_{j_2, θ_2} , and we interpret it as a signature of filamentary structures at a given scale. Indeed for an oriented filamentary structure, we expect the \bar{S}_2 coefficients to peak when $\theta_2 = \theta_1$ and to reach a minimum when $\theta_2 = \theta_1 + \pi/2$. Finally, $\hat{S}_2^{\text{aniso},1}$ and $\hat{S}_2^{\text{aniso},2}$ are measures of the anisotropic properties of the data in second order WST coefficients, here decoupling θ_1 and θ_2 contributions.

b) Fourier perspective

Such a model is the logical consequence of the angular regularity of the coefficients with a few additional assumptions. Indeed, let us rephrase the RWST model in terms of truncated Fourier series. We consider the $\log_2(\bar{S}_1)$ coefficients at a given scale j_1 and write $f_{j_1}(\theta_1)$ the corresponding angular model⁸. If one assumes that there are no more than one privileged direction $\theta^{\text{ref},1}(j_1)$ in the maps we are dealing with, we can write generally $f_{j_1}(\theta_1)$ as a Fourier series expansion using the natural 2π -periodicity of this function:

$$\begin{aligned} f_{j_1}(\theta_1) &= a_0(j_1) + \sum_{k=1}^{+\infty} [a_k(j_1) \cos(k[\theta_1 - \theta^{\text{ref},1}(j_1)]) + \\ &\quad b_k(j_1) \sin(k[\theta_1 - \theta^{\text{ref},1}(j_1)])]. \end{aligned} \quad (\text{III.22})$$

Assuming a mirror symmetry with respect to the potential reference direction, we expect $\{b_k\}$ coefficients to vanish (which is the case in practice for our data). Adding to that the π -periodicity identified previously the angular model reduces to:

$$f_{j_1}(\theta_1) = a_0(j_1) + \sum_{p=1}^{+\infty} a_{2p}(j_1) \cos(2p[\theta_1 - \theta^{\text{ref},1}(j_1)]). \quad (\text{III.23})$$

⁷We note that Eq. (III.20) defines an angular model of the logarithm of the WST coefficients that turns angular modulations of the WST coefficients into additive corrections $\hat{S}_1^{\text{aniso}}(j_1) \cos(2[\theta_1 - \theta^{\text{ref},1}(j_1)])$ to the isotropic amplitude of fluctuations $\hat{S}_1^{\text{iso}}(j_1)$. We use a base 2 logarithm to be consistent with the base 2 definition of scales j_1 and j_2 .

⁸The same kind of reasoning holds for $\log_2(\bar{S}_2)$ coefficients.

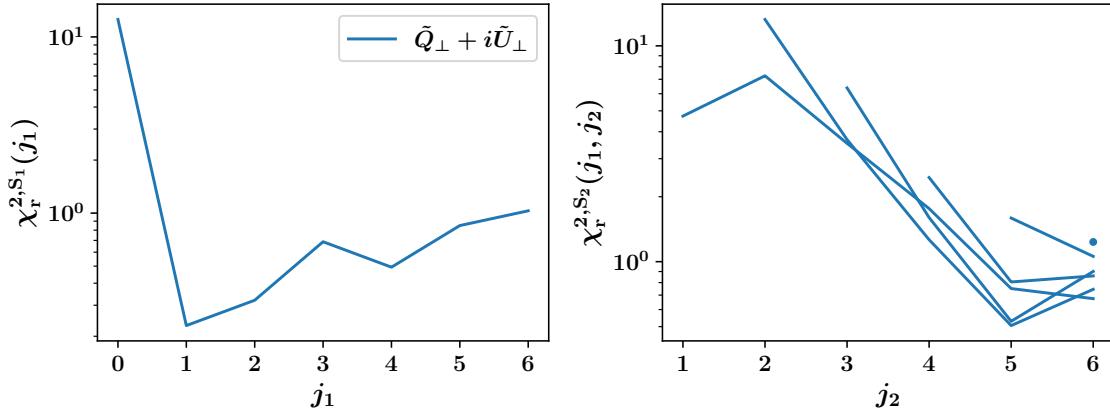


Fig. III.8 Reduced chi square $\chi_r^{2,S_1}(j_1)$ (left) and $\chi_r^{2,S_2}(j_1, j_2)$ (right) associated with the RWST fit of the WST coefficients (see Eqs. (III.20) and (III.21)) for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set. In the right panel, each curve corresponds to a fixed j_1 value with j_2 varying from $j_1 + 1$ to $J - 1 = 6$. For $j_1 = J - 2$, the curve is reduced to a single dot on the figure. We use logarithmic scales for better visibility.

Finally, the smoothness of the patterns presented previously implies a fast decrease of the amplitudes of the harmonics. Truncating the expansion after the second term and writing $a_0(j_1) = \hat{S}_1^{\text{iso}}(j_1)$ and $a_2(j_1) = \hat{S}_1^{\text{aniso}}(j_1)$ we end up with the RWST model of Eq. (III.20).

c) Fit of the model

In practice, for a given data set, this RWST model of the angular dependence is independently fitted to the first order WST coefficients for every scale j_1 , and to the second order WST coefficients for every pair of scales (j_1, j_2) (with $j_2 > j_1$). The accuracy of these multiple fits is quantified with χ^2 statistics as described in [167]. Since it is not possible to properly estimate the full covariance matrix with only 14 samples per coefficient, the uncertainties affecting the WST coefficients for a given data set are simply estimated from the sample variance across the various simulation snapshots.

However an important correlation of the first order WST coefficients across angles for each scale j_1 needs to be taken into account to properly estimate statistical uncertainties. For each sample we compute a mean coefficient across angles for a given scale and subtract this mean before computing the statistical uncertainties⁹. This mitigates most of the correlation between WST coefficients at the same scale j_1 .

Figure III.8 shows the χ^2 values for both $\log_2(\bar{S}_1)$ and $\log_2(\bar{S}_2)$ fits for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set. The χ^2 values are globally close to 1 except at low j_1 for χ_r^{2,S_1} and at low $j_2 - j_1$ for χ_r^{2,S_2} . This deterioration of the goodnesses of the fits is due to a pixellization effect at small scales, and may be fixed by adding adequate lattice terms in the RWST model as described in the next section. Finally, the same RWST model applies to the other data sets used in this work, including $\exp(2i\psi)$ and p data sets, and we get for all of them similar χ^2 values.

In Figs. III.5 and III.6 we show the RWST fit overplotted on a selection of first and

⁹This decreases the effective number of degrees of freedom by one when computing χ^2 values.

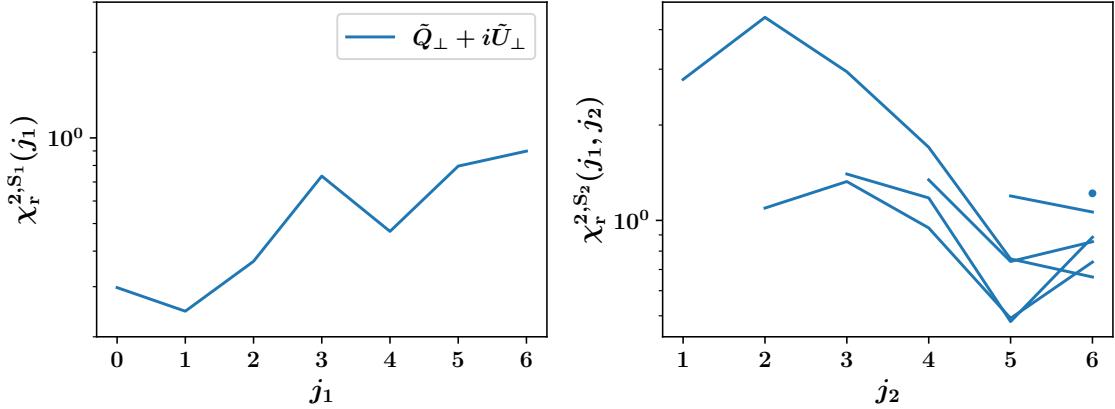


Fig. III.9 Same as Fig. III.8 for the RWST fits of the WST coefficients that take into account lattice terms (see Eqs. (III.24) and (III.25) for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set.

second order WST coefficients, and also show the corresponding normalized residuals. The curves are globally in good agreement and the flaws of the RWST model due to numerical effects at low j_1 and $j_2 - j_1$ appear as clear patterns in the residuals.

In the end, the RWST coefficients define statistical descriptions of the data sets in terms of simple considerations of isotropy and anisotropy with respect to a reference direction. Furthermore, these statistical descriptions exhibit much lower dimensionality, with a total of 127 coefficients (including the \bar{S}_0 coefficient in the description) compared to the original 2801 WST coefficients in the case of complex variables $\tilde{Q} + i\tilde{U}$ and $\exp(2i\psi)$ ¹⁰, thus providing a very large compression of the statistical information contained in the WST coefficients.

d) Extension of the model

Following the appendix C of [167], we may enhance the RWST model defined in Eq. (III.20) and (III.21) by adding so-called *lattice* terms related to pixellization effects at small scales for first order coefficients and a supplementary harmonic of the angular modulation of the second order WST coefficients. This enhanced RWST model of the angular dependency of the WST coefficients becomes, for \bar{S}_1 coefficients:

$$\begin{aligned} \log_2 [\bar{S}_1(j_1, \theta_1)] = & \hat{S}_1^{\text{iso}}(j_1) \\ & + \hat{S}_1^{\text{aniso}}(j_1) \cos(2[\theta_1 - \theta^{\text{ref},1}(j_1)]) \\ & + \hat{S}_1^{\text{lat},1}(j_1) \cos(4\theta_1) + \hat{S}_1^{\text{lat},2}(j_1) \cos(8\theta_1), \end{aligned} \quad (\text{III.24})$$

where the additional lattice terms $\hat{S}_1^{\text{lat},1}$ and $\hat{S}_1^{\text{lat},2}$ quantify angular modulations that are respectively $\pi/2$ and $\pi/4$ -periodic and aligned with the main directions of the lattice.

¹⁰For p maps, the WST descriptions consist of 1401 coefficients.

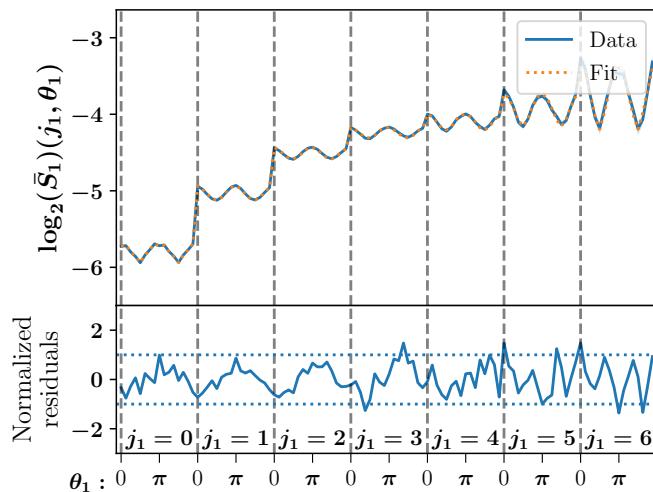


Fig. III.10 Same as Fig. III.5 but for the RWST fit corresponding to the model of Eq. (III.24) including lattice terms.

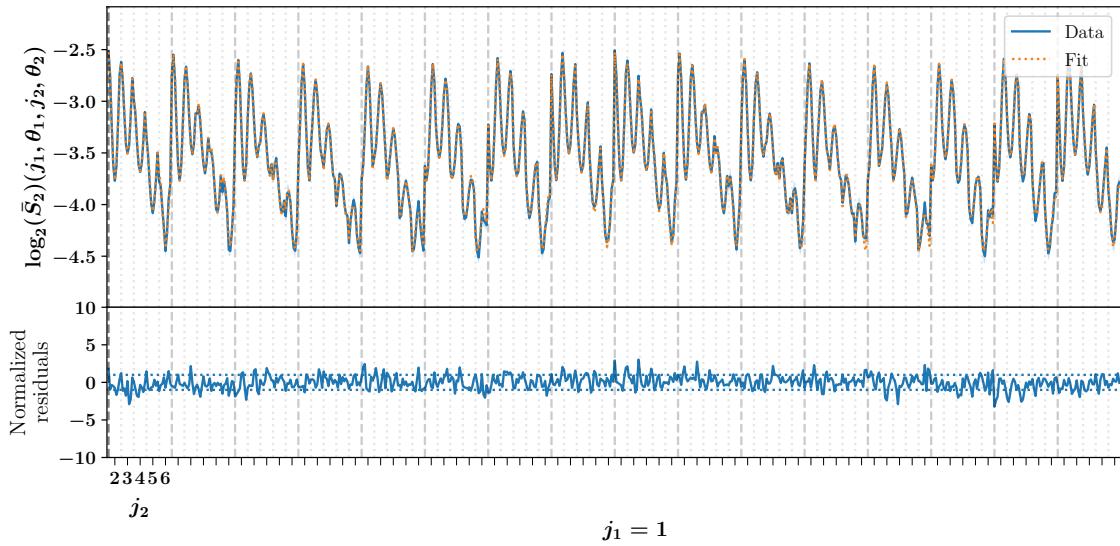


Fig. III.11 Same as Fig. III.6 but for the RWST fit corresponding to the model of Eq. (III.25) including lattice terms.

Similarly the enhanced RWST model of \bar{S}_2 coefficients is:

$$\begin{aligned} \log_2 [\bar{S}_2(j_1, \theta_1, j_2, \theta_2)] &= \hat{S}_2^{\text{iso},1}(j_1, j_2) \\ &+ \hat{S}_2^{\text{iso},2}(j_1, j_2) \cos(2[\theta_1 - \theta_2]) \\ &+ \hat{S}_2^{\text{iso},3}(j_1, j_2) \cos(4[\theta_1 - \theta_2]) \\ &+ \hat{S}_2^{\text{aniso},1}(j_1, j_2) \cos(2[\theta_1 - \theta^{\text{ref},2}(j_1, j_2)]) \\ &+ \hat{S}_2^{\text{aniso},2}(j_1, j_2) \cos(2[\theta_2 - \theta^{\text{ref},2}(j_1, j_2)]), \end{aligned} \quad (\text{III.25})$$

where the additional term $\hat{S}_2^{\text{iso},3}$ measures the amplitude of an additional harmonic of the $\theta_1 - \theta_2$ angular modulation that is $\pi/2$ -periodic.

These additional terms do not affect the values of the RWST coefficients discussed above and significantly improve χ_r^{2,S_1} and χ_r^{2,S_2} at small scales as shown in Fig. III.9. These curves are to be compared to those of Fig. III.8. Figures III.10 and III.11 show greatly improved normalized residuals compared to the ones shown in Figs. III.5 and III.6. In particular at $j_1 = 0$ we no longer observe the strong angular pattern seen in Fig. III.5.

III.3 RWST data analysis and interpretation

In this section we analyze and interpret the RWST statistical descriptions for the 8 data sets built from the same MHD simulation. We relate the coefficients of these descriptions to the physical properties of the simulation.

III.3.1 Isotropic fluctuations in first order coefficients \hat{S}_1^{iso}

Figure III.12 presents $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ coefficients as a function of scale for the $\tilde{Q}_\perp + i\tilde{U}_\perp$, $\tilde{Q}_\parallel + i\tilde{U}_\parallel$, $R[\tilde{Q}_\perp + i\tilde{U}_\perp]$, and $R[\tilde{Q}_\parallel + i\tilde{U}_\parallel]$ data sets. We note that the $\log_2(\bar{S}_0)$ term cancels the normalization of the \bar{S}_1 WST coefficients defined in Eq. (III.10) in order to analyze the statistics of $\tilde{Q} + i\tilde{U}$ (and not the statistics of $(\tilde{Q} + i\tilde{U})/\langle|\tilde{Q} + i\tilde{U}|\rangle$). We see that $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ levels, which measure an amplitude of fluctuations of the signal per scale, are higher for $\tilde{Q}_\perp + i\tilde{U}_\perp$ and $R[\tilde{Q}_\perp + i\tilde{U}_\perp]$ than for the corresponding \parallel data sets. This shows that the amplitudes differ depending on the orientation of the mean magnetic field with respect to the line of sight. We have more fluctuations within $\tilde{Q} + i\tilde{U}$ maps when the mean magnetic field is in the plane of the sky.

The differences between the $\tilde{Q} + i\tilde{U}$ data set and its corresponding randomized counterpart $R[\tilde{Q} + i\tilde{U}]$ (in both \perp and \parallel cases) illustrate the difference between \bar{S}_1 coefficients and the power spectrum. Indeed, $\tilde{Q} + i\tilde{U}$ and $R[\tilde{Q} + i\tilde{U}]$ maps share the same power spectrum but have different \bar{S}_1 coefficients. The fact that the $R[\cdot]$ operator increases the \bar{S}_1 values leaving the power spectrum unchanged shows that it reduces the sparsity of these maps (see discussion in Sect. III.1.2). This feature underlines the non-Gaussianity of the $\tilde{Q} + i\tilde{U}$ data set and we therefore expect its second order coefficients to be higher compared to those of the corresponding randomized data set (see Sect. III.3.3). We note that these differences wear off at large scales. We interpret this as statistical evidence that the non-Gaussianity of $\tilde{Q} + i\tilde{U}$ maps decreases at large scales. This result may reflect

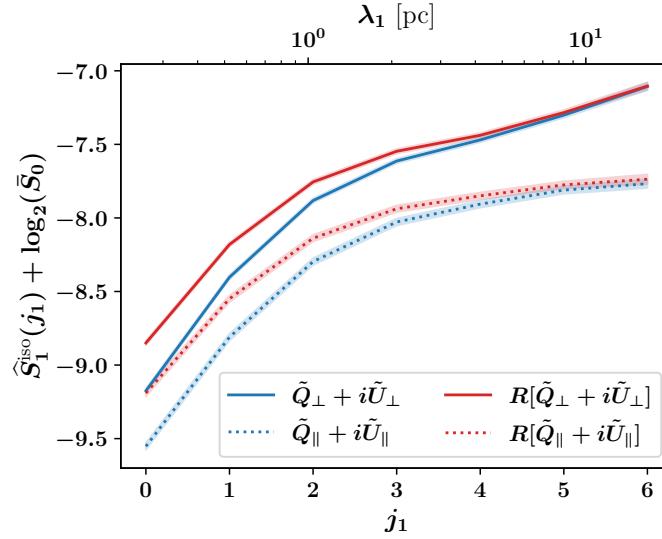


Fig. III.12 $\hat{S}_1^{\text{iso}}(j_1) + \log_2(\bar{S}_0)$ RWST coefficients for the $\tilde{Q}_{\perp} + i\tilde{U}_{\perp}$, $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$, $R[\tilde{Q}_{\perp} + i\tilde{U}_{\perp}]$, and $R[\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}]$ data sets. We use solid lines for \perp data sets and dotted lines for \parallel data sets. For reference, we show at the top of this figure the correspondence between j scale indices and the related wavelengths on the maps (in practice, we have $\lambda_1 = 2^{j_1+1}\pi/k_0$).

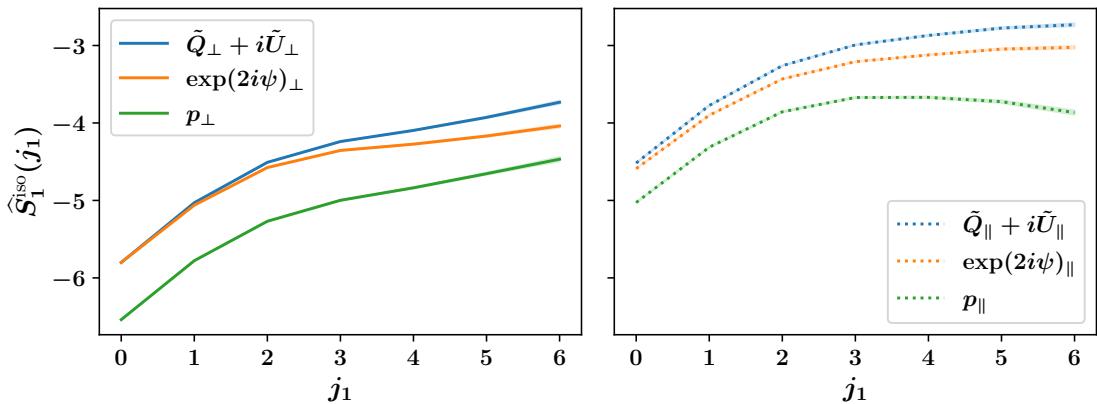


Fig. III.13 $\hat{S}_1^{\text{iso}}(j_1)$ RWST coefficients for $\tilde{Q} + i\tilde{U}$, $\exp(2i\psi)$ and p data sets in the \perp (left, solid lines) and \parallel (right, dotted lines) cases.

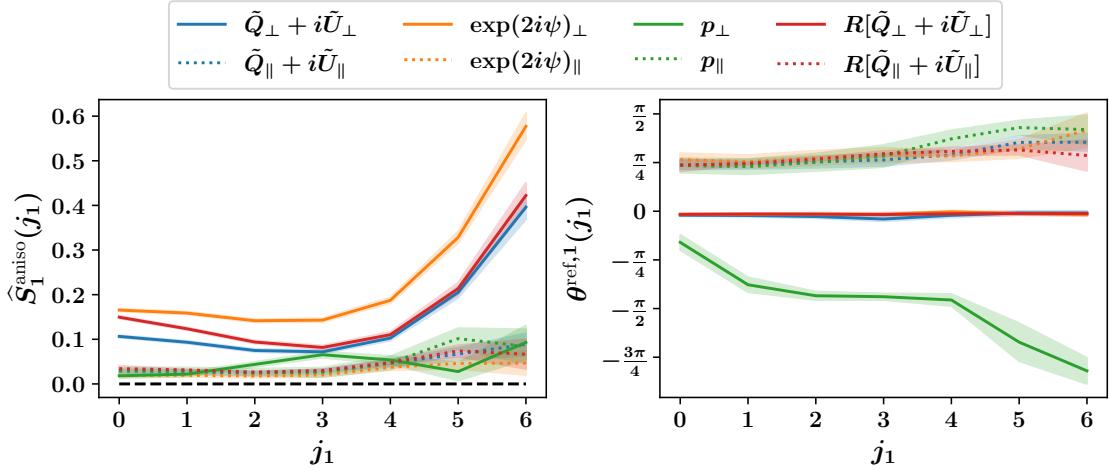


Fig. III.14 $\hat{S}_1^{\text{aniso}}(j_1)$ (left) and $\theta^{\text{ref},1}(j_1)$ (right) RWST coefficients for the $\tilde{Q} + i\tilde{U}$, $\exp(2i\psi)$, p , and $R[\tilde{Q} + i\tilde{U}]$ data sets in the \perp (solid lines) and \parallel (dotted lines) cases.

a characteristic of interstellar turbulence but could also follow from the fact that we start to probe the Gaussian distribution of the turbulent forcing of the simulation.

In Fig. III.13, we compare $\tilde{Q} + i\tilde{U}$ \hat{S}_1^{iso} coefficients to those of p and $\exp(2i\psi)$ for \parallel and \perp data sets. Since we have $\tilde{Q} + i\tilde{U} \approx p \exp(2i\psi)$ we would like to compare the relative contributions of \hat{S}_1^{iso} coefficients of $\tilde{Q} + i\tilde{U}$ maps, but this seems more complicated than expected as we found out that a proper comparison through \hat{S}_1^{iso} coefficients highly depends on the choice of normalization of the WST coefficients for p .

\hat{S}_1^{iso} coefficients roughly range from -4.5 to -3.0 for $\exp(2i\psi)_{\parallel}$ while they range from -5.8 to -4.2 for $\exp(2i\psi)_{\perp}$. These differences indicate larger fluctuations of the polarization angle when the mean magnetic field is along the line of sight compared to when the mean magnetic field is in the plane of the sky. Since the average polarization fraction p is lower when the mean magnetic field is along the line of sight ($\bar{S}_0 \approx 0.03$ for p_{\parallel} compared to $\bar{S}_0 \approx 0.1$ for p_{\perp}) this feature is consistent with the anti-correlation observed between the angle dispersion function \mathcal{S} and the polarization fraction p [103, 58, 188].

III.3.2 Anisotropic fluctuations in first order coefficients \hat{S}_1^{aniso}

\hat{S}_1^{aniso} coefficients measure the angular modulation of the first order WST coefficients. They are presented in Fig. III.14 for all data sets. We see that this anisotropy is much larger for \perp data sets than for \parallel ones, except for the p_{\perp} data set which has a rather low anisotropy level. These differences are not surprising as we expect a stronger anisotropy when the mean magnetic field is in the plane of the sky, while statistical isotropy is expected when integrating along the mean magnetic field. For larger scales, we see an increase of these coefficients for \perp data sets. This trend has already been noticed on observational data of the Polaris flare in total intensity [167] and deserves a closer examination.

As explained in Sect. III.2.1 the large scales of consecutive snapshots could be corre-

lated to some extent. To assess the potential impact of these correlations on our analysis, we have computed separate RWST statistics for three maps of the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set corresponding to snapshots that are sufficiently distant in time to be confident in their independence (we choose $6 \times \delta\tau$ instead of $\delta\tau$). The increase of \hat{S}_1^{aniso} coefficients at large scales remains significant for each map, which demonstrates that this trend is not a consequence of correlations among snapshots.

Reference angles $\theta^{\text{ref},1}$ also presented in Fig. III.14 show that the preferential direction identified for anisotropic \perp data sets (except for p_\perp) is the direction corresponding to $\theta^{\text{ref},1} = 0$ for all scales j_1 . Such a value of the reference angle indicates a statistical tendency for structures, including filaments, to be elongated vertically rather than horizontally, that is, along the y axis in Fig. III.3. This corresponds to an elongation which is orthogonal to the mean magnetic field. This result is to be compared in further works to the abundant literature on the relative orientation between magnetic fields and structures traced by interstellar dust (for a review, see [189]).

The reference angle found for \parallel data set is well defined and approximately equal to $\pi/4$ for all scales while the anisotropy levels are close to zero. These values of $\theta^{\text{ref},1}$ are surprising because we were not expecting any anisotropy for these data sets. By examining the RWST statistics separately for each map, for each of the corresponding data sets, we found out that these surprising values correspond to an intermittent rise of the anisotropy level that appears in a few consecutive snapshots of the simulation. In this case where the level and direction of anisotropy are not coherent over snapshots, the mean coefficient gives an incomplete view of the anisotropic properties of the simulation. Notably, even with significant levels of anisotropy per map, if the reference angles are incoherent between the maps, we expect the mean level of anisotropy to be small. We found out that this is what happens for the p_\perp data set, where \hat{S}_1^{aniso} coefficients are relatively small while $\theta^{\text{ref},1}$ coefficients clearly deviate from zero.

III.3.3 Second order coefficients and non-Gaussianity of the data

$R[\tilde{Q} + i\tilde{U}]$ maps are Gaussian approximations of $\tilde{Q} + i\tilde{U}$ maps, and we have already exhibited differences between these data sets in their first order RWST coefficients in Sect. III.3.1. Similarly, second order RWST coefficients show remarkable differences that highlight the non-Gaussianity of the $\tilde{Q}_\perp + i\tilde{U}_\perp$ and $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data sets. The two dominant second order RWST coefficients $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$ presented in Fig. III.15 display clearly distinct patterns between the original data sets and the randomized ones on the example of the \perp data sets. First, $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$ coefficients are globally smaller for the $R[\tilde{Q} + i\tilde{U}]$ data sets, which is in line with what we had foreseen in Sect. III.3.1. In addition, $\hat{S}_2^{\text{iso},1}$ coefficients for $R[\tilde{Q} + i\tilde{U}]$ show a scale invariance property: $\hat{S}_2^{\text{iso},1}$ coefficients only depend on the difference $j_2 - j_1$. We point out that these scale invariant patterns are signatures of self-similar Gaussian processes [174] and have already been observed for fractional Brownian motions processes in [167].

$\hat{S}_2^{\text{iso},2}$ coefficients also show two distinct trends: the coefficients quickly tend to zero when $j_2 - j_1$ increases for $R[\tilde{Q} + i\tilde{U}]$ while coefficients tend to strictly positive values for the largest $j_2 - j_1$ values for non randomized data sets $\tilde{Q} + i\tilde{U}$. This result is related to the filamentary structure of the non randomized maps, because a filamentary structure involves a modulation of the WST coefficients as a function of the angle difference $\theta_2 - \theta_1$.

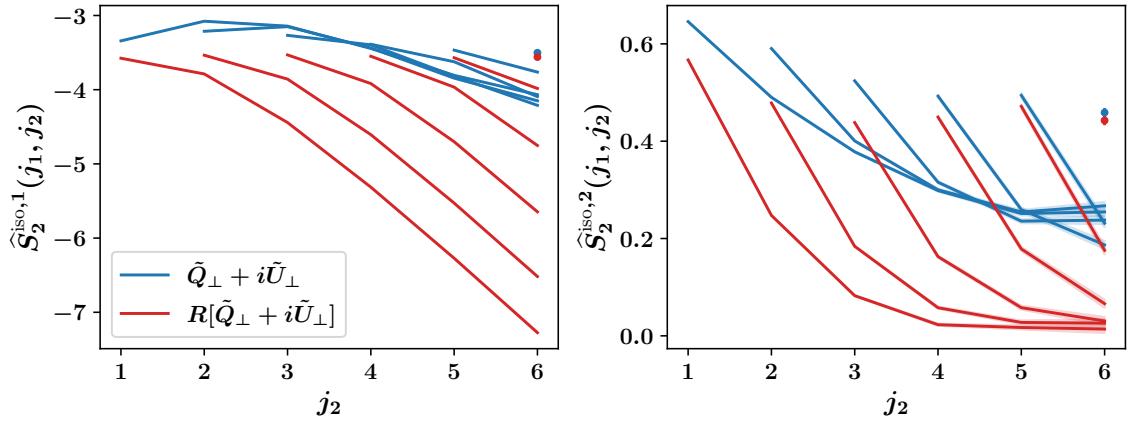


Fig. III.15 $\hat{S}_2^{\text{iso},1}(j_1, j_2)$ (left) and $\hat{S}_2^{\text{iso},2}(j_1, j_2)$ (right) RWST coefficients for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ and $R[\tilde{Q}_\perp + i\tilde{U}_\perp]$ data sets. Each curve corresponds to a fixed j_1 value with j_2 varying from $j_1 + 1$ to $J - 1 = 6$. For $j_1 = J - 2$, the curve is reduced to a single dot on the figure.

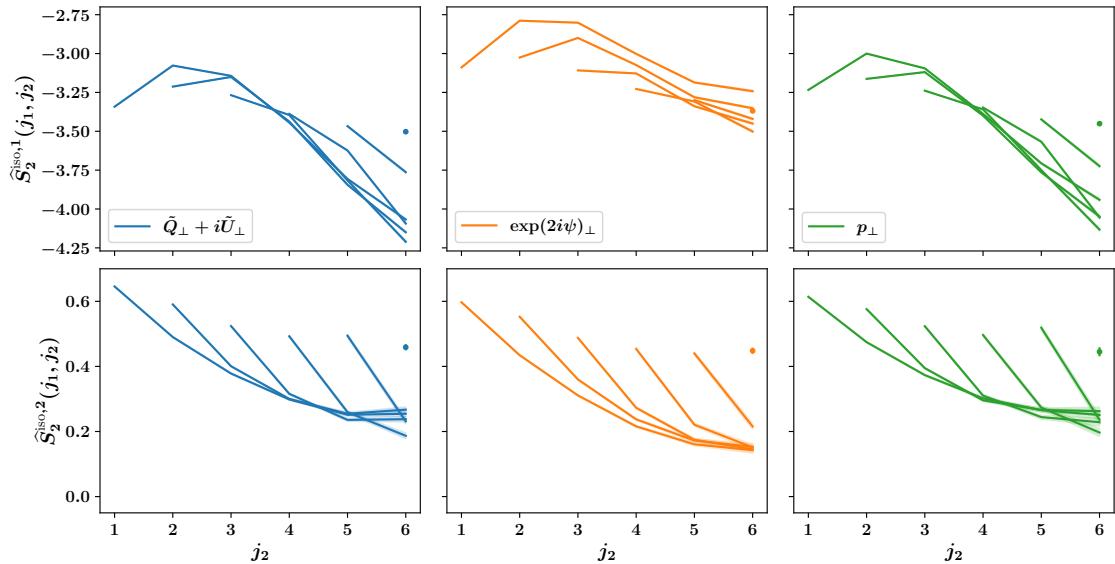


Fig. III.16 $\hat{S}_2^{\text{iso},1}(j_1, j_2)$ (top row) and $\hat{S}_2^{\text{iso},2}(j_1, j_2)$ (bottom row) RWST coefficients for the $\tilde{Q} + i\tilde{U}$, $\exp(2i\psi)$ and p data sets in the \perp case. Each curve corresponds to a fixed j_1 value with j_2 varying from $j_1 + 1$ to $J - 1 = 6$. For $j_1 = J - 2$, the curve is reduced to a single dot on the figure.

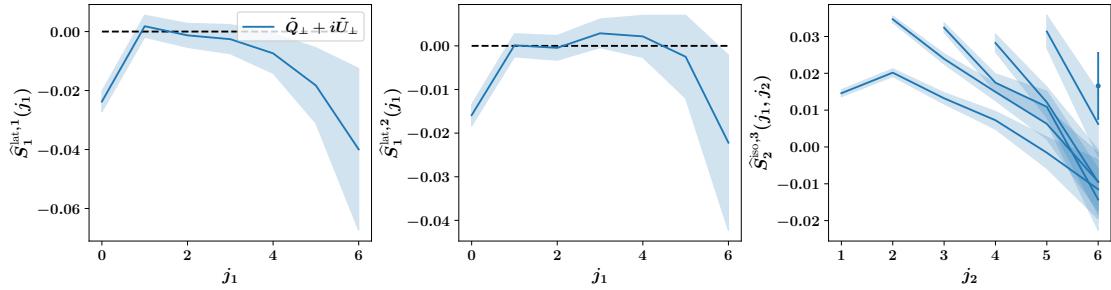


Fig. III.17 $\hat{S}_1^{\text{lat},1}(j_1)$ (left column), $\hat{S}_1^{\text{lat},2}(j_1)$ (middle column) and $\hat{S}_2^{\text{iso},3}(j_1, j_2)$ (right column) RWST coefficients for the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set. In the $\hat{S}_2^{\text{iso},3}$ plot, each curve corresponds to a fixed j_1 value while j_2 ranges from j_1+1 to $J-1=6$. For $j_1 = J-2$, the curve is reduced to a single dot on the figure.

We see the same trend in Fig. III.16 for the $\hat{S}_2^{\text{iso},2}$ coefficients of the other non randomized data sets which all present filamentary structures. Additionally, for all data sets and fixed j_1 values, $\hat{S}_2^{\text{iso},2}$ coefficients decrease as a function of j_2 . We also expect a signal due to the imprint of the wavelets in the $\hat{S}_2^{\text{iso},2}$ coefficients that would increase their values when j_2 is close to j_1+1 .

We see in Fig. III.15 that the differences between the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set and its randomized counterpart decrease for the highest j_1 values. Here again, this shows that the non-Gaussianity of $\tilde{Q} + i\tilde{U}$ maps decreases at large scales. This is consistent with what we have already pointed out in Sect. III.3.1 for \hat{S}_1^{iso} coefficients and we interpret this trend similarly.

Second order anisotropic coefficients $\hat{S}_2^{\text{aniso},1}$, $\hat{S}_2^{\text{aniso},2}$, and $\theta^{\text{ref},2}$ essentially show consistent results with first order anisotropic coefficients both in terms of amplitude and direction of anisotropy. They have generally smaller values compared to those of isotropic coefficients $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$. We do not show these coefficients here as we do not discuss them any further in this work.

III.3.4 Lattice terms

For completeness, we show in Fig. III.17 the additional RWST terms given for the example of the $\tilde{Q}_\perp + i\tilde{U}_\perp$ data set.

III.4 Conclusion and perspectives

In this chapter, we introduced the WST to analyze maps of polarized thermal emission from interstellar dust, using 512×512 pixels Stokes I , Q , and U maps built from a numerical simulation of MHD turbulence designed to reproduce typical properties of the diffuse ISM. To alleviate the fact that Stokes Q and U rely on the definition of an arbitrary reference frame, and to remove the zeroth-order impact of the matter distribution on their properties and thus focus on the statistics of the magnetic field, the WST was applied to complex Stokes maps $\tilde{Q} + i\tilde{U}$ that are normalized by $I + P$. To study the contributions of the polarization fraction p and of the polarization angle ψ to the statistical properties of these complex Stokes maps, we also applied the WST to the corresponding maps of p

and $\exp(2i\psi)$. We finally analyzed "Gaussianized" complex Stokes maps obtained after phase randomization.

The WST gives a low-variance statistical description of these complex and real maps through typically a few thousand coefficients indexed in terms of orientations and scales. These coefficients capture the power spectra of the maps and characterize couplings between oriented scales. WST coefficients for maps of $\tilde{Q} + i\tilde{U}$, p , and $\exp(2i\psi)$ present a striking regularity when taken as functions of the sole angular variables. This is very much in line with what we observed in [167] for column density and total intensity maps, and in fact the same functional form introduced in that paper can be used to fit the angular dependencies of the WST coefficients of polarization maps studied here, thus extending the RWST model introduced in [167]. The RWST yields a statistical description of polarization maps that quantifies their multiscale properties in terms of isotropic and anisotropic contributions, all the while requiring more than one order of magnitude fewer coefficients than the WST.

The RWST analysis allowed us to identify statistical characteristics that exhibit the dependence of the map structure on the orientation of the mean magnetic field and quantify the non-Gaussianity of data.

- The overall level of first order coefficients depends on the orientation of the mean magnetic field with respect to the line of sight. For $\tilde{Q} + i\tilde{U}$ maps, $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ coefficients are larger when the mean magnetic field is in the plane of the sky, while for $\exp(2i\psi)$ maps \hat{S}_1^{iso} is larger when the mean magnetic field is along the line of sight.
- \hat{S}_1^{aniso} coefficients quantify the statistical anisotropy of the maps. When the mean magnetic field is parallel to the line of sight \hat{S}_1^{aniso} coefficients are negligible, while when the mean magnetic field is in the plane of the sky they allow us to identify the direction of anisotropy at each scale. For the MHD simulation we analyzed, this direction is orthogonal to the direction of the mean magnetic field for both the $\tilde{Q}_\perp + i\tilde{U}_\perp$ and $\exp(2i\psi)_\perp$ maps.
- Second order RWST coefficients clearly exhibit the non-Gaussianity of $\tilde{Q} + i\tilde{U}$ maps (although this is also visible to a lesser extent in first order coefficients). While the randomized $R[\tilde{Q} + i\tilde{U}]$ data sets present characteristic properties of self-similar Gaussian fields ($\hat{S}_2^{\text{iso},1}$ is a function of the scale difference $j_2 - j_1$ only, and $\hat{S}_2^{\text{iso},2}$ quickly decreases to zero as $j_2 - j_1$ increases), the $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$ coefficients for the corresponding $\tilde{Q} + i\tilde{U}$ data sets show clearly different patterns. In particular, the strictly positive values of $\hat{S}_2^{\text{iso},2}$ at large $j_2 - j_1$ are interpreted as signatures of the filamentary structure of the maps.

This chapter has demonstrated the interest of the (R)WST to describe the statistical properties of polarization maps of the diffuse interstellar medium. However, it remains unclear how to define statistical models based on these descriptors. Defining such models is the object of the next chapter. This will be the opportunity to quantitatively address the question of the comprehensiveness of the RWST statistics.

To establish the methodology, we have worked with noise-free maps computed from numerical simulations of MHD turbulence. One of the next steps is to apply this methodology to observations of polarized thermal emission from dust. To do this, we need to learn how to handle data noise. Indeed, while signal-to-noise ratios in total intensity for both *Herschel* and *Planck* maps are quite high, this is not the case for currently available polarization data. Studying how data noise affects the WST coefficients demands to repeat the analysis of MHD simulations with noise added to the dust signal. First elements of analysis of the impact of noise on RWST statistics will be introduced in Chapter V, as well as a strategy to retrieve the statistics of the noise-free emission. Once this difficulty is overcome, the RWST might be used to define a metric to compare observations with simulations and phenomenological models. This will be a stepping stone towards a more refined physical interpretation of the RWST coefficients. A main motivation would be to use the RWST to characterize magnetized interstellar turbulence.

Throughout this work we chose to work with Stokes I , Q , and U maps to analyze the polarization of dust thermal emission as astronomers do for Galactic astrophysics. In the framework of CMB data analysis, polarization is more usually characterized through E and B -modes as explained in Chapter I. Thus, it would be interesting to conduct similar RWST analyses on E and B maps. This will be the first step to define a statistical model that is consistent with observational constraints such as the E/B power asymmetry, or the TE and TB correlations [57].

Chapter IV

Generative models from scattering-like descriptions

In the previous chapter, we have introduced the RWST statistics to characterize the non-Gaussian statistical properties of Stokes maps of the polarized emission of interstellar dust. These statistics include an information that goes beyond power spectrum statistics by means of a quantification of interactions between scales. They constitute a low-dimensional and multiscale representation of the data which can be interpreted and related to the physics of the ISM to some extent. In this chapter, we now address the question of the comprehensiveness of these statistics to characterize dust maps, and their ability to define a statistical model of the dust emission. The ultimate, albeit idealistic, goal is to find the optimal set of statistical descriptors that would tell us all there is to know about the statistical properties of the dust emission. And, at the same time, we want to be able to define a statistical generative model based on these descriptors, from which generating new random realizations with properties consistent with the original dust maps would be possible.

Defining a meaningful generative model based on a set of statistical measurements and from which realizations can be efficiently drawn is an active topic of research in data science. Our approach will be that of [190], which makes use of microcanonical gradient descent models to build generative models from WST statistics. We will define in a similar way microcanonical gradient descent models of simulated $\tilde{Q} + i\tilde{U}$ maps based on their RWST statistics first, and then making use of another set of statistics called *wavelet phase harmonics* (WPH), which we will also introduce. Thanks to these models, we will be able to generate new realizations, or *syntheses*, from which an assessment of the exhaustiveness of our statistics will be made possible. In practice, we will compare the synthetic maps with the original ones by means of a set of summary statistics. Obviously, the summary statistics will not include those used to define the model, and whatever we choose, it will remain arbitrary to a certain extent. Here again, we emphasize that this choice should be guided by the applications we have in mind for such models (see Sect. II.1). We will of course also make use of visual assessments, although we will keep in mind that it is unclear what kind of statistics human eyes are sensitive to (see e.g. [191] for a discussion). Moreover, we will be wary of the fact that such visual assessments may greatly depend on the way the syntheses are plotted, notably on the chosen colormap or range of the

colorbar.

This chapter is organized as follows. In a first section, we introduce microcanonical gradient descent models following [190]. Then, in a second section, we use them to define a generative model of simulated $\tilde{Q} + i\tilde{U}$ maps based on RWST statistics and complementary additional constraints, before performing a first assessment of the exhaustiveness of these statistics to characterize dust polarized emission. In a third section, we define a similar model based on another set of statistics called wavelet phase harmonics (WPH). After a formal introduction of the WPH, we proceed similarly to assess the exhaustiveness of the resulting model. Finally, we conclude this chapter by summarizing our results and outlining a few perspectives on this topic.

Computing WPH statistics has motivated the development of a Python package, called PyWPH, allowing GPU accelerated computations of WPH statistics with PyTorch [192] and convenient functions to manipulate the coefficients. This package is introduced in Appendix C.¹

IV.1 Microcanonical models

In this section, we introduce microcanonical models. These are guided by the principle of maximum entropy, which states that the probability distribution that best represents our knowledge of some system is that with largest entropy (in the sense of information theory). Our introduction follows [190] and we refer the reader to this paper and to the references therein for a more detailed presentation. We first define maximum entropy microcanonical models, before discussing microcanonical gradient descent models which were introduced in [190] as approximations of the former. While no longer of maximum entropy, they allow to circumvent the problem of sampling in high-dimensional spaces. We conclude this section by discussing numerical practical considerations to effectively draw samples from microcanonical gradient descent models.

IV.1.1 Maximum entropy microcanonical models

Let us consider a random field X and x one of its realizations. We want to approximate the distribution of X based on this single realization x . In practice, we start by making some statistical measurements on x , which thus define a vector of statistics $\phi(x)$. While the realization x lives in a space of dimension d , let us say \mathbb{R}^d for the purpose of this discussion, $\phi(x)$ typically lives in a space of lower dimension K , let us say \mathbb{R}^K . We define microcanonical sets as ensembles of vectors of \mathbb{R}^d whose statistics are "sufficiently close" to that of x . Formally, we define the microcanonical set Ω_ϵ for $\epsilon > 0$ as:

$$\Omega_\epsilon = \{y \in \mathbb{R}^d / \|\phi(y) - \phi(x)\| \leq \epsilon\}, \quad (\text{IV.1})$$

where $\|\cdot\|$ is some norm on the statistical space \mathbb{R}^K .

In this context, a maximum entropy microcanonical model defined on Ω_ϵ is a probability distribution μ_ϵ supported in Ω_ϵ with maximal entropy. The entropy of a probability distribution μ here refers to its differential entropy, called $H(\mu)$ and defined as:

$$H(\mu) = - \int f_\mu(y) \log f_\mu(y) dy, \quad (\text{IV.2})$$

¹PyWPH is available at <https://github.com/bregaldo/pywph>.

where f_μ is the probability density function (PDF) associated to μ . Assuming that the function ϕ allows Ω_ϵ to be compact, this maximum entropy distribution μ_ϵ is simply the uniform distribution on Ω_ϵ and is defined by its uniform density:

$$f_{\mu_\epsilon}(y) = \frac{1_{\Omega_\epsilon}(y)}{\int_{\Omega_\epsilon} dy}. \quad (\text{IV.3})$$

The relevance of this kind of models directly depends on our choice of statistical measurements, represented by the function ϕ , as well as on the value of ϵ which is a proxy of the volume of the microcanonical set Ω_ϵ . Ideally, we want to choose ϕ and ϵ so that typical samples of X are contained in Ω_ϵ , and conversely, typical samples of μ_ϵ are representative of those of X .

IV.1.2 Microcanonical gradient descent models

Even if we manage to define relevant ϕ and ϵ so that the corresponding maximum entropy microcanonical model correctly approximates the distribution of X on paper, we still need to find an efficient way to draw samples from μ_ϵ . Usual strategies make use of Markov chain Monte Carlo (MCMC) algorithms, however these algorithms reach computational limits when the dimension of the samples d increases.² For the applications of this thesis, these are not an option.

Instead we will make use of approximations of maximum entropy microcanonical models, called microcanonical gradient descent models. These are still supported in a microcanonical set Ω_ϵ , but are no longer of maximum entropy in general. They are defined by the transport of an initial distribution μ_0 through gradient descent towards the microcanonical set Ω_ϵ . In practice this requires to define a loss function $\mathcal{L} : \mathbb{R}^d \rightarrow \mathbb{R}_+$ that will "guide" any sample of μ_0 towards Ω_ϵ . With the following natural choice:

$$\mathcal{L}(y) = \|\phi(y) - \phi(x)\|^2, \quad (\text{IV.4})$$

the gradient descent algorithm defines at each iteration k a mapping $\varphi_k(y) = y - \kappa_k \nabla \mathcal{L}(y)$ with $\nabla \mathcal{L}$ the gradient of \mathcal{L} and κ_k the gradient step at iteration k . With y_0 a sample drawn from μ_0 , this algorithm builds a sequence $\{y_0, y_1, \dots, y_n\}$ such that for all $k \in \llbracket 1, n \rrbracket$, $y_k = \varphi_k \circ \varphi_{k-1} \circ \dots \circ \varphi_1(y_0)$. Provided ϕ agrees with a certain set of conditions (see [190]), y_k will eventually belong to Ω_ϵ for a sufficiently large value of k .

Obviously, one should already be able to efficiently draw samples from μ_0 . In the following, the initial distribution μ_0 will be derived from a Gaussian white noise distribution, which is trivial to sample. It will be either directly equal to a white noise distribution, or a simple transformation of this white noise. Note that, in the case where μ_0 is a Gaussian white noise, it is proven in [190] that typical symmetries of ϕ (e.g. translation or rotation invariance) remain symmetries of the resulting model.

IV.1.3 Numerical considerations

a) Notions of optimization

Sampling a microcanonical gradient descent model thus amounts to a problem of optimization, that is the minimization of $\mathcal{L} : \mathbb{R}^d \rightarrow \mathbb{R}_+$ starting from a given point y_0 .

²In general, Markov chain mixing time depends on the exponential of d [193].

Instead of using a gradient descent algorithm to minimize this function (see e.g. [194]), we will use the L-BFGS algorithm that is implemented in SciPy [195, 196, 197]. This has been empirically proven more efficient in minimizing \mathcal{L} for our applications. Even if the mathematical properties of models defined with a L-BFGS algorithm are less clear, we will reasonably assume that they remain quite similar to that of proper microcanonical gradient descent models.

The L-BFGS algorithm belongs to the class of quasi-Newton methods, which are based on Newton's method. Where the gradient descent algorithm is an iterative algorithm only relying on first-order approximations of \mathcal{L} , Newton's method relies on second-order approximations of \mathcal{L} . Provided \mathcal{L} is twice differentiable at $y \in \mathbb{R}^d$, we can write the second-order Taylor series expansion of \mathcal{L} about y :

$$\mathcal{L}(y + h) \approx \mathcal{L}(y) + \nabla \mathcal{L}(y)^T h + \frac{1}{2} h^T H_{\mathcal{L}}(y) h, \quad (\text{IV.5})$$

with $\nabla \mathcal{L}$ and $H_{\mathcal{L}}$ the gradient and the Hessian matrix of \mathcal{L} , respectively, and $h \in \mathbb{R}^d$ a sufficiently small increment (in norm). Calling y_k the position of the optimizer at iteration k , Newton's method will compute the step h_k that minimizes this expansion, that is $h_k = -[H_{\mathcal{L}}(y_k)]^{-1} \nabla \mathcal{L}(y_k)$ (see [194]), and the position y_{k+1} at next iteration is simply set to $y_{k+1} = y_k + h_k$. Note that Newton's method is well defined when $H_{\mathcal{L}}$ is positive definite, otherwise the algorithm must be adapted. Nevertheless, for the purpose of this discussion, let us remark that while a second-order method is generally more refined than a mere first-order method, it is much more computationally expensive, and all the more if the dimension d is large. Indeed, the Hessian matrix $H_{\mathcal{L}}$ of size $d \times d$ must be computed, stored, and inverted, which may be computationally intractable for our applications where d is typically the number of pixels of an image of 512×512 pixels, meaning $d = 262144$. Quasi-Newton methods circumvent this issue when d becomes too large by approximating the Hessian $H_{\mathcal{L}}$ at each iteration. In particular, the L-BFGS algorithm manages to approximate the inverse of the Hessian with a limited memory cost based on the history of the positions and of the corresponding gradients during the optimization [195]. We refer to [194] for a broader explanation of these optimization techniques.

b) Automatic differentiation

Regardless of the optimization method, we will have to estimate at some point gradients of \mathcal{L} . This loss function will be in practice too complicated for us to derive any analytic expression of its gradient. Thus, we will have to resort to numerical strategies. Common finite-difference schemes would require to estimate d partial derivatives, involving at least $2d$ costly evaluations of \mathcal{L} , and leading necessarily to approximate gradients. Another more realistic alternative is that of automatic differentiation, which is a common technique of machine learning allowing to compute exact gradients by keeping track of the sequence of operations that are successively applied to a given y in order to compute $\mathcal{L}(y)$. The idea behind that is that any complicated function will be at some point reducible to a sequence of elementary operations (addition, subtraction, etc) or functions (exp, log, sin, cos, etc) for which analytic derivatives do exist. By applying the chain rule to this sequence of operations, one can automatically extract the resulting derivatives.

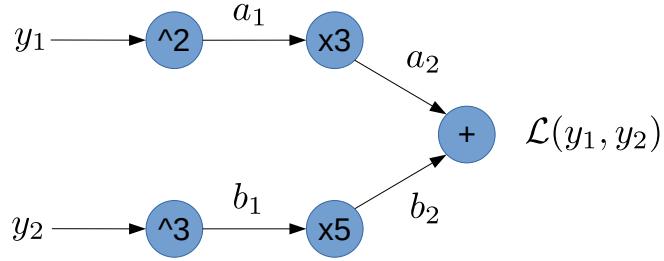


Fig. IV.1 Example of a computation graph associated to the function $\mathcal{L}(y_1, y_2) = 3y_1^2 + 5y_2^3$ for automatic differentiation. The $\hat{\cdot}$ symbol refers to the exponentiation operation.

Let us take a simple example, defining $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}$ as $\mathcal{L}(y_1, y_2) = 3y_1^2 + 5y_2^3$. We have trivially:

$$\frac{\partial \mathcal{L}}{\partial y_1} = 6y_1 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial y_2} = 15y_2^2. \quad (\text{IV.6})$$

However, with automatic differentiation, $\mathcal{L}(y_1, y_2)$ is internally described as a computation graph such as the one depicted in Fig. IV.1, thus giving $\mathcal{L}(y_1, y_2) = a_2(a_1(y_1)) + b_2(b_1(y_2))$. The application of the chain rule gives:

$$\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial a_2} \frac{\partial a_2}{\partial a_1} \frac{\partial a_1}{\partial y_1} \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial y_2} = \frac{\partial \mathcal{L}}{\partial b_2} \frac{\partial b_2}{\partial b_1} \frac{\partial b_1}{\partial y_2}. \quad (\text{IV.7})$$

All the factors of this expression involve standard derivatives which can be automatically and exactly computed. On this example we have:

$$\frac{\partial \mathcal{L}}{\partial a_2} = 1, \quad \frac{\partial a_2}{\partial a_1} = 3, \quad \frac{\partial a_1}{\partial y_1} = 2y_1, \quad (\text{IV.8})$$

$$\frac{\partial \mathcal{L}}{\partial b_2} = 1, \quad \frac{\partial b_2}{\partial b_1} = 5, \quad \frac{\partial b_1}{\partial y_2} = 3y_2^2. \quad (\text{IV.9})$$

Finally, by replacing the expression of these derivatives in Eq. (IV.7), we simply retrieve the gradient of \mathcal{L} as written in Eq. (IV.6). This illustrates how automatic differentiation may extend to more complicated loss functions \mathcal{L} .

In the following, we will make use of automatic differentiation to compute gradients of \mathcal{L} by means of the Python library PyTorch [192]. This library will also allow us to accelerate the computations of \mathcal{L} by employing GPUs.

IV.2 Application with the RWST

In the continuity of the last chapter, we now define microcanonical gradient descent models of simulated $\tilde{Q} + i\tilde{U}$ maps based on their RWST statistics. The immediate interest is twofold: 1) it allows us to assess the exhaustiveness of RWST statistics to characterize the statistical properties of dust polarization maps, 2) it provides a way to generate new realizations with interesting statistical similarities with the original maps. This last point is of strong interest for data augmentation, which is all the more valuable when we know

how costly a MHD simulation can be, or for producing realistic foregrounds for component separation methods. The results of this section are directly taken from [168]. Note that this approach follows what has been done for dust total intensity in [167].

IV.2.1 Generation of synthetic polarization maps

If synthetic total intensity maps generated from a RWST description have already been produced in [167], here I extend and improve this procedure for complex polarization maps. In the following, we generate a synthetic map based on the RWST statistics of the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set, which was introduced in the previous chapter.

The generation of synthetic maps based on microcanonical gradient descent models is an iterative process consisting in the minimization of the loss function \mathcal{L} in pixel space. With 512×512 pixels maps, this optimization problem is thus defined in practice in a complex space of dimension $d = 512^2$. We start from a realization of a complex Gaussian white noise map $x_0 = \tilde{Q}_0 + i\tilde{U}_0$, and successive maps $\{x_k = \tilde{Q}_k + i\tilde{U}_k\}$ are built through the minimization of \mathcal{L} with a L-BFGS algorithm [195, 196]. We employ a modified version of `Kymatio` to compute WST statistics of complex maps with PyTorch, allowing GPU acceleration and automatic differentiation [198, 192].

a) Initial and reference maps

First, we pick one specific map $x^r = \tilde{Q}^r + i\tilde{U}^r$ from the 14 maps of the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set, to serve as a reference for comparison with the synthetic map. The initial map x_0 is drawn from a complex Gaussian white noise with same mean and standard deviation as x^r . We recall that our WST and RWST analyses described in the previous chapter was restricted to the range of scales $j \in \llbracket 0, J - 1 \rrbracket$, consequently, we have no statistical information on scales larger than $\lambda \approx 171$ pixels (see Table III.2) nor on their couplings with smaller scales. Thus, we replace the largest scales of x_0 with those of x^r to address this gap in a deterministic way³. Formally, denoting $\mathcal{F}[x_0]$ the Fourier transform of x_0 , we modify x_0 by setting:

$$\mathcal{F}[x_0](\mathbf{k}) = \mathcal{F}[x^r](\mathbf{k}) \text{ for } |\mathbf{k}| < k_{\min}, \quad (\text{IV.10})$$

where k_{\min} is the wavenumber corresponding to the largest scale probed by the WST, that is $k_{\min} \approx 2\pi/171 \approx 0.037 \text{ pixel}^{-1}$. In practice, 25 Fourier modes out of 512^2 are modified by this procedure. To illustrate, we show in Fig. IV.5 the resulting Q_0 map (top left map). Even if in our case this value of k_{\min} is related to an ad hoc modeling decision (J value), we recall that statistical approaches are not always relevant at all scales for the analysis of the diffuse ISM (see Sect. II.1 for a discussion). Moreover, we have noted that the WST of an image does not adequately characterize its one-point statistics. For this reason, in the following, we impose additional constraints on one-point moments of the synthetic map so that they match those of x^r .

³Note that this deterministic strategy cannot account for the couplings between the largest scales with smaller ones.

b) Loss function

We define the loss function \mathcal{L} of a complex image $x = \tilde{Q} + i\tilde{U}$ as follows:

$$\mathcal{L}[x] = \mathcal{L}_{\text{WST}}[x] + \alpha (\mathcal{L}_{\text{one-point}}[\tilde{Q}] + \mathcal{L}_{\text{one-point}}[\tilde{U}]), \quad (\text{IV.11})$$

with $\mathcal{L}_{\text{WST}}[x]$ the loss function constraining the WST coefficients of x , $\mathcal{L}_{\text{one-point}}[\tilde{Q}]$ (respectively $\mathcal{L}_{\text{one-point}}[\tilde{U}]$) that constraining a few one-point moments of \tilde{Q} (respectively \tilde{U}), and α a weighting coefficient balancing the importance of these two types of constraints. More specifically we set:

$$\begin{aligned} \mathcal{L}_{\text{WST}}[x] = & \frac{1}{N} \left((S_0[x] - S_0^t)^2 + \sum_{j_1, \theta_1} (S_1(j_1, \theta_1)[x] - S_1^t(j_1, \theta_1))^2 \right. \\ & \left. + \sum_{j_1, j_2, \theta_1, \theta_2} (S_2(j_1, \theta_1, j_2, \theta_2)[x] - S_2^t(j_1, \theta_1, j_2, \theta_2))^2 \right), \end{aligned} \quad (\text{IV.12})$$

where the "t" superscript refers to the *target* WST coefficients that the synthetic map should have, and N is the total number of WST coefficients. In our case ($J = 7$ and $L = 8$) we recall that we have $N = 2801$. These target WST coefficients are computed from the RWST coefficients derived from a given data set of $\tilde{Q} + i\tilde{U}$ maps (of which x^r is part) using the RWST model defined in Eqs. (III.20) and (III.21). We see on this loss function that none of the WST coefficients is privileged and that we do not weigh the differences of WST coefficients by any uncertainty on the target coefficients. This is certainly something that can be improved in a future work on WST syntheses. We finally define:

$$\mathcal{L}_{\text{one-point}}[\tilde{Q}] = \frac{1}{3} \left(\left(\frac{\mathcal{M}_2[\tilde{Q}]}{\mathcal{M}_2[\tilde{Q}^r]} - 1 \right)^2 + (\mathcal{M}_3[\tilde{Q}] - \mathcal{M}_3[\tilde{Q}^r])^2 + (\mathcal{M}_4[\tilde{Q}] - \mathcal{M}_4[\tilde{Q}^r])^2 \right), \quad (\text{IV.13})$$

where $\mathcal{M}_2[\tilde{Q}] = \langle (\tilde{Q} - \langle \tilde{Q} \rangle)^2 \rangle$ is the variance of the distribution of pixels values of \tilde{Q} , $\mathcal{M}_3[\tilde{Q}] = \langle (\tilde{Q} - \langle \tilde{Q} \rangle)^3 \rangle / \mathcal{M}_2[\tilde{Q}]^{3/2}$ is its skewness, and $\mathcal{M}_4[\tilde{Q}] = \langle (\tilde{Q} - \langle \tilde{Q} \rangle)^4 \rangle / \mathcal{M}_2[\tilde{Q}]^2$ is its kurtosis.

c) Synthesis from the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set

We generate a synthetic map using the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ RWST description as a target but note that equivalent results for the $\tilde{Q}_{\perp} + i\tilde{U}_{\perp}$ RWST description are also provided in [168]. The optimization algorithm stops when the loss function or its projected gradient stagnates between two consecutive iterations. In practice, we choose $\alpha = 10^{-5}$ and the algorithm stops when $|\mathcal{L}[x_k] - \mathcal{L}[x_{k+1}]| < 10^{-7}$. These numerical values have been adjusted empirically to guarantee satisfactory visual and quantitative results.⁴ This optimization requires 12 evaluations of the loss function (see Fig. IV.2 for a plot of the evolution of

⁴Note that these numerical values are related to the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set, and might need to be modified for other data. Moreover, they are given for a normalized $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ data set, i.e., with maps of unit variance on average.

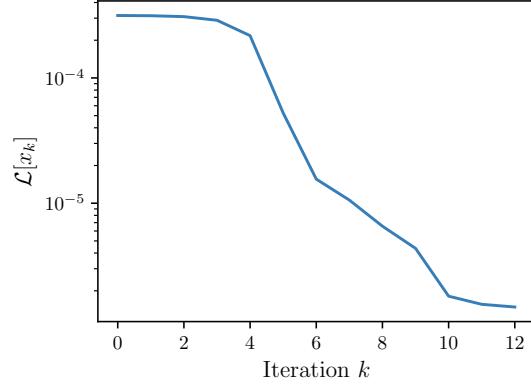


Fig. IV.2 Evolution of the loss function during the generation of the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map shown in Fig. IV.3.

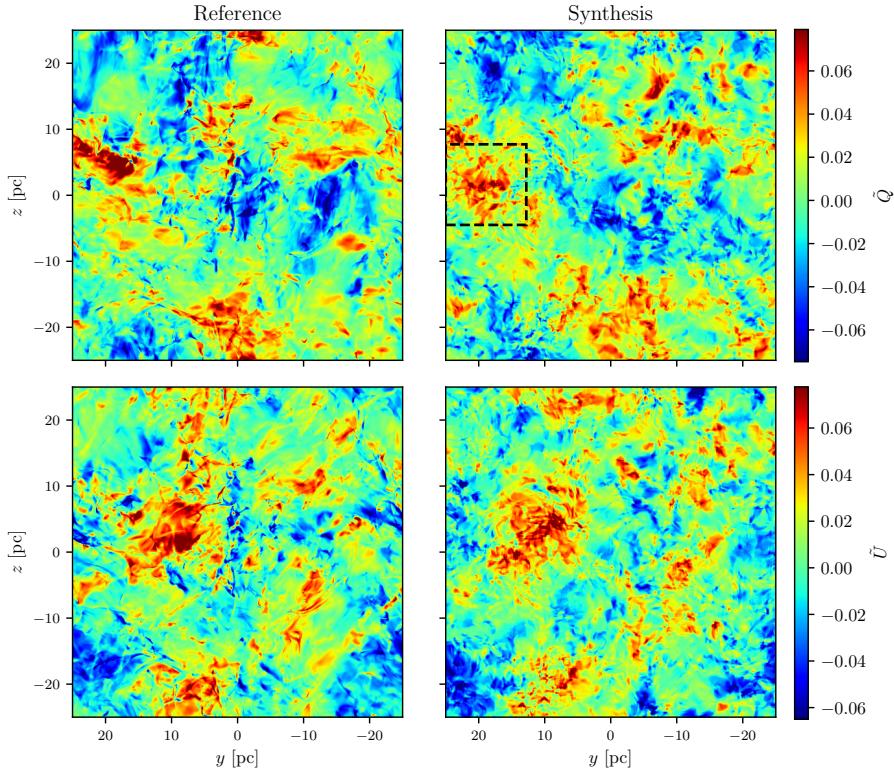


Fig. IV.3 Synthesis of a $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map (right column, with \tilde{Q}_{\parallel} on top and \tilde{U}_{\parallel} on bottom) built from RWST statistics and additional constraints on large-scale components and on a few one-point moments of \tilde{Q} and \tilde{U} maps, to be consistent with those of a reference map (left column). The square plotted in black dashed lines on the top right map delimits the region shown in Fig. IV.4.

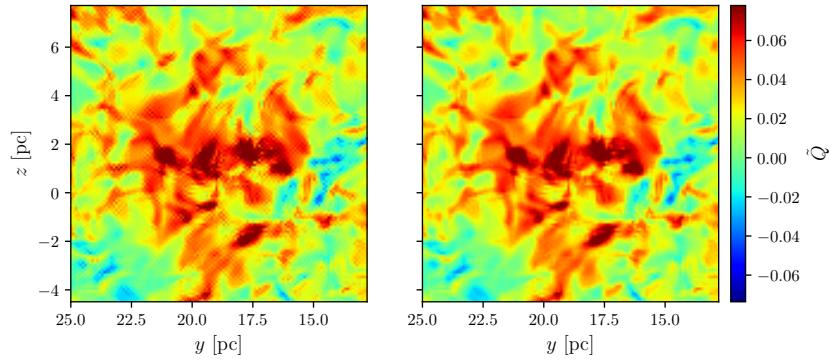


Fig. IV.4 Zoom on the \tilde{Q}_{\parallel} synthetic map shown in Fig. IV.3 before (left) and after (right) the filtering of modes beyond the Nyquist disc. This filter allows to mitigate checkerboard patterns visible in the left panel at pixel scale.

the loss), taking less than 30 seconds on a Quadro P2000 GPU with 4 Gb of memory. The statistical coefficients of the resulting synthesis match the target coefficients with a relative error of less than one percent on average. Once the optimization is done, we post-process the resulting map by filtering modes beyond the Nyquist disc (that is those with a wavenumber higher than the one dimensional Nyquist wavenumber $k_N = \pi \text{ pixels}^{-1}$; see Appendix A) in order to avoid numerical artifacts such as checkerboard patterns shown in Fig. IV.4. Indeed, the WST relies on a bank of Morlet wavelets that does not properly cover frequencies beyond the Nyquist disc, which results in a total loss function \mathcal{L} that poorly constrains these modes (see Appendix B). Figure IV.3 shows the final synthetic $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map (right column) next to its reference map (left column). The overall appearance of the synthetic map is satisfactory. The largest scales are roughly consistent with those of the reference maps⁵, and we also see at intermediate and small scales, which are the truly synthetic scales, consistent filamentary patterns and dynamic ranges between the synthetic and reference maps. Finally, we show in Fig. IV.5 the sequence of maps \tilde{Q}_k computed over the iterations of the optimization. Note that the final synthetic map, corresponding to iteration $k = 12$, is not reproduced in this figure as it is already shown in Fig. IV.3. The initial modified white noise ($k = 0$ map) acts as a random seed for the structures of the final synthesis. Similar results are obtained for the maps \tilde{U}_k .

IV.2.2 One-point and two-point statistics of synthetic maps

By construction, the synthetic map we built has approximately the same WST statistics as prescribed by the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ RWST description, as well as similar large scales and one-point moments as the reference map. We may ask whether elementary one-point and two-point statistics are fully consistent with those of the reference map.

Figure IV.6 shows the power spectra of \tilde{Q} , \tilde{U} , and $\tilde{Q} + i\tilde{U}$ up to the one dimensional Nyquist wavenumber k_N for both the reference maps and the syntheses shown in Fig. IV.3. The power spectra are computed by binning the squared amplitudes of Fourier modes with respect to the modulus of the corresponding wavenumber k . We use a regular binning in

⁵Since the WST constraints leave the largest scales of x_0 unchanged, this shows that the impact of the one-point moments constraints on these scales is minor.

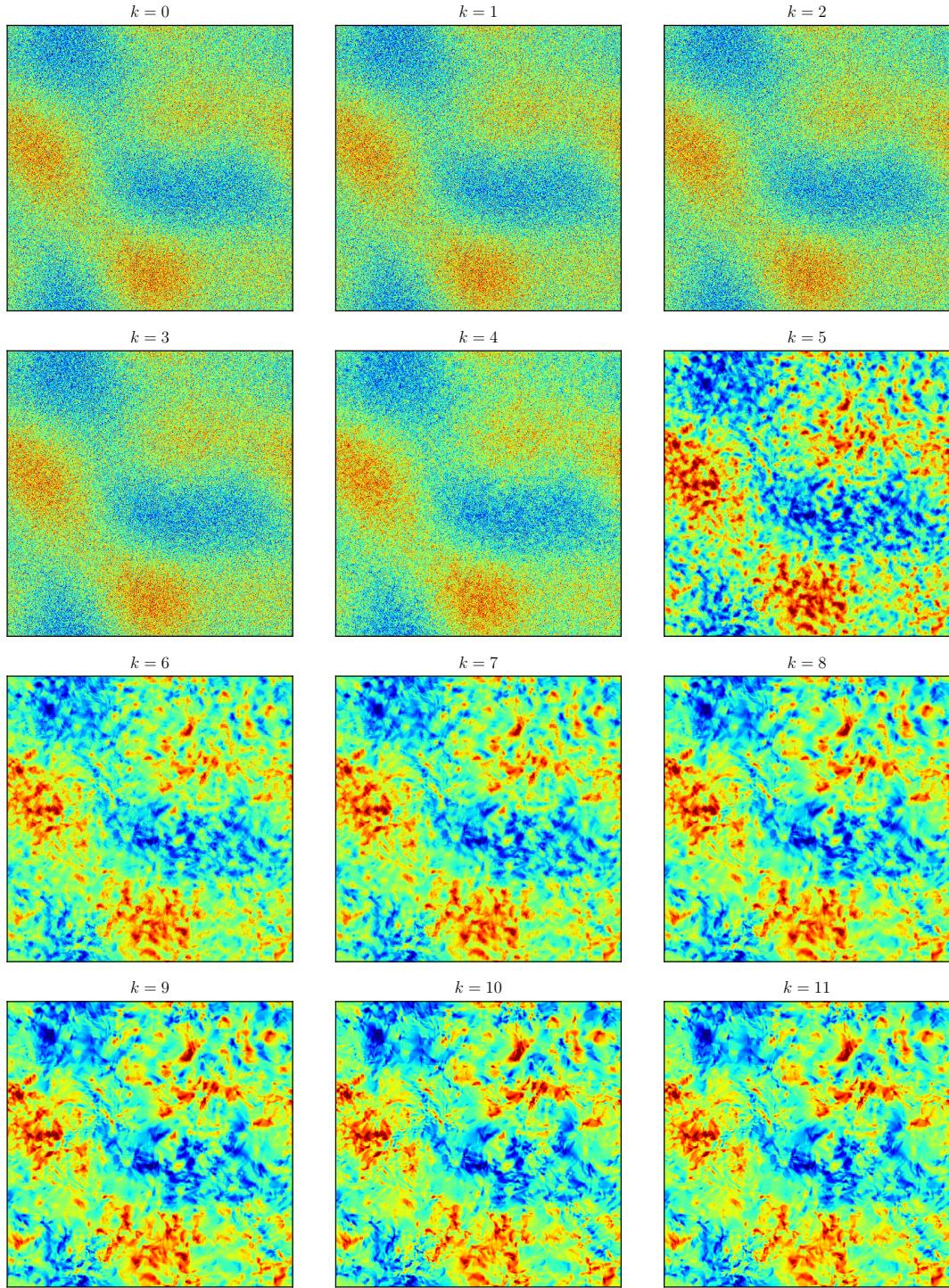


Fig. IV.5 Sequence of maps Q_k computed for the generation of the $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map shown in Fig. IV.3.

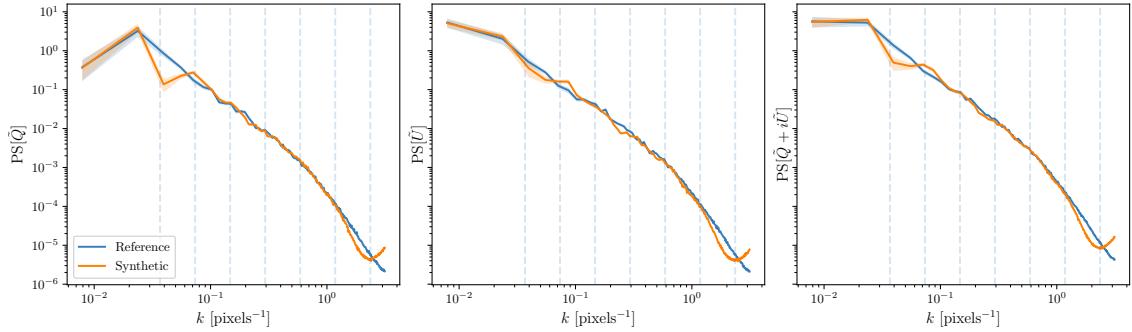


Fig. IV.6 Power spectra of the synthetic $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map shown in Fig. IV.3, compared to those of the reference map, for \tilde{Q} , \tilde{U} , and $\tilde{Q} + i\tilde{U}$ (left, middle, and right, respectively). The vertical dashed lines mark the wavelet central wavenumbers corresponding to the scale indices $j = 0, \dots, J - 1$ (from right to left).

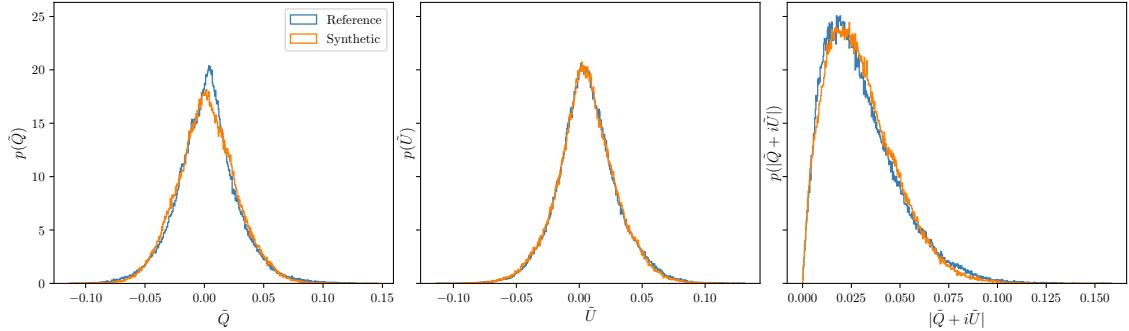


Fig. IV.7 One-point distribution functions of the synthetic $\tilde{Q}_{\parallel} + i\tilde{U}_{\parallel}$ map shown in Fig. IV.3, compared to those of the reference map, for \tilde{Q} , \tilde{U} , and $|\tilde{Q} + i\tilde{U}|$ (left, middle, and right, respectively).

k and the estimations of the power spectra are computed as the means for each bin (see Sect. A.2.2 for additional details).⁶ We also represent standard deviations of the mean per bin. We see that the power spectra as well as their standard deviations are in good agreement for the three variables \tilde{Q} , \tilde{U} , and $\tilde{Q} + i\tilde{U}$ for all scales except the smallest ones. These discrepancies at small scales take the form of a lack followed by an excess of power in the syntheses for wavenumbers approaching k_N . We interpret this as the result of poorly constrained modes close to k_N in the optimization process. It is not surprising to reproduce the power spectrum of $\tilde{Q} + i\tilde{U}$ maps because the \tilde{S}_1 and \tilde{S}_2 coefficients constrain the power spectrum of $\tilde{Q} + i\tilde{U}$ (see discussion in Sect. III.1.2). However, we also correctly reproduce the power spectra of \tilde{Q} and \tilde{U} taken separately, for which there was no direct constraint. Still, we did not investigate cases where \tilde{Q} and \tilde{U} have very different power spectra.

⁶We point out that the chosen estimator of the power spectrum, leads by construction to $\text{PS}[\tilde{Q} + i\tilde{U}] = \text{PS}[\tilde{Q}] + \text{PS}[\tilde{U}]$. This is a direct consequence of the fact that $\hat{\tilde{Q}}(\mathbf{k})\hat{\tilde{U}}^*(\mathbf{k}) = \hat{\tilde{Q}}^*(-\mathbf{k})\hat{\tilde{U}}(-\mathbf{k})$ for any mode \mathbf{k} .

In Fig. IV.7, we compare the one-point distribution functions of the reference maps and the synthetic maps for \tilde{Q} , \tilde{U} , and $|\tilde{Q} + i\tilde{U}|$. These are in fairly good agreement for all the variables, with particularly striking match for \tilde{U} . We also reproduce quite well the tails of these distributions, this must be the consequence of the combined constraints on WST coefficients and one-point moments. Note that one could surely enhance the agreement between the reference and synthetic maps by taking into account higher order moments in the $\mathcal{L}_{\text{one-point}}$ loss functions.

We recall that in this work the RWST statistics comprise 127 coefficients only (see Sect. III.2.4). Adding to that the largest scales that were set in a deterministic way as well as the constraints on the one-point moments of \tilde{Q} and \tilde{U} maps, we end up with a total of 158 coefficients to generate 512×512 complex $\tilde{Q} + i\tilde{U}$ maps that are in good visual agreement with the maps of the original data set and successfully reproduce their one-point and two-point statistics. This shows that microcanonical gradient descent models derived from RWST statistics constitute a promising avenue to model polarization maps of the dust emission, needing only a few hundred statistical coefficients, which can be derived from a single map or from a collection of maps as it was the case here.

IV.3 Application with the WPH

In this section, we define microcanonical gradient descent models of the same data but making use of another set of statistics called wavelet phase harmonics (WPH). The WPH statistics were introduced in [5] to describe phase dependencies across scales with a mathematical formalism inspired by the properties of CNNs. These statistics allow to characterize coherent structures in non-Gaussian random fields, and their ability to define maximum entropy models was first studied in [6]. In [199] models of simulated data of the large-scale structure of the Universe based on their WPH statistics have been introduced, and it was demonstrated quantitatively that the synthetic maps reproduce well the power spectrum, bispectrum, and Minkowski functionals of the input data. We follow in the footsteps of these works to define statistical models of dust emission polarization maps.

After a formal introduction to the WPH statistics, we define microcanonical gradient descent models derived from the WPH statistics of the same simulated $\tilde{Q} + i\tilde{U}$ maps as before. Then we assess how the statistics of synthetic maps compare to those of the original ones, using the results of the previous section as a benchmark.

The computations of WPH statistics are made with PyWPH, a Python package whose design was motivated by this work and that of the following chapter. We refer to Appendix C for a presentation of the package.

IV.3.1 The wavelet phase harmonics

We introduce the WPH statistics from a practical perspective. We use them as interpretable low-dimensional statistics allowing to describe non-Gaussian random fields. However, the reach of this set of statistics is wider, as they were originally introduced in data science to give a complementary mathematical view of key properties of CNNs [5], in line with previous ideas that motivated the definition of the WST [166]. In particular, the phase harmonic operator, which will be defined below, is closely related to rectifier

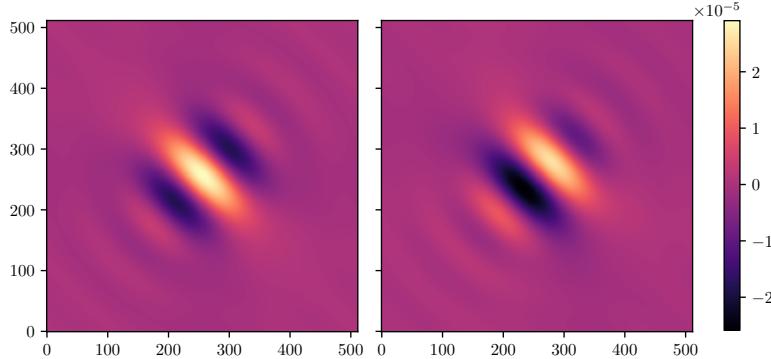


Fig. IV.8 Real part (left) and imaginary part (right) of a 512×512 bump-steerable wavelet $\psi_{6,\pi/4}$. The wavelet is centered on the middle of the map for a better visualization.

units of CNNs, and plays a central role in efficiently capturing phase alignments between different scales [5].

Note that this introduction is partly inspired from that of [200]. As usual, we call X a given random field and x one of its realizations. We also assume X to be statistically homogeneous. The WPH statistics of x correspond to estimates of the WPH moments of X . However, similarly to [199], we complement these statistics by estimates of scaling moments of X . To simplify, in the following, the WPH statistics will thus refer to both the estimates of the WPH and scaling moments. These are defined in the following.

a) Bump-steerable wavelets

Like the WST, the WPH moments rely on the wavelet transform. We recall that with $\{\psi_{\xi_1}, \dots, \psi_{\xi_N}\}$, a bank of wavelets labeled by their central wavevectors ξ_i , the wavelet transform of X corresponds to the set $\{X \star \psi_{\xi_1}, \dots, X \star \psi_{\xi_N}\}$. These convolutions amount to a local bandpass filtering of X on the scales probed by each of the wavelets. Contrary to Sect. III.1.2, we employ bump-steerable wavelets instead of Morlet wavelets following [5]. The mother bump-steerable wavelet is defined in Fourier space as follows:

$$\hat{\psi}(\mathbf{k}) = \exp\left(\frac{-(k - \xi_0)^2}{\xi_0^2 - (k - \xi_0)^2}\right) \cdot 1_{[0, 2\xi_0]}(k) \times \cos^{L-1}(\arg(\mathbf{k})) \cdot 1_{[0, \pi/2]}(|\arg(\mathbf{k})|), \quad (\text{IV.14})$$

with $k = \|\mathbf{k}\|$, $1_A(x)$ the indicator function that returns 1 if $x \in A$ and 0 otherwise, and $\xi_0 = 0.85\pi$ the central wavenumber of the mother wavelet. The bank of wavelets is here again built from dilations and rotations of the mother wavelet ψ , and we consider J dilation indices j ranging from 0 to $J - 1$, and L or $2L$ rotation angles of the form $\{\theta_k = k\pi/L\}$ depending on whether the target field X is real or complex-valued, respectively. We denote by $\xi_{j,\theta} = 2^{-j}\xi_0 \mathbf{u}_\theta$ with $\mathbf{u}_\theta = \cos \theta \mathbf{u}_x + \sin \theta \mathbf{u}_y$ the central wavevector of each wavelet. These wavelets cover most of Fourier space with their respective band-passes (or Fourier half-space for real-valued X). We refer to Appendix B for additional details on bump-steerable wavelets and on the wavelet transform.

In this study, we work with 512×512 maps and choose $J = 7$ and $L = 8$. We show in Fig. IV.8 one example wavelet from the bank.

b) WPH moments

The WPH moments of X are covariances of the phase harmonics of the wavelet transform of X , that is, for a statistically homogeneous X , these are defined as follows:

$$C_{\xi_i, p_i, \xi_j, p_j}(\tau) = \text{Cov} \left([X \star \psi_{\xi_i}(\mathbf{r})]^{p_i}, [X \star \psi_{\xi_j}(\mathbf{r} + \tau)]^{p_j} \right), \quad (\text{IV.15})$$

with $z \mapsto [z]^p = |z| \cdot e^{ip \arg(z)}$ the phase harmonic operator⁷. When applied to a complex z , the phase harmonic operator preserves the modulus of z but multiplies its phase by a factor p . It is a Lipschitz continuous operator with for all $z, z' \in \mathbb{C}^2$, $|[z]^p - [z']^p| \leq \max(|p|, 1)|z - z'|$ [5]. This prevents uncontrolled amplifications, and leads to estimators with reduced variance compared to equivalent moments where the phase harmonic operator would be replaced by a standard exponentiation [6].

These WPH moments are able to capture interactions between different scales of X thanks to the phase harmonic operator. Indeed, the covariance between $X \star \psi_{\xi_i}$ and $X \star \psi_{\xi_j}$ vanishes when the wavelets ψ_{ξ_i} and ψ_{ξ_j} have nonintersecting bandpasses, and it is otherwise a function of the power spectrum of X and of the bandpasses of the wavelets [6, 199]. This is a consequence of the following relation (see [6] for a proof):

$$\text{Cov} \left[X \star \psi_{\xi_i}(\mathbf{r}), X \star \psi_{\xi_j}(\mathbf{r} + \tau) \right] = \int S_X(\mathbf{k}) \hat{\psi}_{\xi_i}(\mathbf{k}) \hat{\psi}_{\xi_j}^*(\mathbf{k}) e^{-i\mathbf{k} \cdot \tau} d\mathbf{k}, \quad (\text{IV.16})$$

with S_X the power spectrum of X . With proper p_i and p_j values, the phase harmonic operator can make $[X \star \psi_{\xi_i}]^{p_i}$ and $[X \star \psi_{\xi_j}]^{p_j}$ comparable in the sense that they share common spatial frequencies, allowing an extraction of high-order information through their covariance.

To illustrate the importance of the phase harmonic operator to measure phase alignment between scales, we show in Fig. IV.9 how the amplitude and phase maps of $x \star \psi_{\xi_1}$ and $x \star \psi_{\xi_2}$ compare for two different wavevectors ξ_1 and ξ_2 , with x a map from the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set (left maps, with \tilde{Q} on top and \tilde{U} on bottom). We choose $\xi_1 = \xi_{3,0}$ and $\xi_2 = \xi_{4,0}$, so that ψ_{ξ_1} and ψ_{ξ_2} probe similar orientations but different scales, with ψ_{ξ_2} probing scales that are twice larger than those probed by ψ_{ξ_1} . The amplitude maps show local variations of the signal x filtered at different scales, with naturally coarser variations in $|x \star \psi_{\xi_2}|$ than in $|x \star \psi_{\xi_1}|$. The phase maps present almost periodic oscillations tending to be vertically aligned. However these maps are incoherent in the sense that the average frequency of oscillation of the phase map $\arg(x \star \psi_{\xi_1})$ is approximately twice that of $\arg(x \star \psi_{\xi_2})$. The phase harmonic operator with $k = 2$ transforms $\arg(x \star \psi_{\xi_2})$ into a phase map (bottom right map) that is much more coherent with $\arg(x \star \psi_{\xi_1})$. Consequently, the sample covariance of these maps does not vanish and quantifies the phase alignment between $x \star \psi_{\xi_1}$ and $x \star \psi_{\xi_2}$.

The covariances depend on the variable τ which introduces a relative shift between $X \star \psi_{\xi_1}$ and $X \star \psi_{\xi_2}$. Inspired by [199], we discretize this τ variable by introducing polar coordinates (n, α) defined relatively to the central wavevector associated to each wavelet. Practically, for each wavelet $\psi_{j,\theta}$, we introduce a grid of $\tau_{n,\alpha}$ vectors defined as

$$\tau_{n,\alpha} = 3n2^j \mathbf{u}_{\theta+\alpha}, \quad (\text{IV.17})$$

⁷These moments do not depend on the \mathbf{r} variable because of the homogeneity of X .

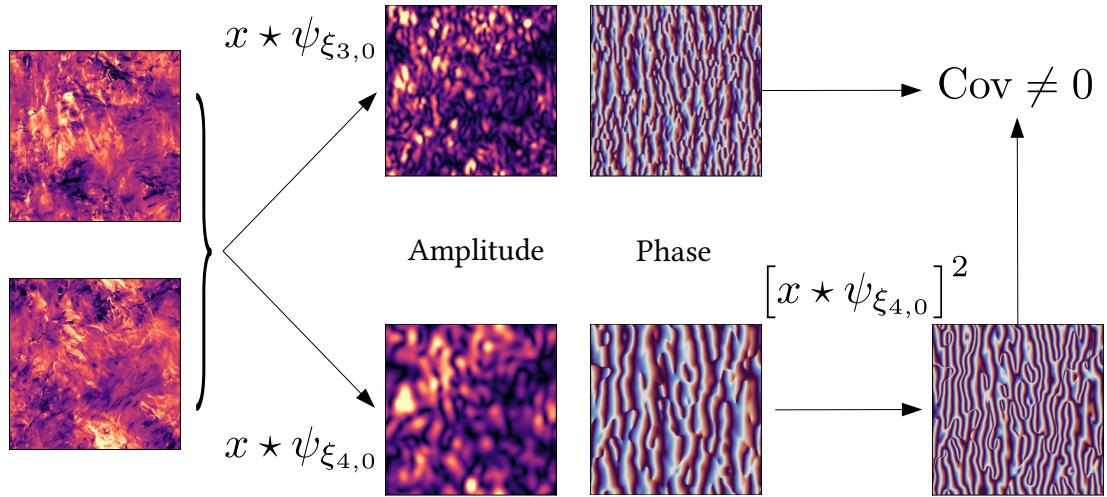


Fig. IV.9 Comparison between the amplitude and phase maps of $x \star \psi_\xi$ for x a complex image shown on the left (real part is above, and imaginary part is below) and two different wavelets $\psi_{3,0}$ and $\psi_{4,0}$. The action of the phase harmonic operator is shown on the phase map of $x \star \psi_{4,0}$.

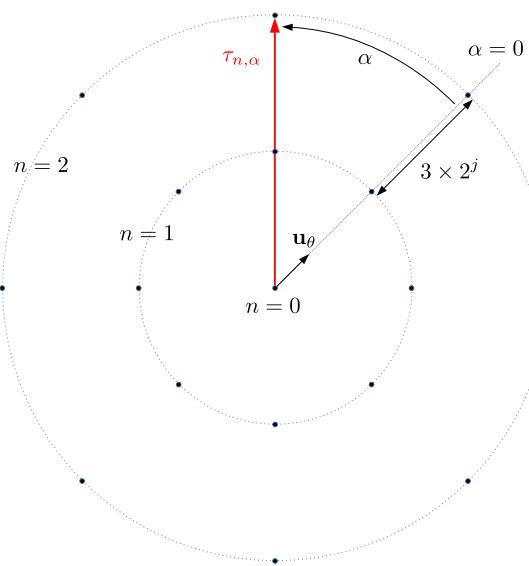


Fig. IV.10 Discrete grid of translation vectors $\tau_{n,\alpha}$ defined relatively to a wavelet $\psi_{j,\theta}$. On this example, we choose $\theta = \frac{\pi}{4}$ and show grid points up to $n = 2$ with $A = 4$.

with $n = \llbracket 0, \Delta_n \rrbracket$ and $\alpha \in \{\alpha_k = \frac{k\pi}{A}, k \in \llbracket 0, 2A - 1 \rrbracket\}$, where Δ_n and A are two integers playing similar roles to J and L but for n and α variables instead of j and θ , respectively (see Fig. IV.10 for an illustration). To avoid a redundancy in the information contained in the coefficients, we discard the translations for which $n > \min(J - 1 - j, \Delta_n)$ similarly to the implementation of [199].

In the following, we choose $\Delta_n = 2$ and $A = 4$, guided by the results of [199]. Note that contrary to [199], α takes values in $[0, 2\pi)$ instead of $(-\pi/2, \pi/2]$. This choice allows to restrict our wavelet bank to wavelets with $\theta \in [0, \pi)$ only when describing real-valued data as explained below.

c) Choice of a subset of WPH moments

In practice, we cannot reasonably estimate the whole set of WPH moments for every possible choice of ξ_1 , ξ_2 , p_1 , p_2 , and $\tau_{n,\alpha}$ variables as this would be obviously computationally excessive and not necessarily useful. Indeed, not all the moments are equally informative, some of them may vanish by construction, or there may be a form of redundancy due to immediate symmetries.

[199] identified a relevant set of WPH moments to build models of real-valued simulated data of the large-scale structure of the Universe⁸. In the present work, we define WPH statistics that are directly inspired from this work while being slightly different. Our choice of subset is made more optimal to describe real-valued data and extended to characterize complex-valued data. We will consider in the following five categories of moments defined as follows:

- the $S^{(1,1)}$ moments, of the form $C_{\xi,1,\xi,1}(\tau) = \text{Cov}[X \star \psi_\xi(\mathbf{r}), X \star \psi_\xi(\mathbf{r} + \tau)]$, at every $\tau_{n,\alpha}$. They measure weighted averages of the power spectrum over the bandpass of ψ_ξ (see Eq. (IV.16)).
- the $S^{(0,0)}$ moments, of the form $C_{\xi,0,\xi,0}(\tau) = \text{Cov}[|X \star \psi_\xi(\mathbf{r})|, |X \star \psi_\xi(\mathbf{r} + \tau)|]$, at every $\tau_{n,\alpha}$. They capture information related to the sparsity of the data in the bandpass of ψ_ξ .
- the $S^{(0,1)}$ moments, of the form $C_{\xi,0,\xi,1}(\tau) = \text{Cov}[|X \star \psi_\xi(\mathbf{r})|, X \star \psi_\xi(\mathbf{r} + \tau)]$, at $\tau = 0$ only. They capture information related to the couplings between the scales included in the same bandpass.
- the $C^{(0,1)}$ moments, of the form $C_{\xi_1,0,\xi_2,1}(\tau) = \text{Cov}[|X \star \psi_{\xi_1}(\mathbf{r})|, X \star \psi_{\xi_2}(\mathbf{r} + \tau)]$, considering $0 \leq j_1 < j_2 \leq J - 1$, at every $\tau_{n,\alpha}$ when $\theta_1 = \theta_2$ and at $\tau = 0$ only when $\theta_1 \neq \theta_2$. They capture information related to the correlation between local levels of oscillation for the scales in the bandpasses associated to ψ_{ξ_1} and ψ_{ξ_2} .
- the C^{phase} moments, of the form $C_{\xi_1,1,\xi_2,p_2}(\tau) = \text{Cov}[X \star \psi_{\xi_1}(\mathbf{r}), [X \star \psi_{\xi_2}(\mathbf{r} + \tau)]^{p_2}]$ with $p_2 = \xi_1/\xi_2 > 1$, considering $0 \leq j_1 < j_2 \leq J - 1$ and $\theta_1 = \theta_2$, at every $\tau_{n,\alpha}$. They capture information related to the statistical phase alignment of oscillations between the scales in the bandpasses associated to ψ_{ξ_1} and ψ_{ξ_2} .

⁸These data are highly non-Gaussian at late times and at scales smaller than $100 h^{-1}\text{Mpc}$.

We identify in the following basic symmetries with respect to the angular variables that will allow us to define optimal ranges for the angular variables, avoiding redundancy in the statistical content of the WPH statistics. We denote, e.g., by $S^{(1,1)}(j, \theta, n, \alpha)$ the moment $C_{\xi_{j,\theta},1,\xi_{j,\theta},1}(\tau_{n,\alpha})$, and use similar notations for the other categories of moments.

For a real-valued statistically homogeneous random field X , and for a choice of wavelets satisfying the symmetry $\psi_{j,\theta+\pi} = \overline{\psi_{j,\theta}}$ (valid when $\hat{\psi}(\mathbf{k}) \in \mathbb{R}$ for all \mathbf{k} , which is notably the case of bump-steerable wavelets ; see Appendix B), we can derive the following list of symmetries:

$$S^{(1,1)}(j, \theta + \pi, n, \alpha) = S^{(1,1)}(j, \theta, n, \alpha), \quad (\text{IV.18})$$

$$S^{(1,1)}(j, \theta, n, \alpha + \pi) = \overline{S^{(1,1)}(j, \theta, n, \alpha)}, \quad (\text{IV.19})$$

$$S^{(0,0)}(j, \theta + \pi, n, \alpha) = S^{(0,0)}(j, \theta, n, \alpha), \quad (\text{IV.20})$$

$$S^{(0,0)}(j, \theta, n, \alpha + \pi) = S^{(0,0)}(j, \theta, n, \alpha), \quad (\text{IV.21})$$

$$S^{(0,1)}(j, \theta + \pi) = \overline{S^{(0,1)}(j, \theta)}, \quad (\text{IV.22})$$

$$C^{(0,1)}(j_1, \theta_1 + \pi, j_2, \theta_2, n, \alpha) = C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha), \quad (\text{IV.23})$$

$$C^{(0,1)}(j_1, \theta_1, j_2, \theta_2 + \pi, n, \alpha) = \overline{C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha + \pi)}, \quad (\text{IV.24})$$

$$C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha + \pi) = \overline{C^{(0,1)}(j_1, \theta_1 + \pi, j_2, \theta_2 + \pi, n, \alpha)}, \quad (\text{IV.25})$$

$$C^{\text{phase}}(j_1, \theta + \pi, j_2, \theta + \pi, n, \alpha) = \overline{C^{\text{phase}}(j_1, \theta, j_2, \theta, n, \alpha + \pi)}. \quad (\text{IV.26})$$

In the case of a complex-valued statistically homogeneous random field X , however, only Eqs. (IV.19) and (IV.21) hold in general.

Proofs of these relations essentially stem from the fact that $[X \star \psi_{j,\theta+\pi}]^p = [\overline{X \star \psi_{j,\theta}}]^p$ for any $p \in \mathbb{R}$. This is the consequence of the commutativity of the complex conjugation with the phase harmonic and convolution operations. Note that due to the peculiar definition of translation vectors $\tau_{n,\alpha}$, which are defined relatively to a given wavelet $\psi_{j,\theta}$ (see Eq. (IV.17)), we have $[X \star \psi_{j,\theta+\pi}]^p(r + \tau_{n,\alpha}) = [\overline{X \star \psi_{j,\theta}}]^p(r - \tau_{n,\alpha})$, where the first $\tau_{n,\alpha}$ vector is implicitly associated to $\psi_{j,\theta+\pi}$ and the second one to $\psi_{j,\theta}$.

For example, let us prove Eq. (IV.25). For X a real-valued random field, we have obviously $X = \overline{X}$, and:

$$C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha + \pi) = \text{Cov}[|X \star \psi_{j_1, \theta_1}(r)|, X \star \psi_{j_2, \theta_2}(r + \tau_{n,\alpha+\pi})] \quad (\text{IV.27})$$

$$= \text{Cov}[|X \star \psi_{j_1, \theta_1}(r)|, X \star \psi_{j_2, \theta_2}(r - \tau_{n,\alpha})] \quad (\text{IV.28})$$

$$= \text{Cov}[|X \star \psi_{j_1, \theta_1}(r)|, \overline{X \star \psi_{j_2, \theta_2+\pi}(r + \tau_{n,\alpha})}] \quad (\text{IV.29})$$

$$= \text{Cov}[\overline{|X \star \psi_{j_1, \theta_1+\pi}(r)|}, \overline{X \star \psi_{j_2, \theta_2+\pi}(r + \tau_{n,\alpha})}] \quad (\text{IV.30})$$

$$= \text{Cov}[|X \star \psi_{j_1, \theta_1+\pi}(r)|, X \star \psi_{j_2, \theta_2+\pi}(r + \tau_{n,\alpha})] \quad (\text{IV.31})$$

$$= \overline{C^{(0,1)}(j_1, \theta_1 + \pi, j_2, \theta_2 + \pi, n, \alpha)}. \quad (\text{IV.32})$$

Proofs of the other symmetries are analogous.

Coefficients that are either equal or related by a complex conjugation operation are said to be redundant. The previous relations show that, in order to avoid redundancy, for real-valued fields, it is sufficient to consider wavelets with $\theta \in [0, \pi)$, while for complex-valued fields, we will need to use the full set of wavelets with $\theta \in [0, 2\pi)$. Moreover, to

	$S^{(1,1)}$	$S^{(0,0)}$	$S^{(0,1)}$	$C^{(0,1)}$	C^{phase}	$L_{j,p}$	Total	Ratio (%)
real field	408	408	56	2944	1768	16	5600	~ 2
complex field	816	816	112	8576	3536	32	13888	~ 5

Table IV.1 – Number of statistical coefficients per class of moments for the parameters used in this work in the cases of both real and complex-valued data. The last column gives the ratio of the total number of coefficients to the number of pixels in a 512×512 pixels image.

avoid further redundancy, for the $S^{(1,1)}$ and $S^{(0,0)}$ moments, it is sufficient to consider $\tau_{n,\alpha}$ vectors with $\alpha \in [0, \pi)$ only for both real and complex-valued fields.

We give in Table IV.1 the resulting number of WPH coefficients for the parameters used in this work ($J = 7$, $L = 8$, $\Delta_n = 2$, $A = 4$) in the cases of both real and complex-valued data. We also give for information the ratio between these numbers of coefficients and the number of pixels in a 512×512 pixels image.

Note that for complex-valued random fields X , one could similarly define pseudo-covariance WPH moments (see Appendix A for a definition of pseudo-covariance). However similar symmetry relations show that these are already contained in regular covariance WPH moments for all categories of moments except C^{phase} moments. To simplify, we will simply ignore pseudo-covariance analogs of C^{phase} moments for complex-valued random fields.

d) Scaling moments

Similarly to [199], we complement the WPH statistics with a small number of coefficients, the estimates of the so-called scaling moments $L_{j,p}$ that better constrain the largest scales that are not probed by the WPH moments. These moments are constructed from a bank of scaling functions $\{\varphi_j\}_{0 \leq j \leq J-1}$, which correspond to dilations of an isotropic Gaussian filter φ defined in Fourier space by the following:

$$\hat{\varphi}(\mathbf{k}) = \exp\left(-\frac{\|\mathbf{k}\|^2}{2\sigma^2}\right), \quad (\text{IV.33})$$

with $\sigma = 0.496 \times 2^{-0.55} \xi_0$ (following [6]). For a real-valued random field X , the scaling moments of X are as follows:

$$L_{j,0} = \text{Cov} [|X \star \varphi_j|, |X \star \varphi_j|], \quad (\text{IV.34})$$

$$L_{j,p} = \text{Cov} [(X \star \varphi_j)^p, (X \star \varphi_j)^p] \quad (\text{for } p > 0). \quad (\text{IV.35})$$

In the case of a complex-valued random field X , we consider the previous scaling moments for both $\text{Re}(X)$ and $\text{Im}(X)$ taken separately.

In this work, we consider $p \in \{0, 1, 2, 3\}$ and $2 \leq j \leq J - 2$ following [199]. The corresponding number of coefficients associated to these moments is also given in Table IV.1.

e) Normalized estimates

We define normalized estimates of the WPH and scaling moments of X based on a single realization x similarly to [6] and [199]. The normalization of these estimates, which has been shown to speed up minimization when sampling from a microcanonical gradient descent model [6], depends on a reference map x_0 .

For a map x , the normalized estimates of the WPH moments $C_{\xi_1, p_1, \xi_2, p_2}(\tau)$ of X , called $\tilde{C}_{\xi_1, p_1, \xi_2, p_2}(\tau)$ are defined as follows:

$$\tilde{C}_{\xi_1, p_1, \xi_2, p_2}(\tau) = \frac{\langle x^{(\xi_1, p_1)}(\mathbf{r}) \overline{x^{(\xi_2, p_2)}(\mathbf{r} + \tau)} \rangle}{\sqrt{\langle |x_0^{(\xi_1, p_1)}|^2 \rangle \langle |x_0^{(\xi_2, p_2)}|^2 \rangle}}, \quad (\text{IV.36})$$

where the brackets stand for a spatial mean on \mathbf{r} , and $x^{(\xi, p)} = [x \star \psi_\xi]^p - \langle [x_0 \star \psi_\xi]^p \rangle$. Note that the definition of $x^{(\xi, p)}$ already depends on x_0 .

Similarly, we define the normalized estimates of the scaling moments of X derived from x by:

$$\tilde{L}_{j,p} = \frac{\langle |x^{(j,p)}|^2 \rangle}{\langle |x_0^{(j,p)}|^2 \rangle}, \quad (\text{IV.37})$$

with

$$x^{(j,0)} = |(x - \langle x_0 \rangle) \star \varphi_j| - \langle |(x - \langle x_0 \rangle) \star \varphi_j| \rangle, \quad (\text{IV.38})$$

$$x^{(j,p)} = ((x - \langle x_0 \rangle) \star \varphi_j)^p - \langle ((x - \langle x_0 \rangle) \star \varphi_j)^p \rangle \quad (\text{for } p > 0). \quad (\text{IV.39})$$

This definition is similar to that of the implementation of [199].

IV.3.2 Generative models of synthetic polarization maps

We produce a synthetic $\tilde{Q} + i\tilde{U}$ map in a completely analogous way to the procedure described in Sect. IV.2.1 but here making use of WPH statistics instead of RWST statistics. We show here results based on the modeling of the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set for better comparison with the results of Sect. IV.2.

This procedure minimizes the following loss \mathcal{L} :

$$\mathcal{L}[x] = \left\| \phi(x) - \phi^t \right\|^2, \quad (\text{IV.40})$$

where ϕ is the operator that computes the WPH statistics of a complex map x as introduced in Sect. IV.3.1, ϕ^t is a set of target WPH statistics, and $\|\cdot\|$ denotes the Euclidean norm in the WPH statistical space. The WPH statistics of x correspond to a vector of complex numbers comprising the normalized estimates of the WPH and scaling moments of x defined in Eqs. (IV.36) and (IV.37). The reference map involved in the normalization is chosen to be the map of the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set that was already taken as reference in Sect. IV.2. The target statistics ϕ^t are averages of the WPH statistics computed for each map of the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set. The normalization of the estimates involved in the derivation of ϕ^t is kept consistent with that of $\phi(x)$.

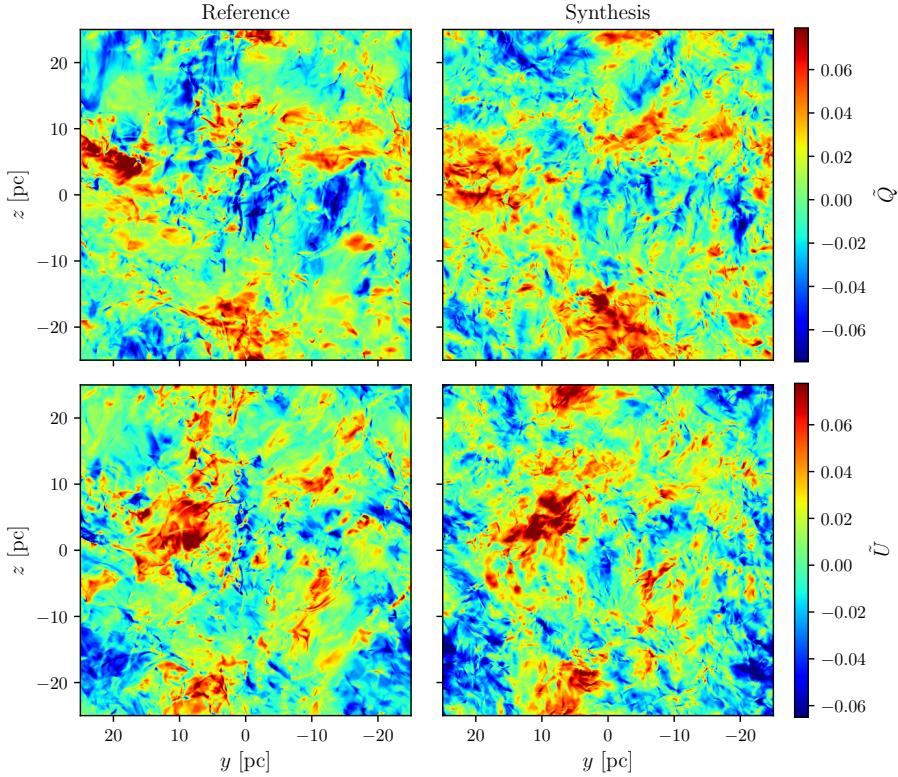


Fig. IV.11 Same as Fig. IV.3 but with a model based on the WPH statistics of the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set. The synthetic map is shown on the right, while the reference map is shown on the left.

The optimization here again starts with a modified realization of a Gaussian white noise, whose largest scales are replaced by those of the reference map as explained in Sect. IV.2.1. In 50 iterations, taking less than 4 minutes on a Tesla V100S GPU with 32 Gb of memory, the loss reaches approximate convergence. We show in Fig. IV.11 the resulting synthesis next to the reference map. This figure is to be compared to Fig. IV.3. Visually, the largest scales between the reference and synthetic maps are consistent, which was obviously expected. The typical filamentary structures of the $\tilde{Q}_\parallel + i\tilde{U}_\parallel$ data set seems to be very well reproduced in the synthesis. However, the density of sharp variations in the synthetic map appears slightly higher than that of the reference map, which by opposition seems more regular, although these differences remain quite subtle and would deserve further quantification. Note that contrary to the synthesis that is derived from RWST measurements, this synthesis does not have any annoying visible numerical artifact such as the checkerboard patterns identified in Fig. IV.4. Consequently, we do not perform any post-processing step as in Sect. IV.2.

IV.3.3 Statistical assessment and comparison with RWST-based models

The quantitative assessment of the quality of this synthesis is similar to that of Sect. IV.2. We show in Figs. IV.12 and IV.13 comparisons of the power spectra and distributions of the pixel values between the reference and synthetic maps, respectively. These figures are

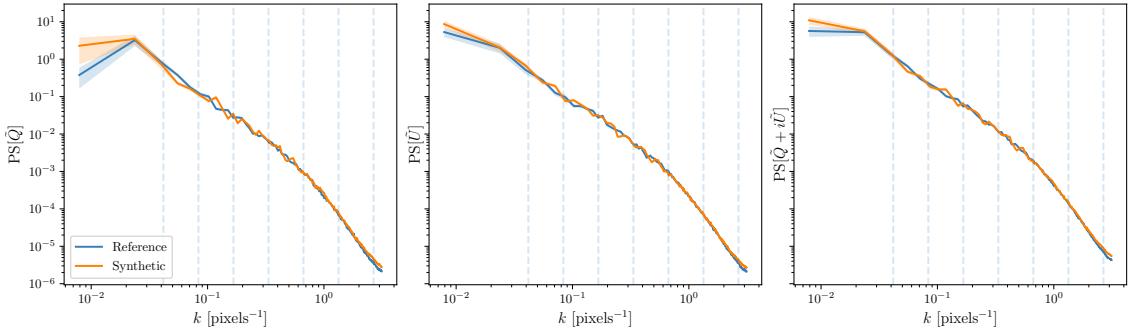


Fig. IV.12 Same as Fig. IV.6 but for a synthesis derived from a WPH-based model.

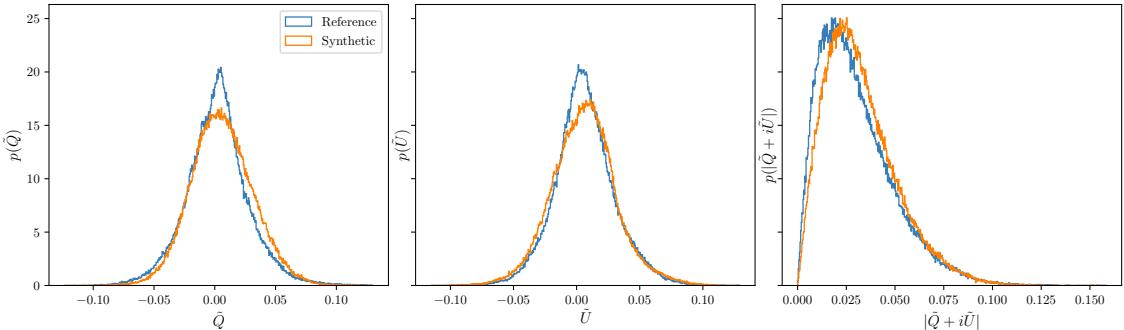


Fig. IV.13 Same as Fig. IV.7 but for a synthesis derived from a WPH-based model.

to be compared to Figs. IV.6 and IV.7.

The agreement of the power spectra is much better compared to the results of Sect. IV.2. In [199], the accuracy of the agreement of power spectrum statistics was shown to be strongly dependent on the Δ_n parameter, and the higher Δ_n is the better these statistics agree. Therefore, this very good agreement must be a consequence of this particular choice of Δ_n (we recall that $\Delta_n = 2$ here). Note however that we discern a very faint discrepancy towards the Nyquist frequency recalling that of Fig. IV.6. This is similarly attributed to gaps in the frequency coverage of the wavelet bank beyond the Nyquist disc and could be corrected easily by extending the wavelet bank.

Distributions of the pixel values exhibit a quite good agreement, notably towards the tails of the distributions. However this agreement remains slightly worse than that of Fig. IV.13. These statistics have been shown in [199] to be partly constrained by the scaling moments, which play an analogous role to the higher-order moments mentioned in Sect. IV.2 to define $\mathcal{L}_{\text{one-point}}$. Further improvement to constrain these distributions is needed, which could involve adjustments in the definition of these scaling moments. Note that there exist other simple strategies to improve this agreement, such as introducing additional constraints in the loss function based on the Kullback-Leibler divergence between the distribution of pixel values of the reference and synthetic maps [201].

IV.4 Discussion

We have introduced in this chapter two types of generative models of simulated $\tilde{Q} + i\tilde{U}$ dust polarization maps, based on their statistical descriptions derived from either the RWST or the WPH. These models correspond to microcanonical gradient descent models that were introduced in [190], and require to solve high-dimensional optimization problems to generate realizations of the models. They allow us to assess the comprehensiveness of a given statistical description by comparing syntheses and original maps by means of a set of summary statistics. Here we have limited such comparisons to two kinds of summary statistics: power spectrum statistics and PDFs of pixel values. We have shown that with either the RWST statistics complemented by additional higher-order statistics or the WPH statistics we reproduce well these summary statistics, although there is still room for improvement. However, this work remains at the proof-of-concept stage. A more refined statistical analysis would be needed, notably employing statistics quantifying the non-Gaussianity of the data. The RWST and the WPH do capture a form of non-Gaussianity, however it would be relevant to determine what kind of standard statistics these do include on the example of simulated dust polarization data. In [199], it was shown that the WPH statistics include an important part of the information captured by the bispectrum or the Minkowski functionals on the example of simulated data of the large-scale structure. Similar analyses would deserve to be conducted on simulated dust polarization data. Finally, we have focused on the modeling of $\tilde{Q} + i\tilde{U}$ maps as the statistical modeling of linear polarization remains the main challenge of upcoming CMB experiments. However let us note that this kind of model may be applied very similarly to intensity maps, or other kind of polarization observables such as $Q + iU$ maps or E or B -modes maps.

This approach would deserve to be thoroughly compared to deep generative models discussed in Chapter II. Perhaps the most salient difference with these approaches is that here we do not need any training step, and are able to define generative models based on a single realization. This avoids the difficulty of defining a relevant training set. This is a true asset in a context where we necessarily face a single realization of our sky, with strong statistical inhomogeneities of the dust emission across the sky.

Chapter V

Statistical denoising and enhanced component separation methods

We have introduced the (R)WST and WPH as descriptive statistics of non-Gaussian data. In the last chapter, we have shown how these may define generative statistical models on the example of simulated $\tilde{Q} + i\tilde{U}$ dust polarization maps. Working with simulated maps was a convenient simplification for our demonstration, allowing obviously much more flexibility and control than with observational data. In particular, we did not have to worry about the relevance of statistical homogeneity assumptions, about annoying boundary conditions, and more importantly for dust polarization, we did not have to worry about the pervasive and destructive impact of data noise. For *Planck* polarization data, and notably at scales and in areas of the sky suited to find primordial *B*-modes (see e.g. the BICEP2 field [2]), noise is a major limitation. It is difficult to model (see e.g. [55] for the *Planck* satellite), and for component separation methods, this naturally leads to higher uncertainties on the resulting estimates of the components.

In this chapter, I address the modeling of the dust polarized emission from noisy observations with the (R)WST or the WPH assuming that the noise distribution is known. Indeed, the accuracy required for the detection of primordial *B*-modes is most certainly unattainable by statistical models derived from physical models of the turbulent magnetized ISM, so that processing of noisy observational data is required. Noise has an impact on statistics, usually by increasing their variability or by introducing potential biases. This needs to be taken into account somehow during the modeling. Here, I introduce a statistical denoising algorithm that aims to retrieve the non-Gaussian statistics of the noise-free dust emission, also called the "true" emission. I implement this algorithm by employing WPH representations of the data, although I could have made use of WST representations in a similar way. We will see that this algorithm may be viewed as a regular image denoising algorithm even if its scope is much wider. Finally, I show that this algorithm extends naturally to more general n -component separation problems, provided statistical models for $n - 1$ of these are known.

This chapter is organized as follows. In a first section, we discuss the impact of noise on the power spectrum, RWST and WPH statistics of simulated $Q + iU$ dust maps, mainly from an empirical perspective. Then, in a second section, we propose a statistical denoising algorithm to mitigate the impact of noise on WPH statistics and retrieve non-

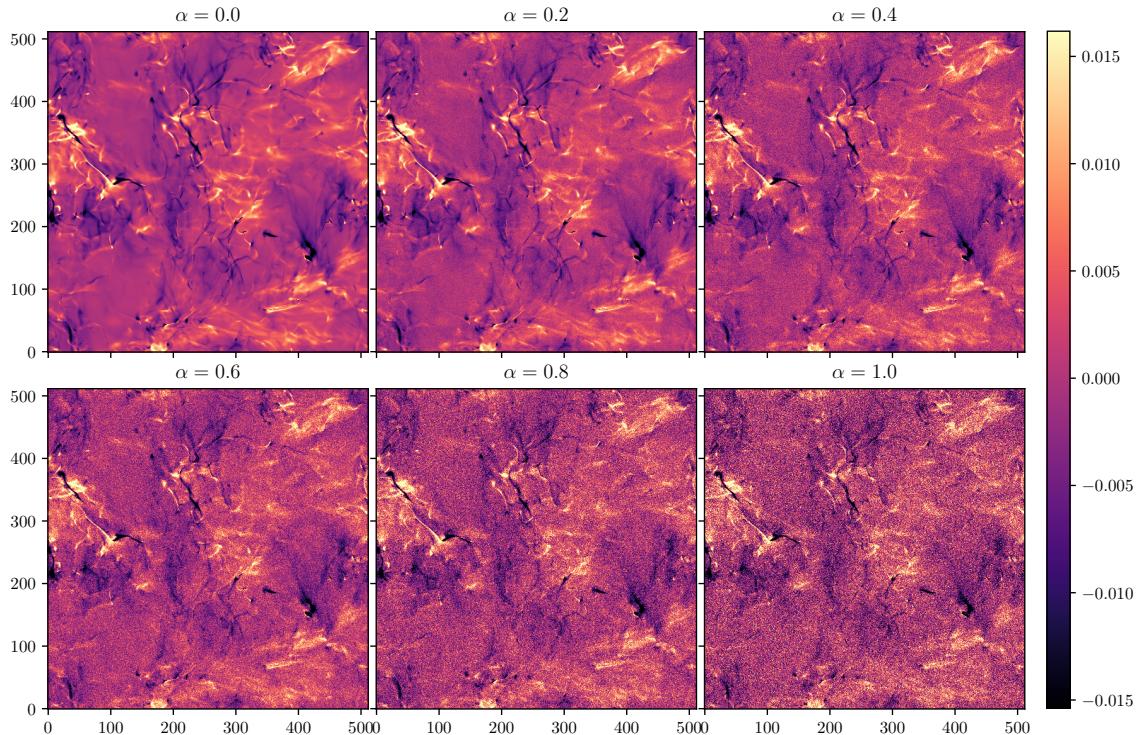


Fig. V.1 Examples of Q maps taken from each data set $Q_\alpha + iU_\alpha$ with $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. α increases from left to right and top to bottom. The noisy maps are built from the same original noise-free Q map shown at the top left. Units are arbitrary, but kept consistent between the maps.

Gaussian statistics of the true emission. After a validation of the algorithm on noisy simulated data, we apply it to the 353 GHz polarized emission observed by *Planck* in the Chamaeleon-Musca field, and discuss our results. Finally, in a third section, we show how this algorithm may be extended to more general component separation problems and present first results on *Planck* polarization observations of the BICEP2 field at 353 GHz.

Part of this chapter reprises the results of [200]. Codes employ PyWST and PyWPH (see Appendix C).

V.1 The impact of noise

In this section, we show empirically how noise impacts the statistics of dust polarization maps on the example of $Q+iU$ noisy simulated data. We derive power spectrum statistics, RWST statistics and WPH statistics for several levels of noise and give first elements to understand these results.

V.1.1 Presentation of the data

We build here noisy mock $Q+iU$ data from the $Q_{||}+iU_{||}$ data set that was introduced in Chapter III. We recall that this data set comprises 14 $Q+iU$ maps denoted by s_1, \dots, s_M

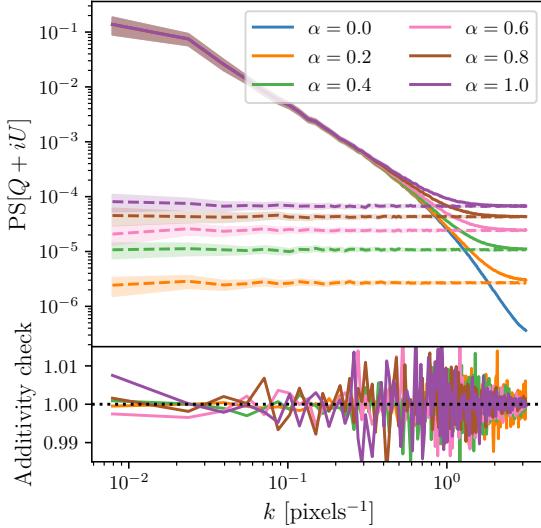


Fig. V.2 *Top panel:* Empirical $Q + iU$ power spectra derived for each of the $Q_\alpha + iU_\alpha$ data sets (solid lines). We also show in dashed lines the power spectra of the noises used to define the $Q_\alpha + iU_\alpha$ data sets. *Bottom panel:* Empirical verification of Eq. (V.4) for each of the noisy $Q_\alpha + iU_\alpha$ data sets ($\alpha > 0$), showing the ratio of the estimate of the left-hand side to that of the right-hand side.

with $M = 14$. We call σ_s the standard deviation of the pixel values computed across the 14 maps. The standard deviation of complex data is taken to be $\sigma_s = \sqrt{\sigma_{\text{Re}(s)}^2 + \sigma_{\text{Im}(s)}^2}$. For each map s_i , we build a noisy map $d_i = s_i + \alpha n_i$, with n_i a realization of a zero-mean complex Gaussian white noise with a standard deviation σ_s (see Appendix A for a definition) and $\alpha \in \mathbb{R}$ a coefficient representing the level of the noise. In this context, we define the signal-to-noise ratio (SNR) of d_i as the ratio of σ_s to the standard deviation of αn_i , so that the SNR of d_i coincides with α^{-1} . For a given α , we call $Q_\alpha + iU_\alpha$ the data set made of the 14 maps d_1, \dots, d_M . Note that we have in particular $Q_0 + iU_0 = Q_\parallel + iU_\parallel$.

In addition to the $Q_0 + iU_0$ data set, we build 5 noisy data sets $Q_\alpha + iU_\alpha$ for $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$. As an illustration, we show in Fig. V.1 how a given Q map is visually affected by these various levels of noise.

V.1.2 Impact on power spectrum statistics

We first discuss the impact of noise on the power spectrum statistics. This is a very instructive example for which an analytical understanding is possible.

For each data set $Q_\alpha + iU_\alpha$ we compute a mean empirical power spectrum by averaging estimates of the power spectrum for each of the 14 maps of the data set. Details on the definition of these estimates can be found in Appendix A. We use a regular binning in k , and the error bars correspond to averages of the standard deviations of the means for each map of the data set. We show in Fig. V.2 the corresponding results in solid lines. We add, in dashed lines, similar results computed for each set of 14 noises involved in the definition of a data set $Q_\alpha + iU_\alpha$.

Let us first remark that the empirical power spectra of the noises are approximately

constant functions, which is expected for white noises. The value of the constant is proportional to α^2 as a consequence of the definition of the power spectrum. Indeed, the autocorrelation function of a stationary random field N verifies for $\alpha \in \mathbb{R}$:

$$\begin{aligned} C_{\alpha N}(\boldsymbol{\tau}) &= \text{Cov} [\alpha N(\mathbf{r}), \alpha N(\mathbf{r} + \boldsymbol{\tau})] \\ &= \alpha^2 \text{Cov} [N(\mathbf{r}), N(\mathbf{r} + \boldsymbol{\tau})] \\ &= \alpha^2 C_N(\boldsymbol{\tau}). \end{aligned} \quad (\text{V.1})$$

We recall that the power spectrum is defined as the Fourier transform of the autocorrelation function (see Appendix A), so that thanks to the linearity of the Fourier transform we get:

$$\text{PS}[\alpha N] = \alpha^2 \text{PS}[N]. \quad (\text{V.2})$$

The empirical power spectrum of $Q_\alpha + iU_\alpha$ evolves with α at scales where the power spectrum of the noise is comparable to that of the true signal, i.e. mainly at small scales. It progressively increases at small scales with the level of the noise, revealing an important property of the power spectrum: it is additive when applied to independent components. Given S and N two independent stationary random fields, with S representing the true signal and N a noise, we have:

$$\begin{aligned} C_{S+N}(\boldsymbol{\tau}) &= \text{Cov} [S(\mathbf{r}) + N(\mathbf{r}), S(\mathbf{r} + \boldsymbol{\tau}) + N(\mathbf{r} + \boldsymbol{\tau})] \\ &= \text{Cov} [S(\mathbf{r}), S(\mathbf{r} + \boldsymbol{\tau})] + \text{Cov} [N(\mathbf{r}), N(\mathbf{r} + \boldsymbol{\tau})] \\ &\quad + \text{Cov} [S(\mathbf{r}), N(\mathbf{r} + \boldsymbol{\tau})] + \text{Cov} [N(\mathbf{r}), S(\mathbf{r} + \boldsymbol{\tau})] \\ &= \text{Cov} [S(\mathbf{r}), S(\mathbf{r} + \boldsymbol{\tau})] + \text{Cov} [N(\mathbf{r}), N(\mathbf{r} + \boldsymbol{\tau})] \\ &= C_S(\boldsymbol{\tau}) + C_N(\boldsymbol{\tau}), \end{aligned} \quad (\text{V.3})$$

since the $\text{Cov} [S(\mathbf{r}), N(\mathbf{r} + \boldsymbol{\tau})]$ and $\text{Cov} [N(\mathbf{r}), S(\mathbf{r} + \boldsymbol{\tau})]$ terms vanish due to the independence of S and N . Finally, the linearity of the Fourier transform leads to:

$$\text{PS}[S + N] = \text{PS}[S] + \text{PS}[N]. \quad (\text{V.4})$$

This additive relation shows that the noise impact on the power spectrum is known a priori, at least in terms of expected values. From an empirical perspective this relation is not exact, although the error tends to zero in the limit of infinite sample¹. We verify empirically Eq. (V.4) for this study by showing in Fig. V.2 (bottom panel) the ratio of the estimate of the left-hand side to that of the right-hand side. This demonstrates that this relation holds empirically at a percent level for each value of α .

For statistics verifying such an additive relation, it is straightforward to define an estimator of the true statistics. Indeed, here $\text{PS}[S]$ can be simply estimated with:

$$\hat{\text{PS}}[S] = \hat{\text{PS}}[S + N] - \hat{\text{PS}}[N], \quad (\text{V.5})$$

where $\hat{\text{PS}}[N]$ and $\hat{\text{PS}}[S + N]$ are estimators of $\text{PS}[N]$ and $\text{PS}[S + N]$, respectively.

Finally, let us mention that provided we have at our disposal two observations $d_1 = s + n_1$ and $d_2 = s + n_2$, with n_1 and n_2 two independent realizations of the noise N , another

¹This is true provided we make use of a consistent estimator of the power spectrum.

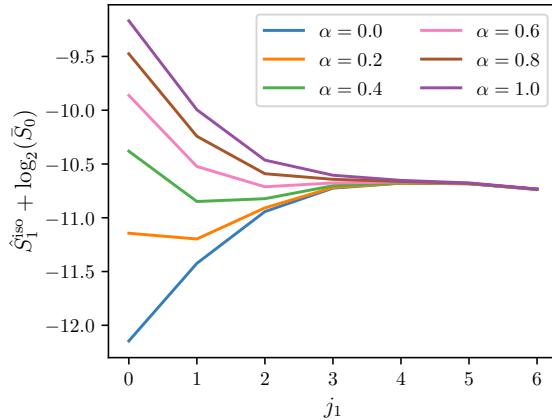


Fig. V.3 $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ coefficients for each of the $Q_\alpha + iU_\alpha$ data sets.

useful unbiased estimate of $\text{PS}[S]$ can be derived from the empirical cross-spectrum of d_1 and d_2 (see Appendix A for a definition). Indeed, one can show that for S , N_1 and N_2 , three independent stationary random fields, the cross-spectrum of $S + N_1$ and $S + N_2$, called $\text{CS}[S + N_1, S + N_2]$ coincides with the power spectrum of S , that is:

$$\text{CS}[S + N_1, S + N_2] = \text{PS}[S]. \quad (\text{V.6})$$

The proof is analogous to that of Eq. (V.4). Thus the empirical cross-spectrum of d_1 and d_2 is an unbiased estimate of the power spectrum of S . For the analysis of CMB data, or more generally in cosmology, this approach is common to circumvent data noise (see e.g. [202]). A major asset of this power spectrum estimation method is that it does not require any noise model. This will be useful in Sect. V.2 when working with observational data.

V.1.3 Impact on RWST statistics

We present a similar analysis of the impact of noise on RWST statistics. We derive the RWST statistics associated with each $Q_\alpha + iU_\alpha$ data set using the same procedure as that introduced in Sect. III.2.4. We focus on the impact of noise on isotropic coefficients as a first step.

a) Impact on first-order coefficients

We show in Fig. V.3 the \hat{S}_1^{iso} coefficients corrected from the normalization related to the \bar{S}_0 coefficients (similarly to Fig. III.12) for each $Q_\alpha + iU_\alpha$ data set.² The results shown in Fig. V.3 are clearly in line with the power spectrum analysis: the $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ coefficients are all the more impacted as the scale j_1 is small, and are increasing functions of the noise level at fixed j_1 value. However, any further analytical study of the noise

²This correction simplifies the interpretation of the \hat{S}_1^{iso} coefficients and avoids potential confusions. Indeed, the \bar{S}_0 coefficients, which correspond to averages of the mean values of the modulus of the 14 maps for each data set (see Sect. III.1.2), increase with the noise level, so that the \hat{S}_1^{iso} coefficients would have shifted downward at large scales for higher noise levels.

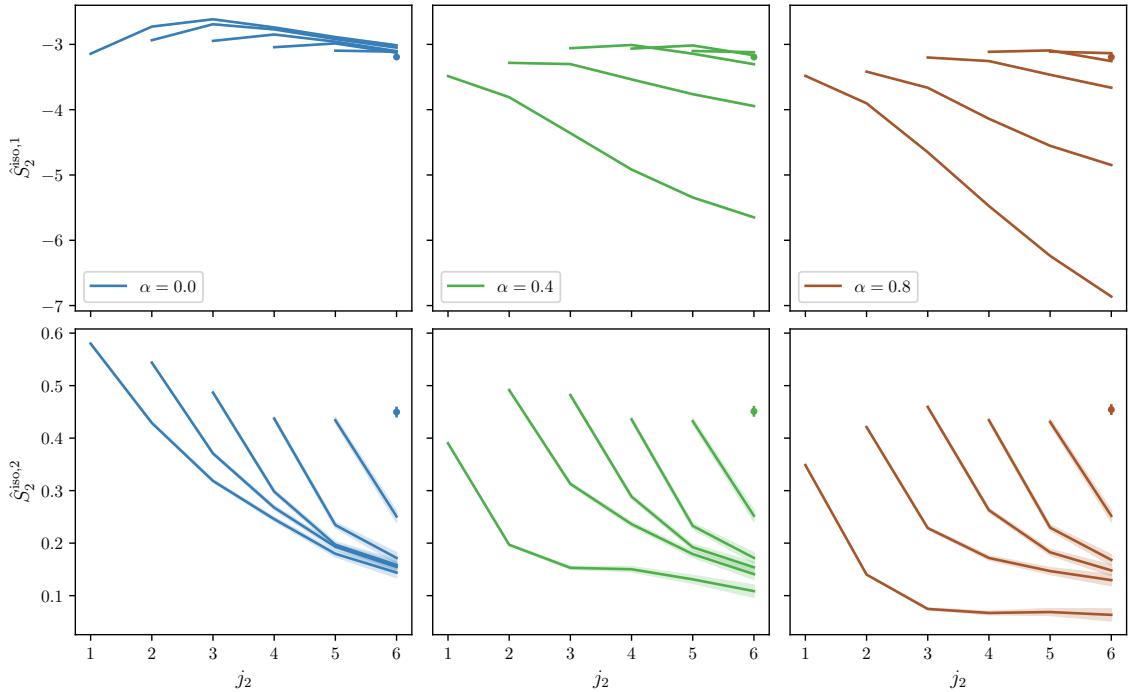


Fig. V.4 $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$ coefficients for the $Q_\alpha + iU_\alpha$ data sets with $\alpha \in \{0.0, 0.4, 0.8\}$. Each curve corresponds to a fixed value of j_1 with j_2 ranging from $j_1 + 1$ to $J - 1 = 6$. For $j_1 = J - 2$, the curve is reduced to a single dot.

impact on RWST first order coefficients turns out to be much more complicated than with the power spectrum statistics. This is mainly due to the fact that these are roughly derived from L^1 norms of wavelet coefficients, which are not as handy as the L^2 norms involved in the definition of the power spectrum. Note however that thanks to the nonexpansivity of the WST (see Sect. III.1.3), the deformation of the RWST statistics remains bounded by the noise level.

b) Impact on second-order coefficients

We show in Fig. V.4 similar results for the $\hat{S}_2^{\text{iso},1}$ and $\hat{S}_2^{\text{iso},2}$ coefficients for a limited choice of α values for better visibility. These second-order coefficients seem more sensitive to noise than first-order coefficients. Looking first at the $\hat{S}_2^{\text{iso},1}$ coefficients, we see that the coefficients with low j_1 values are the most impacted. At high noise level and low j_1 , the evolution of the $\hat{S}_2^{\text{iso},1}$ coefficients as a function of j_2 resembles that of the phase randomized data sets shown in Fig. III.15, with a rapid and approximately linear decrease for increasing values of j_2 . We recall that phase randomization was used as an approximation to Gaussianization. We thus interpret this trend as a sign that the noise "Gaussianizes" the data. Similarly, the impact of noise on the $\hat{S}_2^{\text{iso},2}$ coefficients is the most visible at low j_1 values. It also reminds results of phase randomized data sets shown in Fig. III.15, in particular for high noise levels. Indeed, for $\alpha = 0.8$ and $j_1 = 0$, the $\hat{S}_2^{\text{iso},2}$ tends to zero for increasing j_2 whereas the corresponding $\alpha = 0.0$ curve tends to a strictly positive value. This is similarly interpreted as a sign of the deterioration of the filamentary structure of

the maps.

V.1.4 Impact on WPH statistics

a) Derivation of the statistics

We now derive WPH statistics for each of the $Q_\alpha + iU_\alpha$ data sets. We compute coefficients for $J = 7$, $L = 8$ and $\Delta_n = 0$. This specific choice for Δ_n implies that all the WPH moments are estimated for $\tau = 0$ only. We focus here on estimates of the WPH moments, thus ignoring scaling moments defined in Sect. IV.3.1. Contrary to the analysis of Sect. IV.3, we compute unnormalized estimates of the WPH moments to facilitate interpretations. Given a map x , these are simply defined by:

$$\tilde{C}_{\xi_1, p_1, \xi_2, p_2} = \langle x^{(\xi_1, p_1)} \overline{x^{(\xi_2, p_2)}} \rangle, \quad (\text{V.7})$$

with $x^{(\xi, p)} = [x \star \psi_\xi]^p - \langle [x \star \psi_\xi]^p \rangle$. Unlike their normalized counterparts, note that they do not involve any reference map.

For a given data set $Q_\alpha + iU_\alpha$, we derive WPH statistics as follows. For each map we compute a set of WPH coefficients corresponding to unnormalized estimates of the subset of WPH moments that was introduced in Sect. IV.3.1.³ In order to reduce the number of coefficients, thus simplifying their representation, we choose to focus on the isotropic and parity-invariant properties of the data by reducing the angular dependencies of the coefficients. To do that, similarly to what was done in [199], we average the WPH coefficients over fixed values of $|\theta_2 - \theta_1|$ so that $\delta\theta = |\theta_2 - \theta_1|$ is the only remaining angular dependence of the coefficients, that is we compute:

$$\hat{C}_{j_1, p_1, j_2, p_2, \delta\theta} = \langle \tilde{C}_{\xi_{j_1, \theta_1}, p_1, \xi_{j_2, \theta_2}, p_2} \rangle_{|\theta_1 - \theta_2| = \delta\theta}, \quad (\text{V.8})$$

where $\langle \cdot \rangle_{|\theta_1 - \theta_2| = \delta\theta}$ stands for an average over pairs (θ_1, θ_2) such that $|\theta_1 - \theta_2| = \delta\theta$. We denote by $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, $\hat{S}^{(0,1)}$, $\hat{C}^{(0,1)}$, and \hat{C}^{phase} the coefficients resulting from this angular average for each category of moments. The $C^{(0,1)}$ coefficients were the only ones to involve $\theta_1 \neq \theta_2$ values (see Sect. IV.3.1), so that the $\hat{C}^{(0,1)}$ coefficients are the only reduced coefficients to involve more than one $\delta\theta$ value (for $L = 8$, in practice, $\delta\theta \in \{\frac{k\pi}{8}, k \in \llbracket 0, 7 \rrbracket\}$). On the contrary, the $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, $\hat{S}^{(0,1)}$, and \hat{C}^{phase} coefficients were defined for $\theta_1 = \theta_2$ only, consequently, they do not depend on any angular variable anymore (i.e. $\delta\theta = 0$). Finally, the WPH statistics of the $Q_\alpha + iU_\alpha$ data set correspond to averages of the resulting coefficients across the 14 maps of the data set. Errors on these coefficients correspond to standard deviations of the means.

b) Impact on the $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\hat{S}^{(0,1)}$ coefficients

We show in Fig. V.5 the $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\text{Re}(\hat{S}^{(0,1)})$ coefficients associated with each $Q_\alpha + iU_\alpha$ data set. These only depend on the scale index j . We focus on the real part of the $\hat{S}^{(0,1)}$ coefficients for simplicity, although their imaginary parts could have been shown instead without affecting the following discussion. As discussed in Sect. IV.3.1,

³Note that we first normalize the data set by scaling each map by σ_s^{-1} , although this has no bearing on the discussion.

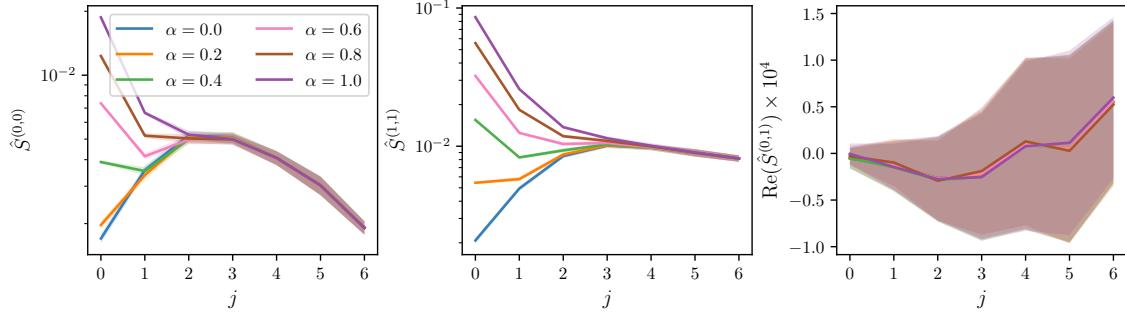


Fig. V.5 $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\text{Re}(\hat{S}^{(0,1)})$ coefficients for each of the $Q_\alpha + iU_\alpha$ data sets.

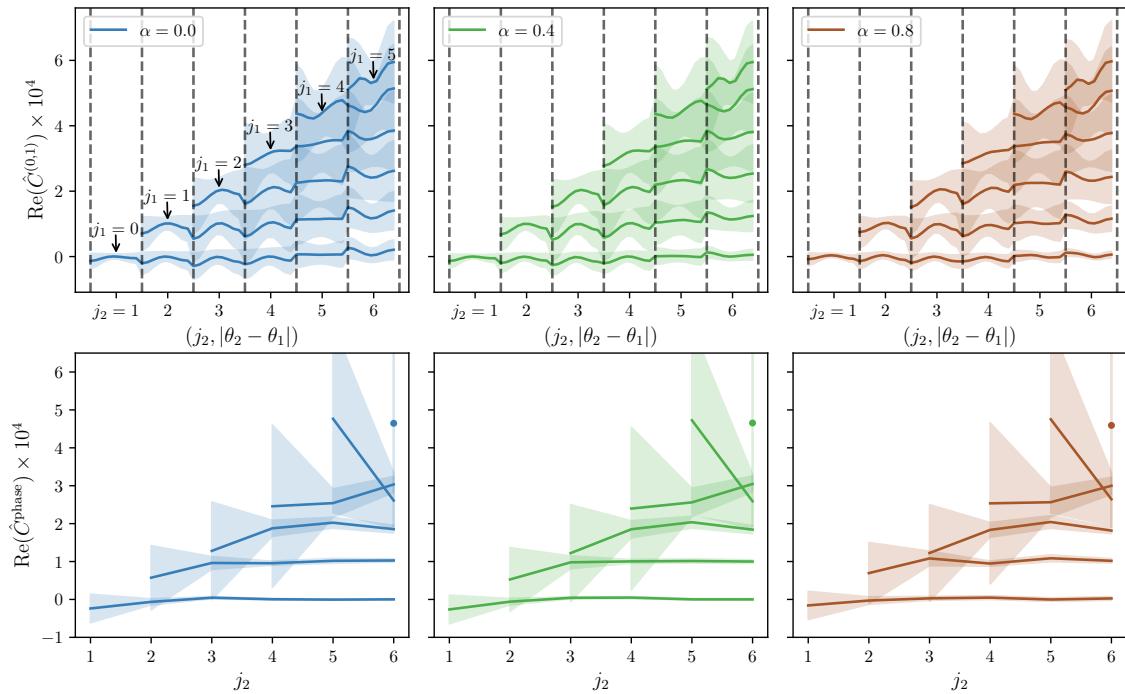


Fig. V.6 $\text{Re}(\hat{C}^{(0,1)})$ (top), and $\text{Re}(\hat{C}^{\text{phase}})$ (bottom) coefficients for the $Q_\alpha + iU_\alpha$ data sets with $\alpha \in \{0.0, 0.4, 0.8\}$. Each curve corresponds to a fixed value of j_1 with j_2 ranging from $j_1 + 1$ to $J - 1 = 6$. For visibility reasons, each curve is shifted upwards by $j_1 \times 10^{-4}$. For the $\text{Re}(\hat{C}^{(0,1)})$ coefficients, the $(j_2, |\theta_2 - \theta_1|)$ dependence is shown in a lexicographical order, and the vertical dashed lines delimit distinct j_2 values. For the $\text{Re}(\hat{C}^{\text{phase}})$ coefficients, at $j_1 = J - 2$ the curve is reduced to a single dot.

the $S^{(1,1)}$ moments are directly related to the power spectrum, corresponding to weighted averages of the power spectrum on the bandpasses of the related wavelets (see Eq. (IV.16)). Consequently, the noise impact on the $\hat{S}^{(1,1)}$ coefficients is essentially equivalent to that already shown in Fig. V.2, although in this representation the x -axis is inverted compared to that of Fig. V.2, as low j values correspond to high k values and vice versa. For similar reasons, the $S^{(1,1)}$ moments are also additive in the sense that for two independent stationary random fields X and N we have:

$$S^{(1,1)}[X + N] = S^{(1,1)}[X] + S^{(1,1)}[N]. \quad (\text{V.9})$$

The proof is completely analogous to the power spectrum case. Therefore, the $\hat{S}^{(1,1)}$ coefficients must approximately verify the same relation. The impact of the noise on the $\hat{S}^{(0,0)}$ coefficients are also related to the results shown for the first order RWST statistics although it is not straightforward. Indeed, a $S^{(0,0)}$ moment of a random field X for $\tau = 0$ can be written:

$$\begin{aligned} S^{(0,0)}(j, \theta) &= \text{Cov}[|X * \psi_{j,\theta}|, |X * \psi_{j,\theta}|] \\ &= \mathbb{E}[|X * \psi_{j,\theta}|^2] - \mathbb{E}[|X * \psi_{j,\theta}|]^2 \\ &= S^{(1,1)}(j, \theta) - S_1(j, \theta)^2, \end{aligned} \quad (\text{V.10})$$

where $S_1(j, \theta)$ is a first order scattering moment associated with X as defined in III.1.2. The $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ coefficients being approximately equivalent to the logarithm of angular averages of the S_1 coefficients, we thus expect the $\hat{S}^{(0,0)}$ coefficients to be consistent with a combination of the results shown for the $\hat{S}_1^{\text{iso}} + \log_2(\bar{S}_0)$ and $\hat{S}^{(1,1)}$ coefficients. Finally, we see that the $\text{Re}(\hat{S}^{(0,1)})$ coefficients seem to be much less impacted by the noise. Also, note that the important error bars on these coefficients show that the variability across samples is much more important for these coefficients than for the $\hat{S}^{(0,0)}$ and $\hat{S}^{(1,1)}$ coefficients.

c) Impact on the $\hat{C}^{(0,1)}$ and \hat{C}^{phase} coefficients

We now show in Fig. V.6 the real parts of the $\hat{C}^{(0,1)}$, and \hat{C}^{phase} coefficients associated with the $Q_\alpha + iU_\alpha$ data sets with $\alpha \in \{0.0, 0.4, 0.8\}$. The $\hat{C}^{(0,1)}$ coefficients depend on $(j_1, j_2, \delta\theta)$, while the \hat{C}^{phase} coefficients depend on (j_1, j_2) only. We focus on the real parts of these coefficients for simplicity. The large error bars on these coefficients also emphasize their important variability. Here again, the coefficients seem to be weakly impacted by noise compared to the $\hat{S}^{(0,0)}$ and $\hat{S}^{(1,1)}$ coefficients although subtle differences remain perceptible. To have a clearer view of these differences, we plot in Fig. V.7 the differences of these coefficients with those of $Q_0 + iU_0$ for the $Q_\alpha + iU_\alpha$ data sets with $\alpha \in \{0.2, 0.4, 0.8\}$. For the $\text{Re}(\hat{C}^{(0,1)})$ coefficients, in the $\alpha = 0.2$ case, the impact seems stronger for coefficients with low j_1 and high j_2 values, while coefficients with high j_1 values are very weakly impacted. When the noise level increases, the impact propagates towards higher j_1 values, and even to the largest ones in the case of $\alpha = 0.8$. For the $\text{Re}(\hat{C}^{\text{phase}})$ coefficients, the impact for a given noise level does not seem to be particularly dependent on scales j_1 and j_2 . The agreement with the $\alpha = 0.0$ case seems to deteriorate uniformly with scales for increasing noise levels.

Note that we expect the noise to "Gaussianize" the data at scales where it is prominent. In [6], it is proved that the $C^{(0,1)}$ and C^{phase} moments of a Gaussian random field

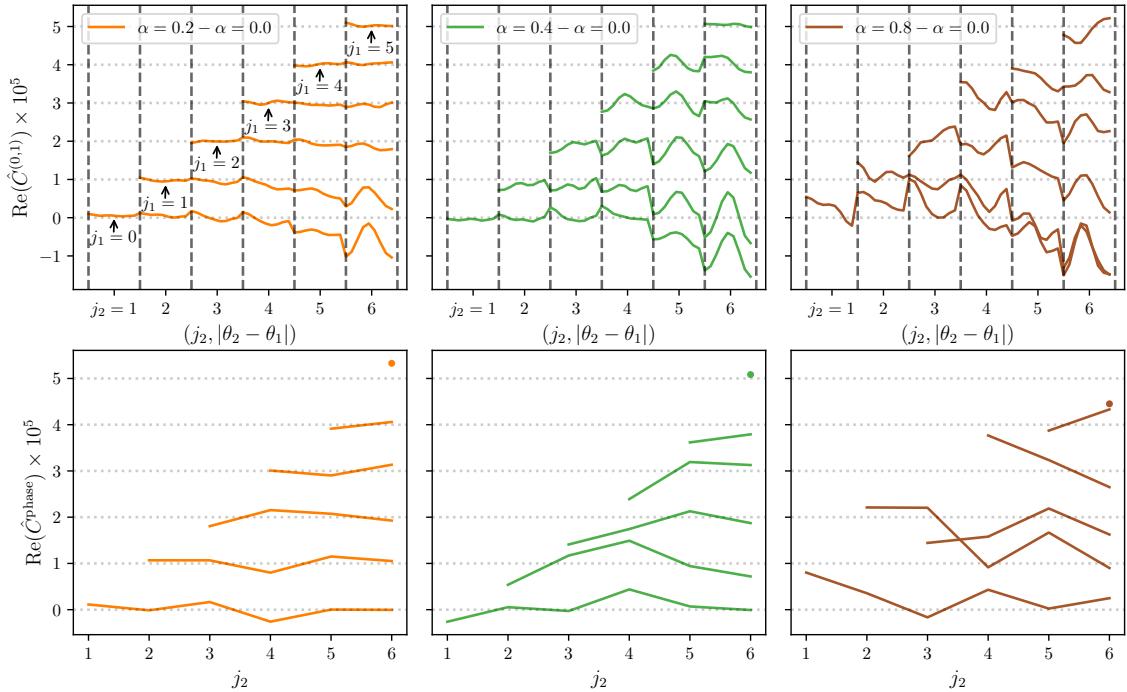


Fig. V.7 $\text{Re}(\hat{C}^{(0,1)})$ (top) and $\text{Re}(\hat{C}^{\text{phase}})$ (bottom) differences between $Q_\alpha + iU_\alpha$ with $\alpha \in \{0.2, 0.4, 0.8\}$ and $Q_0 + iU_0$. Each curve corresponds to a fixed value of j_1 with j_2 ranging from j_1+1 to $J-1=6$. For visibility reasons, each curve is shifted upwards by $j_1 \times 10^{-5}$. The horizontal dotted lines mark null differences per curve. Contrary to Fig. V.6, we do not show the corresponding (large) error bars for better visibility.

vanish provided they involve wavelets with non-overlapping bandpasses. Therefore, we expect the $\hat{C}^{(0,1)}$, and \hat{C}^{phase} coefficients affected by the noise to tend towards zero for increasing noise level. Further examination of these results indicates such a trend.

These numerical results would deserve a complementary mathematical analysis (as far as this is possible), especially to understand the origin of the higher variability across samples of the $\hat{S}^{(0,1)}$, $\hat{C}^{(0,1)}$ and \hat{C}^{phase} coefficients, as well as an explanation for their apparent reduced sensitivity to noise. However, this analysis is beyond the scope of this work, and will have to be undertaken in further studies.

V.2 A statistical denoising algorithm with the WPH

Contrary to power spectrum statistics, (R)WST or WPH statistics are not generally additive when derived from a sum of independent components. For noisy data, we thus do not have a straightforward way to define estimates of the noise-free emission statistics. To address this problem, we introduce in this section a statistical denoising method based on the WPH statistics. It is designed to retrieve the non-Gaussian statistical properties of the noise-free dust emission. However, we will see that it also provides denoised maps,

so that this algorithm may also be used as a regular image denoising algorithm. For the modeling of the dust emission, the motivation for retrieving the statistics of the noise-free emission from noisy data is to define a generative statistical model similar to those introduced in the previous chapter that would not be biased by noise. This is a necessary step for the modeling of dust polarization data where noise cannot be ignored.

After a presentation of our statistical denoising algorithm, we validate it on mock data combining a simulated $Q + iU$ dust map and a *Planck* noise map. Then, we show how it applies on actual observational data, namely the 353 GHz polarized emission observed by *Planck* in the Chamaeleon-Musca field. We also give first elements of comparison of this algorithm with other denoising methods. We finally conclude this section with a more general discussion.

Note that an important part of this section directly reprises results of [200].

V.2.1 Method and validation

a) Description of the method

Basis of the method. We observe a noisy map d that is modeled as follows:

$$d = s + n, \quad (\text{V.11})$$

with s being the target truth map and n being an additive noise signal. The noise n is an unknown realization of a random field N . We assume that we know N , meaning that we are able to generate as many independent realizations of it as needed. We call $\{n_1, \dots, n_M\}$ M of these.

We introduce a method to retrieve the statistical properties of s while denoising d . This statistical denoising consists in iteratively building a denoised map \tilde{s} , starting from d , such that the $\{\tilde{s} + n_i\}_i$ maps and the d map are "sufficiently close" in a given statistical space. This algorithm is inspired by the computation of realizations of microcanonical gradient descent models that were discussed in the previous chapter. In the following, we call ϕ the operator that computes a set of summary statistics from a given map. The algorithm consists in minimizing the following loss function:

$$\mathcal{L}_M(u) = \frac{1}{M} \sum_{i=1}^M \|\phi(u + n_i) - \phi(d)\|^2, \quad (\text{V.12})$$

where $\|\cdot\|$ denotes the Euclidean norm. We choose $u_0 = d$ to initialize the optimizer. The denoised map \tilde{s} corresponds to an approximate minimum obtained by performing this optimization in pixel space, using an L-BFGS optimizer [195]. We note that a limitation of this algorithm relies on the (ideal) assumption that we know N . In practice, any non-modeled statistical property of the noise will be considered to be part of the signal s .

Choice of the operator ϕ . The choice of the operator ϕ is obviously paramount to the quality of the method and must be tailored to the properties of s and n . In the context of ISM polarization data, we expect s to be relatively regular and to exhibit non-Gaussian signatures due to the interactions between scales (e.g., filamentary structures), while the noise is expected to be highly irregular and close to a (possibly spatially-varying) Gaussian white noise.

$S^{(1,1)}$	$S^{(0,0)}$	$S^{(0,1)}$	$C^{(0,1)}$	C^{phase}	$L_{j,p}$	Total	Ratio (%)
960	960	128	11776	5056	40	18920	~ 7

Table V.1 – Number of statistical coefficients per class of moments for the parameters of complex-valued data. The last column gives the ratio of the total number of coefficients to the number of pixels in a 512×512 pixels image.

In this work, the ϕ operator computes WPH statistics. We have seen in the last chapter that WPH statistics are able to define satisfactory generative models of dust polarization maps. Moreover, they have been shown to be relatively exhaustive statistics to characterize non-Gaussian data in [199] as explained in Sect. IV.3.1. We could have chosen similarly (R)WST statistics, however denoising experiments using the WST have demonstrated a higher probability for the resulting denoised maps to exhibit annoying numerical artifacts. Here, we apply ϕ to complex-valued polarization maps $Q + iU$. This operator computes normalized WPH statistics as described in Sect IV.3.1. Contrary to what was done in the last chapter, here we choose $J = 8$, thus increasing the range of scales probed by our statistics. We give in Table V.1 the resulting number of coefficients per class of moments for a given $Q + iU$ map.

Two-step procedure. We have noticed that the reference map used to normalize the WPH coefficients (see Sect. IV.3.1) plays an important role for the quality of the statistical denoising. Indeed, this normalization conditions the relative importance of the coefficients during the optimization. It is not clear yet how to best choose this reference map. However, we propose in the following a denoising procedure in two steps where the purpose of the first step is precisely to find a suitable reference map. This procedure has been proven empirically efficient to retrieve the power spectrum and the PDFs of the increments of the truth map s , as we shall see.

The denoising procedure is as follows. We perform a first denoising of d using a ϕ operator that only takes into account the estimates of the $S^{(1,1)}$ and $S^{(0,0)}$ moments, and normalizing these with the reference map $u_0 = d$. This first step yields a map \tilde{s}_0 , which has notably a power spectrum much closer to that of the truth map s . Then, we perform a second denoising of d using a ϕ operator that includes the whole set of estimates of WPH moments and scaling moments, but here normalizing these estimates with $u_0 = \tilde{s}_0$. We call \tilde{s} the output of this second step.

b) Validation on a simulation

In this section, we assess the performance of our denoising algorithm by applying it to mock data $d = s + n$, emulating a noisy $Q + iU$ *Planck* polarization signal. We show figures of merit based on power spectra, probability density functions (PDFs) of the increments of the maps, and WPH statistics. For the power spectra and the PDFs of the increments, we only show results for Q for simplicity, although those for U are similar.

Building a simulated noisy map. Here s is a simulated $Q + iU$ map taken from the $Q_\perp + iU_\perp$ data set introduced in Sect. III.2. We simulate a *Planck* observation of dust

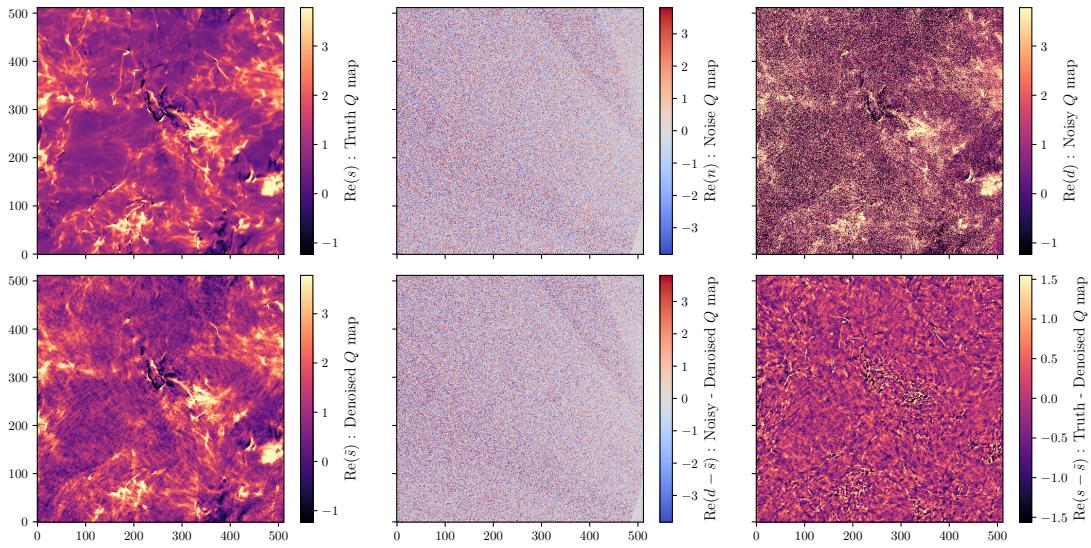


Fig. V.8 *Top row:* Simulated Q maps corresponding to the truth s (left), the noise n (middle), and the resulting noisy map d (right). *Bottom row:* Denoised simulated Q map $\text{Re}(\tilde{s})$ (left) next to the difference noisy-denoised $\text{Re}(d - \tilde{s})$ (middle) and denoising error $\text{Re}(s - \tilde{s})$ (right). Units are arbitrary, but kept consistent between the maps.

polarization using *Planck* instrumental noise maps introduced further in Sect. V.2.2 [55]. We have a total of 300 realizations of this noise at our disposal. We pick one of these, called n , to build d , and we use the remaining $M = 299$ noise maps, labeled $\{n_1, \dots, n_M\}$, for the denoising algorithm.

Here we define the SNR of d as the ratio of the standard deviations of $\text{Re}(s)$ and $\text{Re}(n)$. We adjust this SNR by scaling s , so that the impact of the noise on the power spectrum "resembles" that on the *Planck* map presented in Sect. V.2.2. This is not straightforward since the power spectrum of the simulated map s has a different slope from that estimated for the noise-free emission from the *Planck* map (see Figs. V.9 and V.14). We decide to adjust the SNR so that the scale k_i , at which the power spectra of n and s intersect, coincides with the one from the *Planck* map. Figs. V.9 and V.14 show that $k_i/(2\pi) \approx 0.8 \text{ px}^{-1}$. This procedure leads to $\text{SNR} \approx 0.4$ for the simulated map, which is more than twice lower than the SNR of the observational map discussed in Sect. V.2.2.

We show in Fig. V.8 (top) the simulated Q maps corresponding to the truth s (left), the noise n (middle), and the resulting noisy map d (right). The map $\text{Re}(s)$ exhibits coherent structures such as filaments and large smooth regions that are characteristic of its non-Gaussian statistical properties. On the other hand, $\text{Re}(n)$ seems close to an inhomogeneous white Gaussian noise. The spatial inhomogeneity appears as variations of the local standard deviation due to the *Planck* satellite scanning strategy. Finally, in $\text{Re}(d)$, coherent structures at intermediate and small scales are hard or impossible to identify.

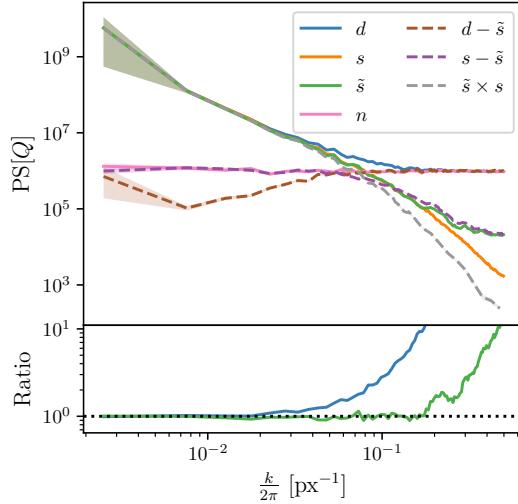


Fig. V.9 Q maps power spectra for d , s , \tilde{s} , n (all in solid lines), $d - \tilde{s}$, and $s - \tilde{s}$, and cross-spectrum between the Q maps of \tilde{s} and s (all in dashed lines). In the bottom panel, we also show the ratio of the power spectra of d and \tilde{s} with that of s .

Visual results. We apply our denoising method to the simulated noisy map d and show, in Fig. V.8 (bottom), the resulting denoised Q map $\text{Re}(\tilde{s})$ (left) next to the difference between the noisy and denoised maps $\text{Re}(d - \tilde{s})$ (middle) and the denoising error $\text{Re}(s - \tilde{s})$ (right). The map $\text{Re}(\tilde{s})$ shows that the noise level has been drastically reduced and that we are able to recover the filamentary structure down to a minimum scale. We can identify the smooth regions of $\text{Re}(s)$ even if there still remains a visible noise. The similarities between $\text{Re}(d - \tilde{s})$ and $\text{Re}(n)$ are striking. The local variations of the standard deviation of the noise are clearly recovered, demonstrating that the inhomogeneity of the noise is not an issue for our method. Finally, $\text{Re}(s - \tilde{s})$ exhibits some remaining structures that match the thinnest filaments appearing in $\text{Re}(s)$, on top of a more diffuse background. This indicates that down to a minimum scale, below which the algorithm struggles to recover features, most of the structures are efficiently reconstructed.

Quantitative results: power spectrum. Figure V.9 compares the power spectra of the six maps shown in Fig. V.8 plus the cross-spectrum between the denoised and truth maps. These power spectra are estimated as usual (see Appendix A). The cross-spectrum is computed in a similar way, with a linearly-spaced binning up to $k/(2\pi) = 0.14 \text{ px}^{-1}$ and logarithmically spaced bins above in order to lower the statistical variance⁴. We first point out that the power spectrum of d coincides with the sum of those of s and n because of the statistical independence between s and n . The power spectra of $\text{Re}(\tilde{s})$ and $\text{Re}(s)$ are in very good agreement with each other up to 0.18 px^{-1} , at which scale the noise power is ten times that of the signal. At smaller scales, where the noise dominates the

⁴Note that this empirical cross-spectrum, called $\text{CS}[\text{Re}(s), \text{Re}(\tilde{s})]$, is related to the empirical power spectra by the following relation: $\text{PS}[\text{Re}(s - \tilde{s})] = \text{PS}[\text{Re}(s)] + \text{PS}[\text{Re}(\tilde{s})] - 2\text{CS}[\text{Re}(s), \text{Re}(\tilde{s})]$.

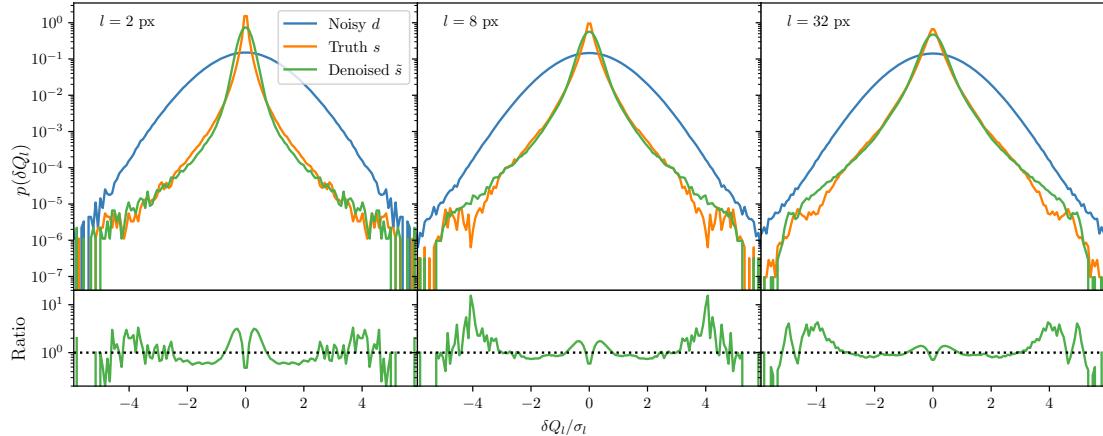


Fig. V.10 PDFs of the increments of Q for d (noisy), s (truth), and \tilde{s} (denoised), computed for three logarithmically spaced lags going from 2 to 32 pixels. For each lag, the increments were normalized by the standard deviation σ_l of the Gaussian that fits the core of the PDF of the noisy map. In the bottom panels, we also show the ratio of the PDFs of \tilde{s} with those of s for each lag.

signal even more, our algorithm is not able to retrieve the true power spectrum but gets closer to it. The power spectrum of $\text{Re}(d - \tilde{s})$ coincides with that of $\text{Re}(n)$ at small scales, and this agreement progressively worsens toward larger scales. This large-scale behavior shows that our algorithm does not remove the already negligible noise, as shown by the superposition of the power spectrum of $\text{Re}(s - \tilde{s})$ and that of $\text{Re}(n)$ at these scales. The cross-spectrum $\tilde{s} \times s$ is slightly below the power spectrum of s at intermediate scales and this discrepancy increases toward the smallest scales. At intermediate scales, where the power spectrum of $\text{Re}(\tilde{s})$ matches that of $\text{Re}(s)$, we suspect this discrepancy to stem from differences between the phases of the Fourier components of s and \tilde{s} that would deserve a further quantification. Nevertheless, the production of a denoised map whose power spectrum coincides with that of s , even though n is ten times more powerful than s , and that retains a significant correlation with s is a striking success of our method.

Quantitative results: PDFs of the increments. To better characterize the non-Gaussianity of \tilde{s} , we compute the PDFs of the increments of Q for the noisy, denoised, and truth maps, and we plot them in Fig. V.10 for three scalar lags (see Sect. II.1.3). The increment $\delta Q_l(\mathbf{r})$ for a scalar lag l and at a position \mathbf{r} corresponds to the set of differences $\delta Q_l(\mathbf{r}) = Q(\mathbf{r}) - Q(\mathbf{r} + \mathbf{l})$ with $l \leq |\mathbf{l}| < l + 1$ in pixel units. Contrary to the case of the noisy map, the distributions of increments for $\text{Re}(s)$ are far from Gaussian for every lag. This is a clear signature of the non-Gaussianity of the data as we expect Gaussian-distributed increments for homogeneous Gaussian data. Our method recovers these statistics and their non-Gaussian tails with limited distortion for each lag, demonstrating its efficiency in retrieving non-Gaussianity in the data.

Quantitative results: WPH. We finally conclude this validation by a WPH analysis. We compare the WPH statistics of d , s , and \tilde{s} . The computation of the WPH statistics is

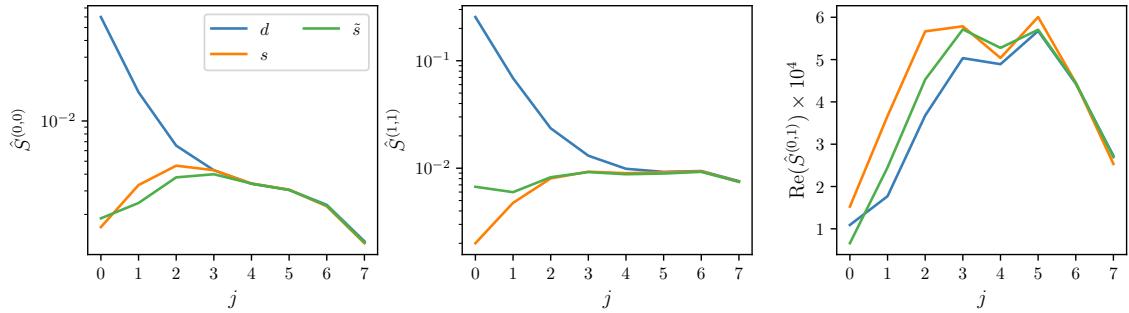


Fig. V.11 $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\text{Re}(\hat{S}^{(0,1)})$ coefficients for d , s , and \tilde{s} .

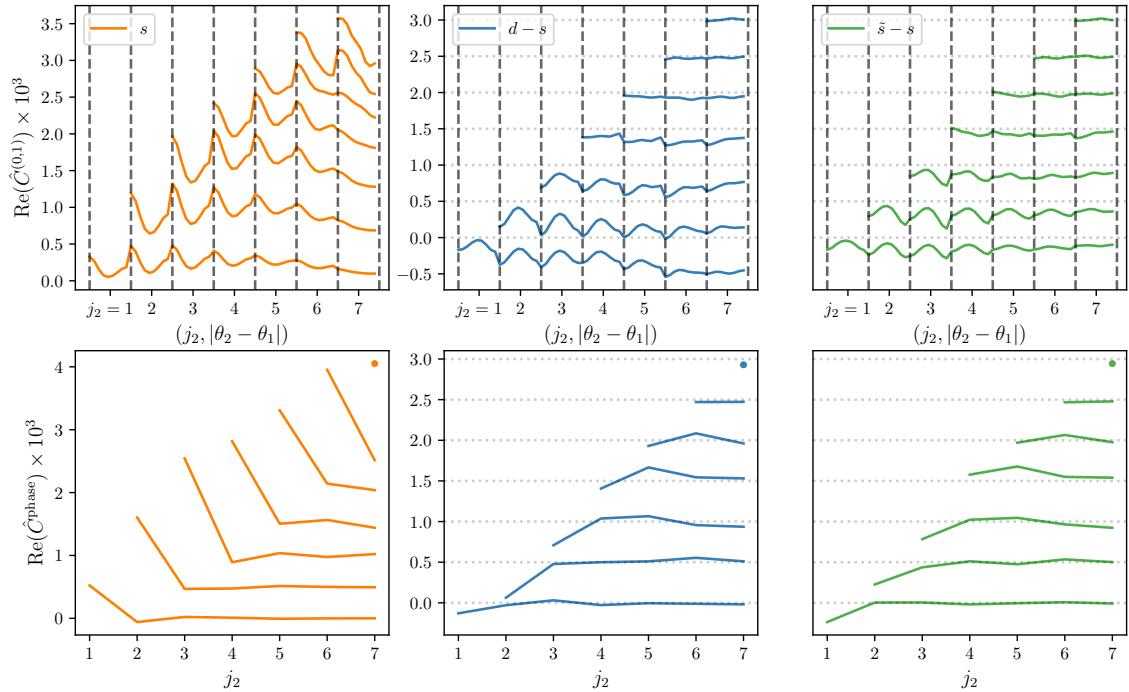


Fig. V.12 $\text{Re}(\hat{C}^{(0,1)})$ (top), and $\text{Re}(\hat{C}^{\text{phase}})$ (bottom) coefficients for d , s , and \tilde{s} . We plot the coefficients corresponding to s in the left column, and the differences of coefficients between d and s (\tilde{s} and s) in the middle (right) column, respectively. Each curve corresponds to a fixed value of j_1 with j_2 ranging from $j_1 + 1$ to $J - 1 = 6$. For visibility reasons, each curve is shifted upwards by $j_1 \times 5 \times 10^{-4}$. For the $\text{Re}(\hat{C}^{(0,1)})$ coefficients, the $(j_2, |\theta_2 - \theta_1|)$ dependence is shown in a lexicographical order, and the vertical dashed lines delimit distinct j_2 values. For the $\text{Re}(\hat{C}^{\text{phase}})$ coefficients, at $j_1 = J - 2$ the curve is reduced to a single dot. In the middle and right columns, the horizontal dotted lines mark null differences for each curve.

similar to what was done in Sect. V.1.4, and we also derive equivalent angular-averaged coefficients to simplify their representation. We show in Fig. V.11 the corresponding $\hat{S}^{(0,0)}$, $\hat{S}^{(1,1)}$, and $\text{Re}(\hat{S}^{(0,1)})$ coefficients. As expected, the $\hat{S}^{(1,1)}$ coefficients show results consistent with Fig. V.9. The true $\hat{S}^{(1,1)}$ coefficients are well retrieved down to $j = 2$. Below this scale, the agreement between \tilde{s} and s worsens and shows that the denoised map \tilde{s} has a significant excess of power at $j = 1$ and $j = 0$. For the $\hat{S}^{(0,0)}$ coefficients, the denoised map has been significantly corrected from the bias introduced by the noise, although small discrepancies remain for $j \leq 3$. Finally, we see on the plot of the $\text{Re}(\hat{S}^{(0,1)})$ coefficients that the denoising globally mitigates the impact of the noise on these coefficients even if this remains subtle.

We plot in Fig. V.12 analog results for the $\text{Re}(\hat{C}^{(0,1)})$ and $\text{Re}(\hat{C}^{\text{phase}})$ coefficients. We display in the left column the coefficients corresponding to the truth map s , and in the middle and right columns, we focus on the differences of the coefficients associated with d and s (middle), and those associated with \tilde{s} and s (right). On these differences, the horizontal dotted lines mark the levels of null differences for each j_1 value. For the $\text{Re}(\hat{C}^{(0,1)})$ coefficients, the denoising clearly improves these differences although a significant bias remains up to $j_1 = 3$ coefficients. The impact of the denoising on the $\text{Re}(\hat{C}^{\text{phase}})$ coefficients is much less visible, and there does not seem to be any significant improvement for the retrieval of the truth statistics.

V.2.2 Application to *Planck* polarization data

a) Presentation of the data

We now apply our denoising method to a $Q + iU$ polarization map of the Chamaeleon-Musca region observed at 353 GHz with the *Planck* satellite (PR3 data⁵ [55]). At this frequency, the CMB is negligible and the dominant components are the dust emission and the noise (see Fig. 34 of [29]). We consider the Q and U maps corresponding to the full mission, and those corresponding to the two half-missions. We also make use of the 300 end-to-end simulated Q and U maps of the noise and the systematics of the instrument for the full mission (the ones used in Sect. V.2.1). We project all of these maps on 512×512 grids with a pixel size of $2.35'$, centered on the region of the Chamaeleon-Musca clouds at Galactic coordinates $(l, b) = (300.26^\circ, -16.77^\circ)$. We use a Gnomonic projection through the HEALPix/healpy⁶ package [121, 203].

In Fig. V.13, we show the projected full-mission Q and U maps that we aim to denoise (left column), as well as the denoised map discussed in the next section (middle column).

b) Denoising results

We apply the denoising method presented in Sect. V.2.1 to the full mission $Q + iU$ map, using the corresponding 300 noises. Figure V.13 (middle column) shows the resulting denoised Q and U maps, respectively. The overall noise level has been clearly mitigated although subtle residuals of the patterns due to the scanning strategy remain, and we can now discern a more complex variety of structures even in the regions where the signal is weak.

⁵https://wiki.cosmos.esa.int/planck-legacy-archive/index.php/Main_Page

⁶<http://healpix.sourceforge.net>

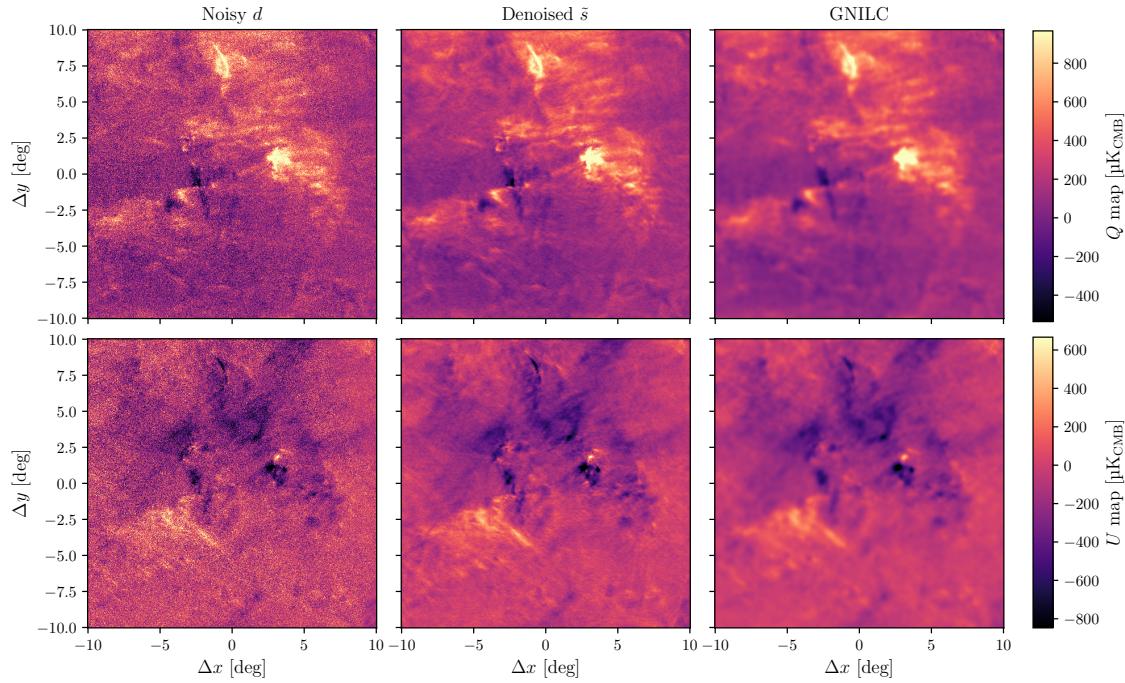


Fig. V.13 Noisy (left column) and denoised (middle column) Q (top row) and U (bottom row) maps of the Chamaeleon-Musca region as observed by the *Planck* satellite at 353 GHz. The noisy maps are denoised as described in Sect. V.2.1. We also show the corresponding GNILC maps (right column) for reference.

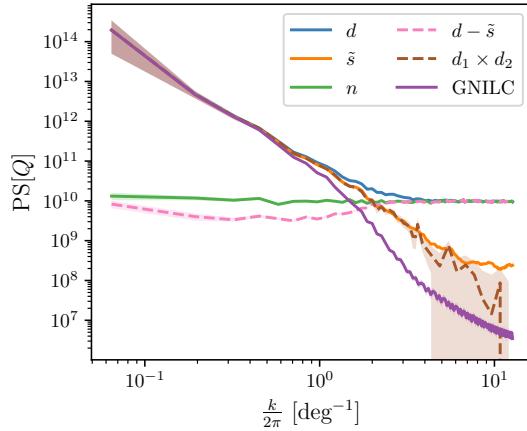


Fig. V.14 Q maps power spectra for d , \tilde{s} , n (all in solid lines), and $d - \tilde{s}$, and cross-spectrum between d_1 and d_2 (both in dashed lines). This cross-spectrum gives an unbiased estimate of the power spectrum of the dust emission. Also shown in a solid line is the power spectrum of the GNILC Q map, corresponding to the same Chamaeleon-Musca field.

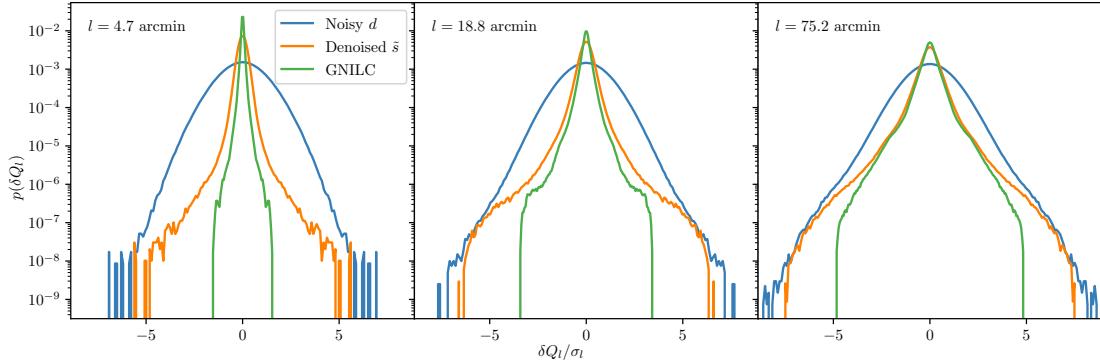


Fig. V.15 PDFs of the Chamaeleon-Musca Q maps increments for d (noisy), \tilde{s} (denoised), and GNILC map, computed for three logarithmically spaced lags ranging from $4.7'$ to $75.2'$. For each lag, the increments were normalized by the standard deviation σ_l of the Gaussian that fits the core of the PDF of the noisy map.

In Fig. V.14, we show a power spectrum analysis of the denoising of the Q map, comparing the power spectra of the noisy and denoised maps $\text{Re}(d)$ and $\text{Re}(\tilde{s})$, of one noise map $\text{Re}(n)$, and of the difference between the noisy and denoised maps $\text{Re}(d - \tilde{s})$, as well as the cross-spectrum between the two half-missions maps d_1 and d_2 .⁷ This cross-spectrum gives an unbiased estimate of the power spectrum of the true signal since the noise signals are independent (see Sect. V.1.2). It is satisfactory to see that the power spectrum of $\text{Re}(\tilde{s})$ is consistent with this cross-spectrum for scales down to $k/(2\pi) \sim 3.4 \text{ deg}^{-1}$. Also, $\text{Re}(d - \tilde{s})$ behaves similarly to what we have observed in Sec. V.2.1, with a power spectrum consistent with that of $\text{Re}(n)$ when the noise dominates the dust signal, and falling below what is expected when the emission of the dust begins to dominate.

Figure V.15 compares the PDFs of the increments of Q for the noisy and denoised maps, and for the same three scalar lags as in Fig. V.10. Based on the results of Sect. V.2.1, we expect the PDFs of the denoised map to be good estimates for the statistics of the truth map. These results show important differences between the noisy and denoised maps that are again a signature of the non-Gaussianity of the true signal. We find similar results for the increments of U .

V.2.3 Comparison to other methods

The literature on image denoising is rich and abundant (see e.g. [204, 205, 206]), and a thorough comparison of our method with other denoising algorithms would be beyond the scope of this work. Nevertheless, to give some first elements of comparison, we compare our method to the following: (1) Wiener filtering and sampling methods, which are widely used in astrophysics; and (2) the GNILC method [67], which was used on *Planck* polarization data and provides a local smoothing kernel in order to optimally remove the noise. Note that comparison (1) is performed on the simulated data used for the validation of the algorithm, while comparison (2) is performed on the observational

⁷Similar results are obtained for U .

data.

a) Wiener filtering and posterior sampling

Note that this subsection is mostly the work of Niall Jeffrey and is adapted from [200]. In terms of notations, in the following, we use bold capital letters to refer to matrices (e.g. \mathbf{S}), unbold lowercase letters to refer to two-dimensional maps (e.g. s , which is consistent with the rest of this chapter), and bold lowercase letters to refer to both deterministic and random column vectors corresponding to one-dimensional arrangements of two-dimensional variables (e.g. \mathbf{s} , column vector of dimension 512^2).

Formalism. Wiener filtering is a typical approach to signal inference for problems of the form given by Eq. (V.11). In this case, the Wiener filtered reconstruction $\tilde{\mathbf{s}}_W$ [207, 208] is given by

$$\tilde{\mathbf{s}}_W = \mathbf{W}\mathbf{d} = \mathbf{S}(\mathbf{S} + \mathbf{N})^{-1}\mathbf{d}, \quad (\text{V.13})$$

where $\mathbf{S} = E(\mathbf{s}\mathbf{s}^\dagger)$ and $\mathbf{N} = E(\mathbf{n}\mathbf{n}^\dagger)$ are the signal and noise covariance matrices, respectively.⁸ For brevity we have assumed $E(\mathbf{s}) = E(\mathbf{n}) = 0$. This is the linear filter that gives the expected minimum variance of residuals $(\mathbf{W}\mathbf{d} - \mathbf{s})$ for known \mathbf{S} and \mathbf{N} .

Furthermore, it is both the maximum a posteriori and mean posterior solution if the noise and signal are drawn from Gaussian distributions. That is, if the real-valued data are distributed with a likelihood

$$p(\mathbf{d}|\mathbf{s}, \mathbf{N}) = \frac{1}{\sqrt{(\det 2\pi\mathbf{N})}} \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{s})^\dagger \mathbf{N}^{-1}(\mathbf{d} - \mathbf{s})\right], \quad (\text{V.14})$$

and the prior distribution for the real-valued signal is

$$p(\mathbf{s}|\mathbf{S}) = \frac{1}{\sqrt{(\det 2\pi\mathbf{S})}} \exp\left[-\frac{1}{2}\mathbf{s}^\dagger \mathbf{S}^{-1} \mathbf{s}\right], \quad (\text{V.15})$$

then the Wiener posterior distribution is given by

$$\begin{aligned} p(\mathbf{s}|\mathbf{S}, \mathbf{N}, \mathbf{d}) &\propto p(\mathbf{d}|\mathbf{s}, \mathbf{N})p(\mathbf{s}|\mathbf{S}) \\ &\propto \exp\left[-\frac{1}{2}(\mathbf{s} - \mathbf{W}\mathbf{d})^\dagger (\mathbf{S}^{-1} + \mathbf{N}^{-1})(\mathbf{s} - \mathbf{W}\mathbf{d})\right]. \end{aligned} \quad (\text{V.16})$$

For the problem described in this work, the noise covariance \mathbf{N} can be estimated from the samples of noise realizations and the signal covariance \mathbf{S} can either be assumed or jointly estimated from the data [209].

Messenger field approach. We avoid inverting the dense \mathbf{W} matrix by implementing a messenger field approach, first described by [210] and now widely used in cosmology (e.g., [211, 212, 213, 214, 215]). We assume that the underlying signal is homogeneous and isotropic, such that the signal covariance \mathbf{S} is diagonal in harmonic space with elements corresponding to the one-dimensional power spectrum. The uncorrelated noise gives a diagonal noise covariance \mathbf{N} in pixel space, so that the messenger field can efficiently iterate between harmonic and pixel space.

⁸ \mathbf{s}^\dagger denotes the complex conjugate transpose of \mathbf{s} .

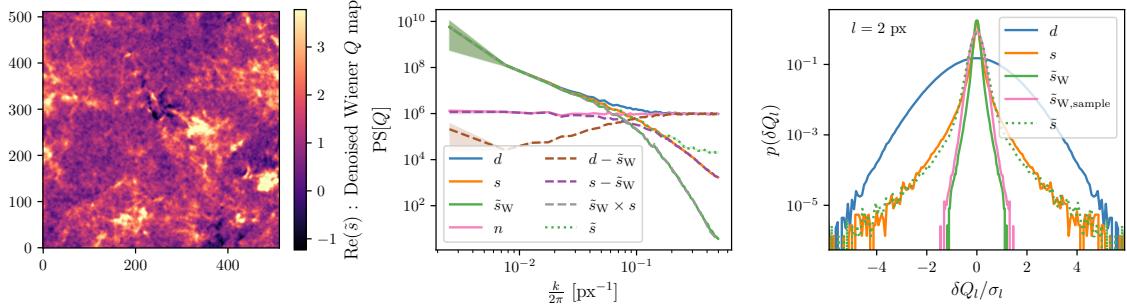


Fig. V.16 *Left:* Wiener filtered map \tilde{s}_W . *Middle:* Same as Fig. V.9, but for \tilde{s}_W . The power spectrum of \tilde{s}_W is suppressed at small scales as expected. *Right:* Same as Fig. V.10, but for $l = 2\text{px}$ with the Wiener filtered (mean) map and a Wiener posterior sample.

As we are concerned with polarization data, we use a spin-2 harmonic transformation between $\{\mathbf{s}_Q, \mathbf{s}_U\} = \{\text{Re}(\mathbf{s}), \text{Im}(\mathbf{s})\}$ and $\{\mathbf{s}_E, \mathbf{s}_B\}$, using the curl-free E -mode and divergence-free B -mode representation (see Sect. I.2.1). The signal covariance in harmonic space is therefore a concatenation of $\{\mathbf{S}_E, \mathbf{S}_B\}$. This formulation preserves the relevant Q - U correlation.

Even if the Wiener filtered reconstructed signal \tilde{s}_W is the maximum of the posterior distribution $p(\mathbf{s}|\mathbf{d})$, functions $f(\tilde{s}_W)$ do not correspond to the maximum $p(f(\mathbf{s})|\mathbf{d})$. For example, the two-point statistics (e.g., variance or power spectra) of \tilde{s}_W are not unbiased estimates of the power spectrum of \mathbf{s} . Instead, we can draw sample images \mathbf{s}_i from the posterior $p(\mathbf{s}|\mathbf{d})$ (Eq. (V.16)), so that the transformed samples $f(\mathbf{s}_i)$ are correctly distributed according to $p(f(\mathbf{s})|\mathbf{d})$. These realizations are generated by amending the messenger field algorithm [210].

Results. Figure V.16 shows the Wiener filtered reconstruction of the simulated Q map (left). The low amplitude power spectrum of the residual map $\text{Re}(s - \tilde{s}_W)$ demonstrates that the pixel values are close to the truth, whereas the power spectrum of $\text{Re}(\tilde{s}_W)$ is biased low (center), which is expected. To compare the methods, we compared with functions of samples s_i from the Wiener posterior. Though not plotted here, as we input the true power spectrum (which was not done for the WPH denoising method), the realizations s_i have the correct power spectra with relatively small sample variance.

As a goal of this denoising work is to retain the statistical non-Gaussianity in the signal, which could be represented in the WPH coefficients, we again compare the PDF of increments δQ_l . The right panel of V.16 shows the comparison between the data $\text{Re}(d)$, the signal $\text{Re}(s)$, the Wiener filtered (mean) map $\text{Re}(\tilde{s}_W)$, and a realization $\text{Re}(\tilde{s}_{W,\text{sample}})$ drawn from the Wiener posterior. We see that neither the Wiener mean map $\text{Re}(\tilde{s}_W)$ nor the sample $\text{Re}(\tilde{s}_{W,\text{sample}})$ capture the non-Gaussianity in terms of δQ_l increments as successfully as our WPH denoising method. Indeed, for $|\delta Q_l|/\sigma_l \gtrsim 1$, there is a clear divergence from the true PDF, showing that the tails of the true statistics are not recovered at all.

Furthermore, inspection of the WPH coefficients of the Wiener posterior samples $\phi(s_i)$ shows them to be noticeably further from those of the truth map than $\phi(\tilde{s})$. These

preliminary results are encouraging, but not surprising. The Gaussian prior distribution in the statistical model of the Wiener posterior leads to a poor recovery of the non-Gaussian structures that are intrinsic to the polarized dust emission.

b) GNILC

The GNILC method is a wavelet-based component separation method designed to extract the emission of the Galactic foregrounds from the *Planck* full-sky maps (see Sect. I.2.3). It has been applied to polarization maps Q and U [29, 58] to disentangle the thermal dust polarized emission from the CMB polarization and instrumental noise over the entire sky (see Sect. I.2.3 for additional details). We show in Fig. V.13 (right column) the resulting Q and U maps for the Chamaeleon-Musca field observed at 353 GHz. We note that at this frequency, the CMB can be safely ignored, so that the main effect of the GNILC algorithm is to denoise the dust emission. In these maps, we clearly see that the smallest scales, most contaminated by the noise, have been filtered out. Compared to GNILC, our denoised maps thus include a wider variety of structures at intermediate and small scales.

A quantitative comparison based on the power spectrum and the PDFs of the increments for the Q maps is shown in Figs. V.14 and V.15. The power spectrum of the GNILC map plummets at small scales and exhibits a lack of power compared to that of \tilde{s} and the cross-spectrum $d_1 \times d_2$. The PDFs of the increments also show important discrepancies, especially in the tails, between the \tilde{s} map and the GNILC map. Assuming that the statistics of \tilde{s} give a reasonable order of magnitude of the true statistics, based on the validation on simulated data, this suggests a distortion of these statistics by the GNILC method. This semi-quantitative comparison demonstrates the superiority of our method for recovering the true power spectrum and, *a priori*, non-Gaussian properties.

V.2.4 Discussion

We have introduced a new method for the denoising of *Planck* polarization data with an algorithm inspired by the generation of random synthetic maps from WPH statistics that characterize the spatial structure of dust emission. This method takes advantage of the strong statistical differences between the signal of interest (non-Gaussian and regular) and the noise (close to Gaussian and irregular) by performing an optimization that constrains the statistical properties of the denoised map plus noise realizations.

We applied our method to mock $Q + iU$ noisy data designed to emulate typical *Planck* polarization maps of dust in the diffuse ISM. The denoised map has a power spectrum that coincides with the true power spectrum down to a minimum scale where the power of the noise is ten times that of the signal, while being highly correlated with the true signal. It recovers the PDFs of the increments for various isotropic lags, demonstrating that our method is able to retrieve non-Gaussianity in the data. A complementary WPH analysis demonstrates that the denoising mitigates the impact of the noise on these statistics, although significant discrepancies remain. Finally, we applied this method to a 353 GHz *Planck* observation of the Chamaeleon-Musca field. To give some first elements of comparison, we have compared our results to those obtained with a Wiener approach (for the mock data), and those of the GNILC component separation method (for the observational data). For the Wiener approach, we find that neither the Wiener filtered image nor realizations drawn from the Wiener posterior distribution are able to retrieve

the PDFs of the increments, even when the true power spectrum was given as input. The comparison with GNILC shows that our method better recovers the true power spectrum and the PDFs of the increments.

Our method has been introduced as a statistical denoising method, but we have shown that it can also be used as a regular denoising algorithm. Moreover, this should be more generally applicable to component separation problems. In particular, we expect this method to be efficient at disentangling dust emission from the CMB, cosmic infrared background, and noise in *Planck* maps, provided that we have accurate models for each of these contaminants at our disposal. Therefore, it could hopefully enhance the scientific outcome of *Planck* data and other CMB experiments. This idea will be developed in the next section.

One of the main motivations of our work is to build a generative model of the dust polarization signal that takes into account the non-Gaussianity of the data. Such a model derived from the analysis of *Planck* data may be used to simulate the dust polarization sky. Our denoising algorithm constitutes a step toward this modeling, in the sense that WPH statistics of the dust emission, corrected to a first approximation from noise contamination, may be used to define a generative model as explained in the previous chapter. However we have seen that an important part of the WPH coefficients of the denoised map remains significantly affected by the noise. Further work is needed to quantify the impact of the noise on this generative model.

V.3 Toward enhanced component separation methods

A denoising problem can be viewed as a two-component separation problem, so that the previous algorithm acts in this sense as a component separation method. However our separation is designed to be statistical, meaning that contrary to *Planck* component separation methods that were introduced in Sect. I.2.3 we are not primarily interested in the deterministic maps of the components, but rather in their statistical properties. In this section, we present a first attempt to generalize this algorithm to the more general problem of a n -component statistical separation. We assume that we have at our disposal statistical models for $n - 1$ components, and that we are interested in retrieving the non-Gaussian statistical properties of the last one.

The extension of the algorithm is very natural, and it is discussed first. Then, we provide a first application to a $Q + iU$ polarization map at 353 GHz in the BICEP2 field as observed by the *Planck* satellite. The observed data is modeled as a sum of three components: a dust component, a CMB component, and a noise component. Based on models of the CMB and noise components, we focus on the retrieval of the statistical properties of the dust component. We conclude this section by discussing a few perspectives of this method.

V.3.1 Extension of the statistical denoising algorithm

In a n -component separation problem, the observed signal d is modeled as a sum of components s_i :

$$d = s_1 + \cdots + s_n, \quad (\text{V.17})$$

with potentially one of these components corresponding to a noise signal. These components are modeled as realizations of some random fields S_1, \dots, S_n . We assume that we know $n - 1$ of them, let us say S_2, \dots, S_n , meaning that we are able to generate as many independent realizations of these random fields as needed. The goal is to estimate the statistical properties of S_1 , which are unknown.

We generalize our previous statistical algorithm introduced in Sect. V.2.1 as follows. We consider M_2, \dots, M_n realizations of S_2, \dots, S_n , respectively. For each random field S_i , we call these realizations $\{s_{i,1}, \dots, s_{i,M_i}\}$. Given a statistical operator ϕ , the generalized algorithm consists in minimizing the following loss function:

$$\mathcal{L}(u) = \frac{1}{M_2 \dots M_n} \sum_{j_2=1}^{M_2} \dots \sum_{j_n=1}^{M_n} \|\phi(u + s_{2,j_2} + \dots + s_{n,j_n}) - \phi(d)\|^2. \quad (\text{V.18})$$

Here again we use a L-BFGS optimizer initialized with $u_0 = d$. In practice in the following, ϕ will also correspond to WPH statistics, and we use the same procedure as that introduced in Sect. V.2.1.

Note that, although this algorithm is presented as a n -component separation problem, it remains formally completely equivalent to a two-component separation problem. Indeed, defining $t = s_2 + \dots + s_n$, and $T = S_2 + \dots S_n$ we have simply retrieved an equivalent formulation of the method described in Sect. V.2.1. However, this formulation remains a useful starting point for defining more complex losses involving for example additional terms of the form $\|\phi(u + s_{2,j_2} + \dots + s_{n-1,j_{n-1}}) - \phi(d - s_{n,j_n})\|^2$, which could improve the recovery of the target statistics.

V.3.2 Application to the BICEP2 field

This section presents very preliminary results, which were obtained in collaboration with Jonathan Aumont and Léo Vacher.

a) Presentation of the data

We apply our generalized method to a $Q + iU$ polarization map of the BICEP2 field observed at 353 GHz with the *Planck* satellite (PR3 data⁹ [55]). Contrary to the Chamaeleon-Musca field, in this field and at 353 GHz, dust emission is no longer largely dominant at all scales compared to the CMB, and we expect comparable power between the two at small scales. We consider the Q and U maps corresponding to the full mission, and those corresponding to the half-missions. We also make use of the 300 end-to-end simulated Q and U maps of the noise and the systematics of the instrument for the full mission, as well as of 300 simulated CMB maps FFP10 (Full Focal Plane simulations, v. 10). We project these maps on 256×256 grids with a pixel size of $8'$, centered on the region observed by the BICEP2 experiment, that is at Galactic coordinates $(l, b) = (316^\circ, -57^\circ)$.

⁹https://wiki.cosmos.esa.int/planck-legacy-archive/index.php/Main_Page

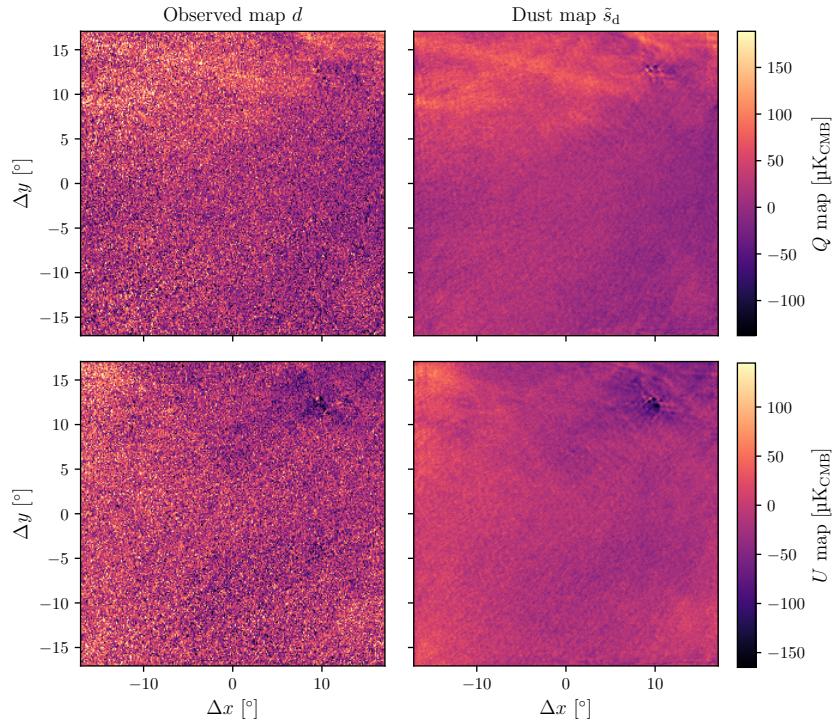


Fig. V.17 Observed (left column) and estimated dust contribution (right column) Q (top row) and U (bottom row) maps of the BICEP2 region as observed by the Planck satellite at 353 GHz.

b) Results

We apply a similar procedure to that of Sect. V.2 although here the loss takes the following form:

$$\mathcal{L}(u) = \frac{1}{M_2 M_3} \sum_{i=1}^{M_2} \sum_{j=1}^{M_3} \|\phi(u + s_{\text{CMB},i} + n_j) - \phi(d)\|^2, \quad (\text{V.19})$$

where $M_2 = M_3 = 300$, $s_{\text{CMB},1}, \dots, s_{\text{CMB},M_2}$ are the 300 CMB realizations, and n_1, \dots, n_{M_3} are the 300 noise realizations. To keep the computations reasonable, we limit the double sum to a diagonal summation, that is we only sum over $i = j$ values.

We show in Fig. V.17 the resulting dust map \tilde{s}_d (right column) next to the observational total map d (left column). The overall noise level has been clearly mitigated, and we now better discern the underlying dust signal. As a first quantitative assessment, we show in Fig. V.18 a power spectrum analysis that is similar to that of Sect. V.2.2. We plot the power spectra of the Q maps corresponding to the total map d and dust map \tilde{s}_d , as well as power spectra corresponding to one specific noise map n and one CMB map s_{CMB} . We also show the power spectrum of $d - \tilde{s}_d$, which should approximately coincide with the sum of the power spectra of n and s_{CMB} . However note that the power spectrum of the noise remains largely dominant over that of the CMB, so that this sum is dominated by the power of the noise. Finally, we show the cross-spectrum between the two half-mission maps d_1 and d_2 , which estimates the power spectrum of the sum of the dust and CMB signals. The power spectrum of \tilde{s}_d appears to be satisfactorily consistent

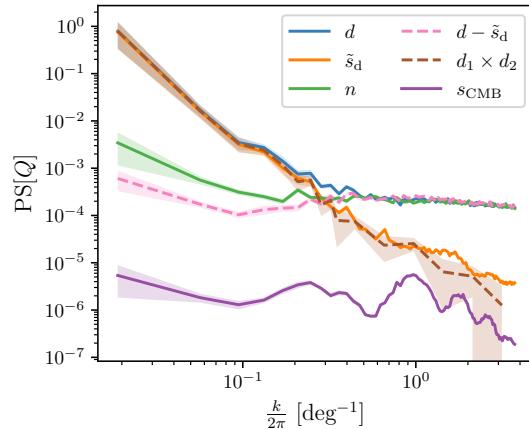


Fig. V.18 Q maps power spectra for d , \tilde{s}_d , n , s_{CMB} (all in solid lines), and $d - \tilde{s}$, and cross-spectrum between d_1 and d_2 (both in dashed lines). This cross-spectrum gives an unbiased estimate of the power spectrum of the sum of the dust and CMB emissions.

with this cross-spectrum up to the degree scale as expected. Indeed, over this range of scales, the contribution of the CMB is negligible in the empirical cross-spectrum, so that this cross-spectrum is a good estimate of the power spectrum of the dust emission. This agreement worsens for scales smaller than the degree scale, showing that an excess of power remains in the estimated dust map \tilde{s}_d . On the other hand, the power spectrum of $d - \tilde{s}_d$ agrees with the sum of the power spectra of n and s_{CMB} with a good approximation for most scales, as expected. However, at the largest scales, significant discrepancies show that, similarly to the results of Sect. V.2, the statistical component separation procedure has a weak impact on the largest scales, for which the dust signal is largely dominant over the noise and CMB components.

V.3.3 Perspectives

These results, while very preliminary, illustrate one of the motivations of this thesis, that is the improvement of component separation methods for the detection of primordial B -modes. In this context, one has to retrieve the statistical properties of the CMB emission with a sufficiently high accuracy to hopefully detect discrepancies with the CMB model involved in this example, i.e. a fiducial model with null B -mode signal. We have been working at 353 GHz here for convenience, as the SNR remains sufficiently high for our method to give interesting results. However, at this frequency the signal remains dominated by the dust emission. For the detection of this B -mode signal, main target frequencies are lower, typically around 100 GHz (see e.g. [2]), so that dust and CMB are much more comparable in terms of power (see Chapter I). However, at 100 GHz, the SNR of *Planck* data in the BICEP2 region is too low to perform a similar study.

Let us recall that this approach ignores an important aspect of polarization data, namely its spectral dimension. Chapter II has underlined the richness of dust spectral dimension. Therefore, ignoring it by tackling this statistical component separation prob-

lem using a single frequency channel is an unnecessary complication. Dust polarization data is correlated across the frequency channels, and we need to take advantage of this. WPH moments characterize the non-Gaussian statistical properties of a single random field X , however, since they rely on covariances, it is very natural to define analogous cross-moments to characterize the joint non-Gaussian statistical properties of two random fields X and Y . For X and Y two statistically homogeneous random fields, these cross-WPH moments of X and Y may be defined as follows:

$$C_{\xi_i, p_i, \xi_j, p_j}[X, Y](\boldsymbol{\tau}) = \text{Cov} \left([X \star \psi_{\xi_i}(\mathbf{r})]^{p_i}, [Y \star \psi_{\xi_j}(\mathbf{r} + \boldsymbol{\tau})]^{p_j} \right). \quad (\text{V.20})$$

The definition of a consistent subset of cross-WPH moments / standard WPH moments to characterize the joint statistical properties of two random fields is needed. Once this is done, the next step would be to perform joint statistical component separations of multi-frequency data. This will be the subject of future work.

Another comparable problem, although it concerns total intensity data, is that of the detection of the non-Gaussian statistical properties of the cosmic infrared background (CIB). Indeed, the CIB emission remains entangled with the dust emission, and disentangling these two components while taking into account the non-Gaussian properties of the data remains an open problem. Numerical experiments have already been conducted in that direction with a similar algorithm to the one used here [216]. This shows the interest of such statistical component separation algorithms beyond the CMB context.

Conclusion and perspectives

I have introduced new statistics for the description and modeling of polarized thermal emission maps from interstellar dust, namely the (R)WST and the WPH. These provide low-variance multiscale representations of data with non-Gaussian statistics, with a particular focus on the characterization of the interactions between scales. These representations shed new light on the statistical properties of the magnetized ISM, and may be related to the physics of the ISM, although this remains to be explored fully. In particular, the RWST, which is built from a fit of the angular dependence of the WST coefficients, yields interpretable coefficients quantifying the multiscale properties of polarization maps in terms of isotropic and anisotropic contributions. Moreover, the (R)WST and WPH statistics allow us to build generative statistical models from a very limited set of maps. Polarization maps drawn from such models are visually and quantitatively consistent with the input data, in statistical terms. This demonstrates the relevance of these statistics to characterize dust polarization maps. Finally, I have exhibited the detrimental impact of noise on the WST and WPH statistics, thus highlighting the complications inherent in the modeling of observational data. To address this problem, I have designed a statistical denoising algorithm to recover the statistical properties of the noise-free emission. Two publicly available Python packages complement this work, namely PyWST and PyWPH, which allow respectively for WST and WPH statistical analyses of two-dimensional data.

The perspectives of this work are numerous. First, it should be noted that an important part of this work focused on methodological issues. This methodology should now be applied more widely to observational data (e.g. Advanced ACT data [217]). However, a current limitation is that (R)WST/WPH analyses can only be conducted within the flat-sky approximation, i.e., to a small enough part of the sky. Indeed, for larger areas, and up to full-sky analyses, the extension of the WST and WPH on the sphere remains to be done. A useful starting point could be the *needlet* formalism that has already been introduced in the literature for wavelet analysis on the sphere [68, 69].

Second, throughout this work, an important aspect of dust polarization data has been ignored, namely its spectral dimension. Dust polarization data is correlated across frequency channels, and we need to take advantage of this for the improvement of component separation methods. The WPH statistics are naturally extendable to the characterization of interactions between the scales of two different maps since they rely on covariances. The definition of relevant cross-WPH moments mentioned in Chapter V has already been initiated. An immediate application of these moments would be to improve the statistical denoising algorithm introduced in Chapter V by leveraging this multi-frequency information.

We have seen that this statistical denoising algorithm may be extended to more general

statistical component separation problems. The extension to data involving a combination of dust, CMB, and noise signals was demonstrated as a proof-of-concept using *Planck* data. Yet, this would deserve further analysis, notably involving a validation on simulated data. In another on-going project I am also involved in, the WPH is used to separate dust and CIB anisotropies in *Herschel* far-infrared observations [216] in a similar way as the denoising problem addressed in Chapter V.

Improving component separation methods for the detection of primordial B -modes can take an alternative way to what has been outlined in Chapter V. I have also been involved in a project that is introduced in [201], where WPH-based generative models of simulated maps of the polarized emission of dust play a central role for CMB B -mode inference. Within a likelihood-free Bayesian inference framework, the authors manage to estimate the pixel-level posterior probability of a simulated B -mode map based on a single frequency simulated observation for a BICEP-like sky patch. This further legitimates the interest of such generative models of polarized dust emission.

Finally, let us remark that in this thesis, I have focused on the analysis of polarization data through Stokes parameters I , Q and U , mostly applying the (R)WST and WPH to $Q + iU$ or $\tilde{Q} + i\tilde{U}$ maps. However, for the quest of primordial B -modes, the natural variables are E and B , and a similar statistical analysis involving these variables would be relevant. In particular, it would be interesting to study how well generative models based on the (R)WST/WPH statistics would capture observational properties of the dust foreground, such as the E/B asymmetry, and the TE and TB correlations.

Appendix A

Random fields: definitions and elementary properties

Random processes, also known as stochastic processes, are mathematical objects of probability theory that are commonly used to model systems varying randomly in a tremendous variety of domains (e.g. physics, signal processing, biology, finance). In cosmology, typical applications include the modeling of the CMB [33], or the description of the fluctuations of density of the large-scale structure of the Universe (see e.g. [218]). In astrophysics of the ISM, in addition to being crucial for the spatial modeling of the dust emission (see Chapter II for a general discussion), they are for example also used for the modeling of the temporal fluctuations of temperature of small dust grains (see e.g. [219]).

In this appendix, we give elementary definitions and notions revolving around these mathematical objects that are relevant to this thesis. In particular, we formally introduce the power spectrum, which is extensively discussed in this thesis, and Gaussian random fields (GRFs), which are the most common random fields used for the statistical modeling of the dust emission. This appendix does not aim to be either an exhaustive introduction to the subject, or perfectly rigorous from a mathematical point of view. For a more complete introduction on the subject, we refer the interested reader to introductory textbooks such as [220] on which part of this appendix is based.

A.1 Basic concepts

A.1.1 Random processes and random fields: definitions

Let us consider a probability space (Ω, \mathcal{A}, P) where Ω is a sample space, \mathcal{A} is a σ -algebra, and P a probability measure.

Definition A.1.1. A *random process* is a collection of random variables indexed by a set T called the index set, or parameter set. These random variables all take their values in the same measurable space E .

Practically, a random process X is a function $X : T \times \Omega \rightarrow E$, meaning that for each outcome ω of the sample space Ω , X assigns the function $X(\cdot, \omega) : T \rightarrow E$, which is called a *realization* of X . Thus, whereas a mere random variable assigns an element of E to a

given outcome, a random process assigns a function. Obviously, for all $t \in T$, $X(t, \cdot)$ is a random variable.

The index set usually corresponds to a parametrization of time, with typically $T = \mathbb{R}$ or $T = \mathbb{N}$ for a continuous or discrete time, respectively. But it also commonly parametrizes space, which can be the Euclidean space \mathbb{R}^n or more convoluted manifolds. Space can also be assumed discrete or continuous. When T represents space, we rather call X a *random field*.

In this thesis we only deal with random fields. The index set T may be the sphere \mathbb{S}^2 when modeling observations on the whole sky (e.g. sky map of the CMB or of the dust emission), or a rectangular domain $[a, b] \times [c, d]$ when the flat-sky approximation is valid. When discrete measurements require to assume a discrete space, typically for an image of size $M \times N$, we may have $T = \llbracket 0, M - 1 \rrbracket \times \llbracket 0, N - 1 \rrbracket$, or $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ if periodic boundary conditions are assumed. The random fields we will consider will be either real or complex-valued, so that we have $E = \mathbb{R}$ or $E = \mathbb{C}$, respectively.

In the following, we will discuss real or complex-valued random fields only. To simplify notations we will not show the dependence on the outcome, for example denoting $X : \mathbb{R}^2 \rightarrow \mathbb{R}$ a real-valued random field indexed by \mathbb{R}^2 . Also, we will refer to random fields with capital letters (e.g. X), and to their realizations with lowercase letters (e.g. x).

A.1.2 Elementary analysis tools

The distribution of a random field is determined by the ensemble of its finite-dimensional distributions. These are defined as follows.

Definition A.1.2. The *finite-dimensional distributions* of a random field X are the joint distributions of $X(\mathbf{r}_1), \dots, X(\mathbf{r}_n)$ with $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$ and $n \in \mathbb{N}$.

The finite-dimensional distributions are usually described in terms of joint cumulative distribution functions (CDFs), or in terms of joint probability density functions (PDFs) when they exist. We consider in the following random fields for which these joint PDFs do exist. For a random field X , a given n , and $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$, the joint PDF of the random vector $(X(\mathbf{r}_1), \dots, X(\mathbf{r}_n))$, also referred to as a n -th order PDF of X , is denoted $f_X(x_1, \dots, x_n; \mathbf{r}_1, \dots, \mathbf{r}_n)$, with $x_1, \dots, x_n \in E$.

In general deriving n -th order PDFs, either analytically or numerically, is often intractable, especially when n increases, so that one typically resorts to the computation of simpler quantities to characterize the properties of a random field. We consider in the following a random field X for which all the following quantities are well defined. Here again, we refer to textbooks for details on the validity of the definitions.

The most simple and natural quantities to derive from a given random field X are the mean function, and the autocorrelation (or alternately autocovariance) function. These are defined as follows.

Definition A.1.3. The *mean function* of X , denoted $\mu_X(\mathbf{r})$, is for each \mathbf{r} the expected value of the random variable $X(\mathbf{r})$, that is:

$$\mu_X(\mathbf{r}) = \mathbb{E}[X(\mathbf{r})] = \int_E x f_X(x; \mathbf{r}) dx. \quad (\text{A.1})$$

Definition A.1.4. The *autocorrelation function* of X , denoted $R_X(\mathbf{r}_1, \mathbf{r}_2)$, is the expected value of the products of random variables $X(\mathbf{r}_1)X(\mathbf{r}_2)$, that is:

$$R_X(\mathbf{r}_1, \mathbf{r}_2) = E \left[X(\mathbf{r}_1) \overline{X(\mathbf{r}_2)} \right] = \int_E x_1 \overline{x_2} f_X(x_1, x_2; \mathbf{r}_1, \mathbf{r}_2) dx_1 dx_2, \quad (\text{A.2})$$

where \bar{x} is the complex conjugate of x .

Definition A.1.5. The *autocovariance function* of X , denoted $C_X(\mathbf{r}_1, \mathbf{r}_2)$, is the covariance¹ between the random variables $X(\mathbf{r}_1)$ and $X(\mathbf{r}_2)$, that is:

$$\begin{aligned} C_X(\mathbf{r}_1, \mathbf{r}_2) &= \text{Cov}[X(\mathbf{r}_1), X(\mathbf{r}_2)], \\ &= E[(X(\mathbf{r}_1) - E[X(\mathbf{r}_1)]) (\overline{X(\mathbf{r}_2)} - \overline{E[X(\mathbf{r}_2)]})]. \end{aligned} \quad (\text{A.3})$$

The autocovariance function is obviously simply related to the autocorrelation and mean functions as:

$$C_X(\mathbf{r}_1, \mathbf{r}_2) = R_X(\mathbf{r}_1, \mathbf{r}_2) - \mu_X(\mathbf{r}_1)\overline{\mu_X(\mathbf{r}_2)}. \quad (\text{A.4})$$

The mean function and the autocorrelation/autocovariance function of a random field are often referred to as first and second order statistics, respectively. These are perhaps the most immediate quantities to look at when studying a random field.

For a complex-valued random field, some complementary second order information might be measured by the pseudo-autocovariance function.²

Definition A.1.6. The pseudo-autocovariance function of X , denoted $C'_X(\mathbf{r}_1, \mathbf{r}_2)$, is the covariance between the random variables $X(\mathbf{r}_1)$ and $\overline{X(\mathbf{r}_2)}$, that is:

$$\begin{aligned} C'_X(\mathbf{r}_1, \mathbf{r}_2) &= \text{Cov}[X(\mathbf{r}_1), \overline{X(\mathbf{r}_2)}], \\ &= E[(X(\mathbf{r}_1) - E[X(\mathbf{r}_1)]) (X(\mathbf{r}_2) - E[X(\mathbf{r}_2)])]. \end{aligned} \quad (\text{A.5})$$

Obviously in the case of a real-valued random field, the pseudo-autocovariance and the autocovariance functions are identical.

We also define for two random fields X and Y , the cross-correlation function and the cross-covariance function as follows.

Definition A.1.7. The cross-correlation function R_{XY} , and the cross-covariance function C_{XY} of a pair of random fields are defined as:

$$R_{XY}(\mathbf{r}_1, \mathbf{r}_2) = E \left[X(\mathbf{r}_1) \overline{Y(\mathbf{r}_2)} \right], \quad (\text{A.6})$$

$$C_{XY}(\mathbf{r}_1, \mathbf{r}_2) = \text{Cov}[X(\mathbf{r}_1), Y(\mathbf{r}_2)]. \quad (\text{A.7})$$

And we have similarly:

$$C_{XY}(\mathbf{r}_1, \mathbf{r}_2) = R_{XY}(\mathbf{r}_1, \mathbf{r}_2) - \mu_X(\mathbf{r}_1)\overline{\mu_Y(\mathbf{r}_2)}. \quad (\text{A.8})$$

Finally, we can define simple higher order statistics. Common ones are n -point correlation functions.

Definition A.1.8. For a random field X and $n \in \mathbb{N}^*$, the n -point correlation function is the function $C_{n,X} : T^n \rightarrow E$ defined for all $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$ as:

$$C_{n,X}(\mathbf{r}_1, \dots, \mathbf{r}_n) = E[X(\mathbf{r}_1) \dots X(\mathbf{r}_n)]. \quad (\text{A.9})$$

¹We recall that the covariance of two random variables X and Y , called $\text{Cov}[X, Y]$, is defined by $\text{Cov}[X, Y] = E[(X - E(X))(Y - E(Y))]$.

²Note that we could also define on the same model a pseudo-autocorrelation function.

A.1.3 Common properties

We will often deal with random fields whose distribution are invariant under the action of symmetries of T , typically translations or rotations. We review some related definitions and their immediate implications.

A translation invariant random field is called a stationary random field. This vocabulary, while usually referring to temporal shift invariance when T describes time, will be employed likewise to refer to spatial shift invariance when T describes space. In this work, a stationary random field will also be called equivalently a *statistically homogeneous* random field.

Definition A.1.9. A random field $X = (X(\mathbf{r}))_{\mathbf{r} \in T}$ is (strict sense) *stationary* if and only if for all $\boldsymbol{\tau} \in T$, $(X(\mathbf{r} + \boldsymbol{\tau}))_{\mathbf{r} \in T}$ has the same distribution as X .

This is equivalent to the fact that all the finite-dimensional distributions of X are invariant to global translations, meaning in practice that for any given n -th order PDF, for all $\boldsymbol{\tau} \in T$, we have $f_X(x_1, \dots, x_n; \mathbf{r}_1 + \boldsymbol{\tau}, \dots, \mathbf{r}_n + \boldsymbol{\tau}) = f_X(x_1, \dots, x_n; \mathbf{r}_1, \dots, \mathbf{r}_n)$.

Showing that a random field is strict sense stationary can be inextricable without knowing all the finite-dimensional distributions of X . A looser and more practical form of stationarity, called wide sense stationarity, is thus usually employed.

Definition A.1.10. A random field X is *wide sense stationary* (WSS) if and only if its mean, autocovariance and pseudo-autocovariance functions (or equivalently its mean, autocorrelation and pseudo-autocorrelation functions) are invariant to global translations, that is if and only if:

$$\mu_X(\mathbf{r} + \boldsymbol{\tau}) = \mu_X(\mathbf{r}) \quad \text{for all } \boldsymbol{\tau}, \mathbf{r} \in T, \quad (\text{A.10})$$

$$C_X(\mathbf{r}_1 + \boldsymbol{\tau}, \mathbf{r}_2 + \boldsymbol{\tau}) = C_X(\mathbf{r}_1, \mathbf{r}_2) \quad \text{for all } \boldsymbol{\tau}, \mathbf{r}_1, \mathbf{r}_2 \in T, \quad (\text{A.11})$$

$$C'_X(\mathbf{r}_1 + \boldsymbol{\tau}, \mathbf{r}_2 + \boldsymbol{\tau}) = C'_X(\mathbf{r}_1, \mathbf{r}_2) \quad \text{for all } \boldsymbol{\tau}, \mathbf{r}_1, \mathbf{r}_2 \in T. \quad (\text{A.12})$$

Obviously strict sense stationarity implies WSS, but the converse is not true. We will say that a random field is non-stationary when it is not WSS.

For a WSS random field X , the mean function is a constant function, and the (pseudo-)autocovariance effectively only depends on one variable, so that we simplify the notations by defining:

$$\mu_X = \mu_X(\mathbf{r}) \quad \text{for all } \mathbf{r} \in T, \quad (\text{A.13})$$

$$C_X(\boldsymbol{\tau}) = C_X(\mathbf{r}, \mathbf{r} + \boldsymbol{\tau}) \quad \text{for all } \boldsymbol{\tau} \in T, \quad (\text{A.14})$$

$$C'_X(\boldsymbol{\tau}) = C'_X(\mathbf{r}, \mathbf{r} + \boldsymbol{\tau}) \quad \text{for all } \boldsymbol{\tau} \in T. \quad (\text{A.15})$$

The autocorrelation function R_X of a WSS random field X also verifies two useful basic properties.

Proposition A.1.1. For a WSS random field X we have for all $\boldsymbol{\tau} \in T$:

1. $R_X(\boldsymbol{\tau}) = \overline{R_X(-\boldsymbol{\tau})}$, and an identical relation for C_X ,
2. $|R_X(\boldsymbol{\tau})| \leq R_X(\mathbf{0})$, meaning that the maximum of the modulus of the autocorrelation function is at the origin.

Another important class of invariant random fields are isotropic random fields, that are invariant to rotations.

Definition A.1.11. A random field $X = (X(\mathbf{r}))_{\mathbf{r} \in T}$ is *isotropic* if and only if for any rotation g of T , $(X(g\mathbf{r}))_{\mathbf{r} \in T}$ has the same distribution as X .

We also define self-similar random fields, whose distributions are simply transformed under scaling symmetries.

Definition A.1.12. A random field $X = (X(\mathbf{r}))_{\mathbf{r} \in T}$ is *self-similar* of order $H > 0$ if and only if for all $\lambda > 0$, $(X(\lambda\mathbf{r}))_{\mathbf{r} \in T}$ has the same distribution as $\lambda^H X = (\lambda^H X(\mathbf{r}))_{\mathbf{r} \in T}$.

Note that we could also define weaker forms of these definitions by analogy with the definition of wide sense stationarity.

Finally, we end this section by mentioning the notion of ergodicity. When observing realizations of a WSS random field, we may wonder how to estimate its mean or its autocorrelation function, or even other types of statistics. Obviously, if one observes many independent realizations of the same random field, it will be straightforward to estimate these statistics by computing averages over the realizations, called ensemble averages. However, if one only observes a single realization of the random field, the problem might seem intractable. An intuitive way to estimate, for example, the mean of a WSS random field X given a single realization x is to compute the spatial average $\langle x \rangle_A = \frac{1}{|A|} \int_A x(\mathbf{r}) d\mathbf{r}$, where A is some bounded domain of T and $|A|$ its volume. This is motivated by the intuition that when the volume of A goes to infinity, the spatial average should converge to the ensemble average. Yet, there is a priori no guarantee for that. A random field that satisfies this property is said to be ergodic (in the mean, on this example).

Definition A.1.13. A WSS random field X is said to be *ergodic* if ensemble averages involving the process can be estimated through spatial averages of any realization x of the process. Typically, one can restrict the definition of ergodicity to weaker forms, such as ergodicity in the mean, meaning $\langle x(\mathbf{r}) \rangle = E[X(\mathbf{r})]$, or ergodicity in the autocorrelation function, meaning $\langle x(\mathbf{r})x(\mathbf{r} + \boldsymbol{\tau}) \rangle = E[X(\mathbf{r})X(\mathbf{r} + \boldsymbol{\tau})]$.

A.2 Spectral analysis

A.2.1 Power spectrum: definition and properties

In this thesis, we extensively discuss the power spectrum statistics, which is introduced as a standard (although not comprehensive) statistics to describe the properties of data. For a WSS random field, it is simply defined as the Fourier transform of the autocorrelation function, thus describing the same statistical content as the autocorrelation function but from a spectral point of view.

Definition A.2.1. For a WSS random field X , the *power spectrum*, or *power spectral density*, of X , denoted S_X , is defined as the Fourier transform of the autocorrelation function $R_X : T \rightarrow E$, that is:

$$S_X = \mathcal{F}[R_X], \quad (\text{A.16})$$

where \mathcal{F} is the (potentially multi-dimensional) Fourier transform on T .³

Depending on the nature of T , \mathcal{F} may refer to the usual (continuous) Fourier transform, or may be the discrete Fourier transform (DFT), or other. For example, for a random field X defined on $T = \mathbb{R}^2$, \mathcal{F} is the two-dimensional Fourier transform and we have for all $\mathbf{k} \in \mathbb{R}^2$:

$$S_X(\mathbf{k}) = \int_{\mathbb{R}^2} R_X(\boldsymbol{\tau}) e^{-i\mathbf{k}\cdot\boldsymbol{\tau}} d\mathbf{k}. \quad (\text{A.17})$$

On the other hand, if T describes the pixels of a random image X of size $M \times N$ with periodic boundary conditions, that is with $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, \mathcal{F} is the DFT (see Appendix B for more details) and the power spectrum of X is, for all $\boldsymbol{\xi} \in \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$:

$$S_X(\boldsymbol{\xi}) = \sum_{\boldsymbol{\tau} \in T} R_X(\boldsymbol{\tau}) e^{-2i\pi(k_x \tau_x/M + k_y \tau_y/N)}. \quad (\text{A.18})$$

Let us also remark that the definition of the power spectrum necessarily depends on the choice of convention chosen to define the Fourier transform.

We also define the cross-spectrum as the Fourier transform of the cross-correlation function.

Definition A.2.2. For a pair of random fields X and Y , the *cross-spectrum*, or *cross spectral density*, between X and Y , denoted S_{XY} , is defined as the Fourier transform of the cross-correlation function, that is:

$$S_{XY} = \mathcal{F}[R_{XY}]. \quad (\text{A.19})$$

We now restrain the following discussion to the case of a stationary random field X defined on $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, which typically models the statistical properties of an image of size $M \times N$ assuming periodic boundary conditions. This kind of model is standard for image analysis when stationarity can be assumed, and it will often be used or implicitly assumed in this thesis. In this case, the power spectrum of X is defined as in Eq. (A.18). From this definition, one can derive the useful following proposition (see e.g. [221] for a proof).

Proposition A.2.1. For a WSS random field defined on $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, we have for all $\boldsymbol{\xi} \in T$:

$$S_X(\boldsymbol{\xi}) = \frac{1}{MN} \mathbb{E} \left[|\hat{X}(\boldsymbol{\xi})|^2 \right]. \quad (\text{A.20})$$

This connects the power spectrum to a second-order moment derived from the DFT of X , and in practice we will estimate the power spectrum of a random image X by means of this relation.

Fourier modes derived from the DFT are conveniently indexed by $\boldsymbol{\xi} \in T$, however one can define a more interpretable variable corresponding to a physical wave vector. With Δ the spatial step, which has obviously the dimension of a distance, we define the

³A classical alternative definition is to define the power spectrum of a random process X , here assumed to be defined on $T = \mathbb{R}$ to simplify, as $S_X(f) = \lim_{t_0 \rightarrow +\infty} \mathbb{E} \left[|\hat{X}_{t_0}(f)|^2 \right] / 2t_0$, where \hat{X}_{t_0} is the Fourier transform of a truncated version of X that is null when $|t| \geq t_0$. Thanks to the Wiener-Khintchine-Einstein theorem this second definition does coincide with the first one [220].

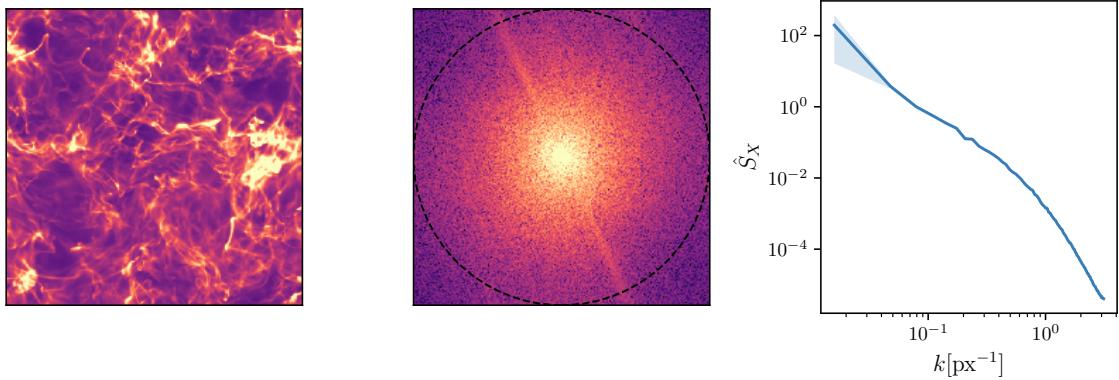


Fig. A.1 *Illustration of the power spectrum estimation algorithm, showing an example image (left, this comes from the I_{\parallel} data set introduced in Chapter III), the map of the squared modulus of its DFT centered on its null frequency and with a logarithmic color scale (middle), and the power spectrum deduced from the middle map as described in Sect. A.2.2. The circle drawn in dashed line (middle) delimits the Nyquist disc.*

wave vector \mathbf{k} corresponding to $\xi \in T$ as $\mathbf{k} = 2\pi\Delta^{-1}(\tilde{\xi}_x/M, \tilde{\xi}_y/N)$, with $\tilde{\xi}_x = \xi_x$ for $0 \leq \xi_x \leq M/2$ and $\tilde{\xi}_x = \xi_x - M$ for $M/2 < \xi_x \leq M - 1$, and a similar definition for $\tilde{\xi}_y$. This way, the wave vector \mathbf{k} is such that $k_x, k_y \in [-\pi/\Delta, \pi/\Delta]$. We call $\pm\pi/\Delta$ the Nyquist wavenumber, corresponding to the highest spatial frequency one can probe along one dimension of some discrete data. Note that for rectangular images, we can have wave vectors \mathbf{k} such that $|\mathbf{k}| > \pi/\Delta$. We will say that these modes are "out of the Nyquist disc".

A.2.2 Power spectrum: estimation

Power spectrum estimation is a vast topic, which is extensively covered in the literature. We refer for instance to textbooks [220] and [222] for a broader outlook on this subject. Here we introduce a very simple method to estimate a power spectrum from an image with periodic boundary conditions. We make use of this method to estimate all the power spectra shown in this manuscript.

Let us consider an image x of size $M \times N$. We assume that this image can be rightfully modeled as a realization of a random field X defined on $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ (thus implicitly enforcing periodic boundary conditions on X). Moreover, we assume X to be stationary and isotropic. These are in general strong assumptions, so one has to be careful on their implications for any statistical analysis.

With Δ the spatial step of the data, we first introduce a binning of the interval $[0, \pi/\Delta]$, that is a finite strictly increasing sequence $\{k_0, \dots, k_Q\}$ with $k_0 = 0$ and $k_Q = \pi/\Delta$. We associate to this binning a sequence of bin centers $\tilde{T} = \{\tilde{k}_0, \dots, \tilde{k}_{Q-1}\}$ defined by $\tilde{k}_m = (k_m + k_{m+1})/2$. We now define the estimator of the power spectrum of X , often simply called "the power spectrum of x " or "the empirical power spectrum of X ", as

$\hat{S}_X : \tilde{T} \rightarrow [0, +\infty)$ such that for all $\tilde{k}_m \in \tilde{T}$:

$$\hat{S}_X(\tilde{k}_m) = \frac{1}{MN} \frac{1}{|\Gamma_{\tilde{k}_m}|} \sum_{\mathbf{k} \in \Gamma_{\tilde{k}_m}} |\hat{x}(\mathbf{k})|^2, \quad (\text{A.21})$$

where $\Gamma_{\tilde{k}_m} = \{\mathbf{k} \mid k_m \leq |\mathbf{k}| < k_{m+1}\}$ and $|\Gamma_{\tilde{k}_m}|$ its cardinality.⁴ We also estimate uncertainties $\sigma_{\hat{S}_X}$ for \hat{S}_X computing the standard deviation of the mean within each bin, that is:

$$\sigma_{\hat{S}_X}(\tilde{k}_m) = \frac{1}{|\Gamma_{\tilde{k}_m}|} \left[\sum_{\mathbf{k} \in \Gamma_{\tilde{k}_m}} \left(\frac{1}{MN} |\hat{x}(\mathbf{k})|^2 - \hat{S}_X(\tilde{k}_m) \right)^2 \right]^{1/2}. \quad (\text{A.22})$$

We illustrate this algorithm in Fig. A.1 showing how it applies to a simulated intensity map extracted from the I_{\parallel} data set that is introduced in Chapter III. The power spectrum (right) is built from azimuthal averages on the map of the squared modulus of the DFT of the initial image (middle). Only modes within the Nyquist disc, which is delimited by the circle drawn in dashed line in the middle panel, are used to estimate the power spectrum shown in the right panel. Note that we see on this example a sharp anisotropy in the map of the squared Fourier modulus, showing that on this example the isotropic assumption might be questionable with a single realization.

A.2.3 Bispectrum: definition

One can define, by analogy with the power spectrum, higher order descriptive statistics based on the Fourier transform of a stationary random field. For reference, we introduce here the bispectrum, which is a quite common statistics to assess the non-Gaussianity of a stationary random field. Indeed, when it is defined as the Fourier transform of the third-order cumulant, it is expected to be null for a stationary gaussian random field (see below). For convenience, we define the bispectrum here in a slightly different way, although it does not change its statistical content at all.

Definition A.2.3. The *bispectrum* of a stationary random field X , called B_X , is defined as the Fourier transform of the three-point correlation function $C_{3,X}$, that is:

$$B_X = \mathcal{F}[C_{3,X}]. \quad (\text{A.23})$$

For a stationary random field, $C_{3,X}$ is effectively a function of two variables, that is $C_{3,X}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = C_{3,X}(\mathbf{r}, \mathbf{r} + \boldsymbol{\tau}_1, \mathbf{r} + \boldsymbol{\tau}_2)$ for any $\mathbf{r} \in T$, and the Fourier transform applies both on $\boldsymbol{\tau}_1$ and on $\boldsymbol{\tau}_2$.

For a stationary random image X of size $M \times N$ defined on $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ we have, similarly to the case of the power spectrum, the useful following proposition.

Proposition A.2.2. For a stationary random field defined on $T = \mathbb{Z}/M\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, we have for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in T$:

$$B_X(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \frac{1}{MN} \mathbb{E} \left[\hat{X}(\boldsymbol{\xi}_1) \hat{X}(\boldsymbol{\xi}_2) \overline{\hat{X}(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2)} \right]. \quad (\text{A.24})$$

Note that, for this property to be valid rigorously in the case of a complex-valued random field, we must redefine $C_{3,X}$ as $C_{3,X}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \mathbb{E} [\overline{X(\mathbf{r}_1)} X(\mathbf{r}_2) X(\mathbf{r}_3)]$.

⁴Mind the conflict of notation between the hat symbol denoting both a statistical estimator \hat{S}_X and a Fourier mode $\hat{x}(\mathbf{k})$.

A.3 Gaussian random fields

Gaussian random fields (GRFs) are important examples of random fields. They are widely used in statistical modeling, and their properties can be easily derived analytically. By opposition, we will say that a random field is non-Gaussian if it is not a GRF.

A.3.1 Definition

Before giving the definition of a GRF, we recall those of Gaussian random variables and Gaussian random vectors.

Definition A.3.1. A real-valued *Gaussian* (or *normal*) random variable X is a random variable with a Gaussian distribution, whose PDF is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \text{ for all } x \in \mathbb{R}, \quad (\text{A.25})$$

with $m = \mathbb{E}[X]$ the mean of X , and $\sigma^2 = \text{Var}[X]$ its variance.

When X is Gaussian (or normally) distributed with mean m and variance σ^2 , we denote this by $X \sim \mathcal{N}(m, \sigma^2)$.

Definition A.3.2. A real-valued random vector (X_1, \dots, X_n) is said to be *Gaussian* if and only if every linear combination of its components $\alpha_1 X_1 + \dots + \alpha_n X_n$ with $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ is a real-valued Gaussian random variable.

For a real-valued random (column) vector $\mathbf{X} = (X_1, \dots, X_n)^T$, one can define the mean vector $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ and the covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$. These elements fully characterize the distribution of \mathbf{X} and we note $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. When the covariance matrix $\boldsymbol{\Sigma}$ is not singular, the joint PDF of \mathbf{X} has an analog expression to that of Eq. (A.25).

Proposition A.3.1. The joint PDF of a real-valued Gaussian random vector (X_1, \dots, X_n) when its covariance matrix $\boldsymbol{\Sigma}$ is not singular reads:

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right). \quad (\text{A.26})$$

Complex-valued Gaussian random variables and vectors are defined from their real-valued analogs as follows.

Definition A.3.3. A complex-valued random variable X is said to be *Gaussian* if and only if $(\text{Re}(X), \text{Im}(X))$ is a real-valued Gaussian random vector. Similarly, (X_1, \dots, X_n) is a complex-valued *Gaussian* random vector if and only if the real-valued random vector $(\text{Re}(X_1), \dots, \text{Re}(X_n), \text{Im}(X_1), \dots, \text{Im}(X_n))$ is Gaussian.

Complex-valued Gaussian random vectors can thus be viewed as real-valued Gaussian random vectors and characterized accordingly. However, within a fully complex formalism, a complex-valued Gaussian random vector $\mathbf{X} = (X_1, \dots, X_n)^T$ is defined by its mean vector $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$, its covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\overline{\mathbf{X} - \boldsymbol{\mu}})^T]$, and

its pseudo-covariance (or relation) matrix $\Sigma' = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$, and we write $\mathbf{X} \sim \mathcal{CN}(\boldsymbol{\mu}, \Sigma, \Sigma')$.

Now let us define in this context a Gaussian random field, which can be either real-valued ($E = \mathbb{R}$) or complex-valued ($E = \mathbb{C}$).

Definition A.3.4. A Gaussian random field (GRF) $X : T \rightarrow E$ is a random field such that every linear combination of the form $\alpha_1 X(\mathbf{r}_1) + \dots + \alpha_n X(\mathbf{r}_n)$ with $n \in \mathbb{N}$, $\alpha_1, \dots, \alpha_n \in E$ and $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$ is a Gaussian random variable.

Equivalently, $X : T \rightarrow E$ is a GRF if and only if for all $n \in \mathbb{N}$ and $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$, $(X(\mathbf{r}_1), \dots, X(\mathbf{r}_n))$ is a Gaussian random vector.

A.3.2 Characterization

A GRF X is determined by its mean function, its autocovariance function, and its pseudo-autocovariance function (which is equal to the autocovariance function in the real-valued case). Indeed, its finite-dimensional distributions are all multivariate Gaussian distributions depending on μ_X , C_X , and C'_X only. More precisely, the PDF of the Gaussian random vector $(X(\mathbf{r}_1), \dots, X(\mathbf{r}_n))$ only depends on the mean vector $(\mu_X(\mathbf{r}_1), \dots, \mu_X(\mathbf{r}_n))$, and the autocovariance and pseudo-autocovariance matrices defined by $\Sigma_{ij} = C_X(\mathbf{r}_i, \mathbf{r}_j)$ and $\Sigma'_{ij} = C'_X(\mathbf{r}_i, \mathbf{r}_j)$, respectively. Note that a direct consequence of this characterization is that if a GRF X is WSS then it is strict sense stationary.

In this thesis, we often claim that to characterize non-Gaussianity we need to go beyond power spectra measurements. It is a consequence of this previous characterization. Indeed, for real-valued GRFs, this implies that a stationary real-valued GRF is characterized by its mean and power spectrum. For complex-valued GRFs, this is true only when the pseudo-autocovariance function is null, which is usually assumed.⁵

Proposition A.3.2. A stationary GRF X that is either real-valued or complex-valued with null pseudo-autocovariance function is entirely defined by its mean μ_X and its power spectrum S_X .

Let us finally mention a useful result of probability theory for Gaussian random vectors, called Isserlis' theorem or Wick's probability theorem. This theorem emphasizes the comprehensiveness of second-order statistics to characterize a GRF.

Theorem A.3.1 (Isserlis, Wick). If (X_1, \dots, X_n) is a zero-mean real-valued Gaussian random vector, then:

$$E[X_1 \dots X_n] = \sum_{p \in P_n} \prod_{\{i,j\} \in p} E[X_i X_j], \quad (\text{A.27})$$

where P_n is the set of all partitions of $[1, n]$ into subsets of two elements.

Let us consider (X_1, \dots, X_n) a zero-mean real-valued Gaussian random vector. Note that for odd n value, that is $n = 2k + 1$, we obviously have $P_{2k+1} = \emptyset$, and this theorem thus implies that:

$$E[X_1 X_2 \dots X_{2k+1}] = 0. \quad (\text{A.28})$$

⁵A null pseudo-autocovariance functions may result from circular symmetry, that is the invariance of the distribution of the random field to a multiplication by a global phase factor $e^{i\phi}$ with $\phi \in \mathbb{R}$, which is relatively common.

On the other hand, for $n = 2k$, this theorem states that $\mathbb{E}[X_1 \dots X_{2k}]$ may be written as a sum of $|P_{2k}| = (2k - 1)!!$ second-order moments. For example, we have for $n = 4$:

$$\mathbb{E}[X_1 X_2 X_3 X_4] = \mathbb{E}[X_1 X_2] \mathbb{E}[X_3 X_4] + \mathbb{E}[X_1 X_3] \mathbb{E}[X_2 X_4] + \mathbb{E}[X_1 X_4] \mathbb{E}[X_2 X_3]. \quad (\text{A.29})$$

For a GRF, the consequence of this is that any n -point correlation function is either null (for odd n), or, can be written in terms of the two-point correlation function (for even n). This result implies that the bispectrum of a zero-mean GRF is always zero, underlining its relevance to characterize non-Gaussianity. Note that this theorem has extensions in the case of complex-valued Gaussian random vectors (see e.g. [223]).

A.3.3 Examples

We now discuss a few typical examples of GRFs which are mentioned in this manuscript.

The most simple GRF we can define is the Gaussian white noise. It is simply a zero-mean stationary GRF X that is completely spatially uncorrelated. Therefore for any $\mathbf{r}_1, \mathbf{r}_2 \in T$ with $\mathbf{r}_1 \neq \mathbf{r}_2$, $X(\mathbf{r}_1)$ and $X(\mathbf{r}_2)$ are independent Gaussian random variables, and $C_X(\mathbf{r}_2 - \mathbf{r}_1) = R_X(\mathbf{r}_2 - \mathbf{r}_1) = 0$. Alternatively, Gaussian white noises can be defined in an equivalent way as follows.

Definition A.3.5. A Gaussian white noise $X : T \rightarrow E$ is a zero-mean stationary GRF with a constant power spectrum S_X .

We also give the definition of fractional Brownian fields which have been notably employed for the modeling of the dust emission (see Chapter II). These are well known examples of self-similar random fields (see e.g. [224]).

Definition A.3.6. A fractional Brownian field $X : T \rightarrow E$ is a Gaussian random field defined for a given $H \in (0, 1)$ by:

- $\mu_X(\mathbf{r}) = 0$ for all $\mathbf{r} \in T$,
- $C_X(\mathbf{r}_1, \mathbf{r}_2) = \frac{V}{2} \left(\|\mathbf{r}_1\|^{2H} + \|\mathbf{r}_2\|^{2H} - \|\mathbf{r}_1 - \mathbf{r}_2\|^{2H} \right)$ for all $\mathbf{r}_1, \mathbf{r}_2 \in T$ and $V \in \mathbb{R}^+$.

The real number H is called the Hurst exponent of X . When $V = 1$, we say that X is a standard fractional Brownian field.

When T is an interval of \mathbb{R} , a fractional Brownian field is simply called a fractional Brownian motion. However, in this thesis, we will abusively call fractional Brownian motions (fBm) any fractional Brownian field.

A fBm is an example of H -self-similar random field. While it is clearly non-stationary, its increments, on the other hand, are stationary, meaning that for any $\mathbf{r}_1, \boldsymbol{\tau} \in T$, $(X(\mathbf{r}_1 + \boldsymbol{\tau}) - X(\mathbf{r}_2 + \boldsymbol{\tau}))_{\mathbf{r}_2 \in T}$ and $(X(\mathbf{r}_1) - X(\mathbf{r}_2))_{\mathbf{r}_2 \in T}$ have the same distribution.

A.3.4 Simulation

Another asset of GRFs is that they are relatively easy to simulate. We discuss such simulation algorithms here in the case of real-valued GRFs, and we refer to [225] for additional details.

Obviously, even if the index set T is infinite, a computer will necessarily have to approximate T through the definition of a relevant finite number of grid points $\mathbf{r}_1, \dots, \mathbf{r}_n \in T$. This way, the simulation of a real-valued GRF X defined on T amounts to the simulation of a real-valued Gaussian random vector $\tilde{\mathbf{X}} = (X_1, \dots, X_n)^T$ with $X_i = X(\mathbf{r}_i)$. In the following we identify X to its discrete approximation $\tilde{\mathbf{X}}$.

In the general case, with $\boldsymbol{\mu}$ the mean vector of \mathbf{X} , and $\boldsymbol{\Sigma}$ its covariance matrix, one can simply simulate \mathbf{X} provided one knows a square root \mathbf{A} of $\boldsymbol{\Sigma}$, that is a matrix \mathbf{A} such that $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$.⁶ Indeed the random vector $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$ with $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ and $Z_1, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ has by construction the same distribution as \mathbf{X} since we trivially have:

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\boldsymbol{\mu}] + \mathbb{E}[\mathbf{A}\mathbf{Z}] = \boldsymbol{\mu} + \mathbf{A}\mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}, \quad (\text{A.30})$$

$$\mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^T] = \mathbf{A}\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]\mathbf{A}^T = \boldsymbol{\Sigma}. \quad (\text{A.31})$$

However, in practice finding a square root of $\boldsymbol{\Sigma}$ can be a quite computationally demanding step. It usually requires to find the Cholesky decomposition of $\boldsymbol{\Sigma}$ which involves algorithms with a complexity $\mathcal{O}(n^3)$ (for the most general case).

Fortunately, there exist much more efficient algorithms to simulate GRFs provided they are stationary. Let us illustrate this on the example of the generation of a stationary Gaussian random image X defined on a square, with periodic boundary conditions, that is with $T = \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ (see [225]). In that case, T is a finite set so we do not need to make discretization approximations. To simplify, let us assume that X is zero-mean since a non-zero mean can be trivially added a posteriori. We first arrange the grid points $(i, j) \in T$ in the order $(0, 0), (0, 1), \dots, (0, N-1), (1, 0), \dots, (N-1, N-1)$, so that $\boldsymbol{\Sigma}$ can be expressed as a $n \times n$ matrix with $n = N^2$ and $\Sigma_{kl} = \mathbb{E}[X(\lfloor k/N \rfloor, k \bmod N)X(\lfloor l/N \rfloor, l \bmod N)]$. This way, the covariance matrix $\boldsymbol{\Sigma}$ is a symmetric block-circulant matrix with circulant blocks, meaning it can be written as:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 & \dots & \mathbf{C}_N \\ \mathbf{C}_N & \mathbf{C}_1 & \mathbf{C}_2 & \dots & \mathbf{C}_{N-1} \\ & \dots & & & \\ \mathbf{C}_2 & \mathbf{C}_3 & \dots & \mathbf{C}_{N-1} & \mathbf{C}_1 \end{pmatrix}, \quad (\text{A.32})$$

with \mathbf{C}_i a circulant matrix of size $N \times N$, which is thus defined by its first row. This matrix is entirely defined by its first row. Let us arrange this first row in a $N \times N$ matrix \mathbf{G} . The eigenstructure of a matrix such as $\boldsymbol{\Sigma}$ is well known and one can write:

$$\boldsymbol{\Sigma} = \bar{\mathbf{P}} \text{diag}(\boldsymbol{\gamma}) \mathbf{P}, \quad (\text{A.33})$$

with \mathbf{P} the Kronecker product of two DFT matrices \mathbf{F} (see Appendix B for a definition), and $\boldsymbol{\gamma}$ the vector of eigenvalues. When arranged in a square matrix $\boldsymbol{\Gamma}$ of size $N \times N$, this vector of eigenvalues can be determined by $\boldsymbol{\Gamma} = N\bar{\mathbf{F}}\mathbf{G}\mathbf{F}$. Finally, $\mathbf{A} = \bar{\mathbf{P}} \text{diag}(\sqrt{\boldsymbol{\gamma}})$ is a square root of $\boldsymbol{\Sigma}$, from which one can easily generate realizations of X as explained above. Such algorithms thus avoid an expensive Cholesky decomposition, by relying only on two-dimensional DFT computations. With the FFT2 algorithm, the complexity of

⁶Note that a square root always exists as the covariance matrix is a positive semidefinite symmetric real matrix.

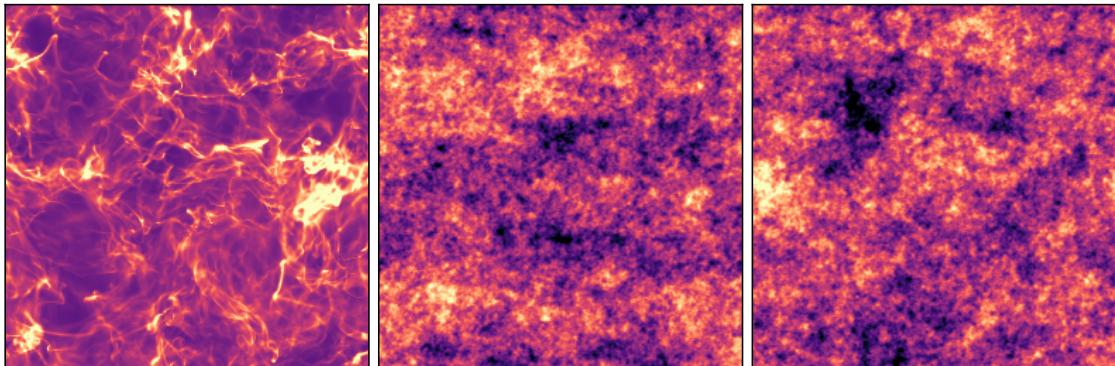


Fig. A.2 *Simulation of a GRF (middle) based on the empirical autocorrelation function estimated from the left image. We show in the right panel for comparison the phase randomized image of the left one.*

this simulation method becomes $\mathcal{O}(n \log(n))$, which is a great improvement of complexity compared to the $\mathcal{O}(n^3)$ complexity of Cholesky decomposition algorithms.

In a more general case, without necessarily assuming periodic boundary conditions for X , or square images, a common and more sophisticated algorithm also valid in higher dimensions is that of circulant embedding [226]. This algorithm consists in embedding the target covariance matrix in a larger matrix that is block-circulant with circulant blocks. Then, similarly to above, the eigenstructure of this new matrix can be efficiently determined with the FFT. This approach has been extended to the case of complex-valued GRFs in [227, 228].

A.3.5 Gaussianization

We now discuss two methods of "Gaussianization", that is, given some data x , the generation of a realization \tilde{x} of a GRF \tilde{X} having consistent first and second-order statistics with those of x .

The first method makes use of the simulation techniques described in the previous subsection, and is the most rigorous. Then, we discuss an approximate method, called phase randomization, that is employed in Chapter III for its simplicity.

a) Rigorous method

Given an image x , assumed to be real-valued and with a zero-mean for simplicity, we wish to generate new Gaussian realizations \tilde{x} that have similar second-order statistics to those of x . We assume x to be a realization of a stationary random field X , the case of non-stationary data being intractable in general.

As explained in Sect. A.3.4, we are able to generate realizations of stationary GRFs based on a covariance matrix Σ describing their second-order statistics. Therefore, we only need to define a relevant Σ derived from x , meaning we need to estimate the auto-covariance function of X .

Assuming an ergodic X , we estimate C_X with the following natural estimator \hat{C}_X :

$$\hat{C}_X(\mathbf{r}_2 - \mathbf{r}_1) = \frac{1}{|T|^2} \sum_{\mathbf{r}_1 \in T} \sum_{\mathbf{r}_2 \in T} x(\mathbf{r}_1)x(\mathbf{r}_2). \quad (\text{A.34})$$

In practice, with an additional periodic boundary condition assumption on X , we efficiently compute \hat{C}_X with the FFT using the fact that:

$$\hat{C}_X = \frac{1}{|T|} \mathcal{F}^{-1} \left[|\mathcal{F}[x]|^2 \right], \quad (\text{A.35})$$

where \mathcal{F} is the DFT. This allows a complexity in $\mathcal{O}(|T| \log(|T|))$ instead of $\mathcal{O}(|T|^2)$ for a naive algorithm directly relying on Eq. (A.34).

We present in Fig. A.2 a numerical application of this method. We show the test image (left) used for this example (same one as that introduced in Sect. A.2.2), next to the resulting Gaussianization of the image (middle). This relies on the estimation of the autocovariance function C_X as explained here and the subsequent generation of a GRF realization with the technique described in Sect. A.3.4. Note that here the test map has highly non-Gaussian statistics, so that the effect of this Gaussianization is to strongly damage the statistical properties of the test map. In particular, the filamentary structure of the image is completely suppressed by this procedure.

b) Approximate method: phase randomization

We introduce another Gaussianization method, called phase randomization, which is more straightforward to implement and gives satisfactory results for the purposes of this thesis (see Chapter III).

Given a real-valued image x of size $N \times N$, assuming periodic boundary conditions for simplicity, so that its domain of definition is $T = \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, we define a random image y in Fourier space by:

$$\hat{y}(\boldsymbol{\xi}) = |\hat{x}(\boldsymbol{\xi})| e^{i\phi(\boldsymbol{\xi})}, \quad (\text{A.36})$$

for every mode $\boldsymbol{\xi} \in T$ and with $\phi(\boldsymbol{\xi})$ drawn from a uniform distribution on $[0, 2\pi]$. In fact, ϕ must verify some additional constraints to properly define a real-valued image y . In [229], where such phase-randomized fields are studied, ϕ is called a realization of a *uniform random phase*. With $T_0 = \{(0, N/2), (N/2, 0), (N/2, N/2)\}$ if N is even, and $T_0 = \emptyset$ otherwise, it is defined by the following constraints:

1. $\forall \boldsymbol{\xi} \in T \setminus T_0, \phi(-\boldsymbol{\xi}) = -\phi(\boldsymbol{\xi})$,
2. $\forall \boldsymbol{\xi} \in T, \phi(\boldsymbol{\xi})$ is drawn uniformly and independently in $[0, 2\pi]$ (obviously the independence holds for modes that are not related by 1.),
3. $\forall \boldsymbol{\xi} \in T_0, \phi(\boldsymbol{\xi})$ is drawn uniformly and independently in the set $\{0, \pi\}$.

In addition to ensuring that y will be real-valued, note that these relations also ensure that the mean of y will be equal to the absolute value of the mean of x (since condition 1. gives $\phi(0) = 0$).

The motivation behind such a definition is twofold. First, the phases of a stationary GRF are known to be independently drawn from a uniform distribution on $[0, 2\pi)$ (see

e.g. [187]). Second, this procedure will retain estimates of the power spectrum such as defined in Sect. A.2.2. However let us emphasize that from a mathematical point of view, random fields Y defined from a single x as described here are not stationary GRFs. Indeed, the moduli of Fourier modes of a GRF cannot be deterministic. Typically, when stationary they must follow Rayleigh distributions.

We show in Fig. A.2 (right) an example of phase-randomized image. We see that it is visually consistent with a more rigorous Gaussianization as described above.

Appendix B

Elements of Fourier and wavelet analysis

The idea that signals can generally be represented by a sum of oscillating modes of multiple frequencies is key to signal processing. In Fourier analysis, signals are viewed as sums of simple trigonometric functions called *Fourier modes*, while in wavelet analysis, they are decomposed into *wavelets*.

In this appendix, we first recall classical definitions and results revolving around the Fourier transform. We give results without proof, referring to [230] and references therein for proofs and additional details. Then, we give some technical complements on the wavelets used for this thesis, and illustrate their action when they are convolved to images.

We denote by $L^p(\mathbb{R}^n)$, with $p \geq 1$, the set of measurable functions $f : \mathbb{R}^n \rightarrow \mathbb{C}$ for which the p -norm defined by $\|f\|_p = (\int_{\mathbb{R}^n} |f(\mathbf{r})|^p d\mathbf{r})^{1/p}$ is finite. We recall that these sets $L^p(\mathbb{R}^n)$ are all vector spaces. Functions of $L^1(\mathbb{R}^n)$ are simply called *integrable* functions. For $f \in L^2(\mathbb{R}^n)$, we define the energy of f by its squared 2-norm $\|f\|_2^2$, which is thus a finite positive number. In $L^2(\mathbb{R}^n)$ the 2-norm $\|\cdot\|_2$ coincides with the norm associated to the inner product defined for all f and g in $L^2(\mathbb{R}^n)$ by:

$$\langle f, g \rangle = \int_{\mathbb{R}^n} f(\mathbf{r}) \overline{g(\mathbf{r})} d\mathbf{r}, \quad (\text{B.1})$$

and $(L^2(\mathbb{R}^n), \langle \cdot, \cdot \rangle)$ defines a Hilbert space.

B.1 The Fourier transform

B.1.1 Definition on $L^2(\mathbb{R}^n)$ and elementary properties

a) Definition

Physical fields are usually well modeled by functions of $L^2(\mathbb{R}^n)$, which are by definition functions of finite energy. This motivates the definition of the Fourier transform on this specific functional space. In addition to this, we will see that, in this space, the Fourier transform also has very convenient properties.

The definition of the Fourier transform in $L^2(\mathbb{R}^n)$ is not straightforward, as it usually first requires the definition of the Fourier transform in $L^1(\mathbb{R}^n)$.

Definition B.1.1. With f an integrable function, that is $f \in L^1(\mathbb{R}^n)$, the Fourier transform of f is the function $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{C}$ defined for all $\mathbf{k} \in \mathbb{R}^n$ by:

$$\hat{f}(\mathbf{k}) = \int_{\mathbb{R}^n} f(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}. \quad (\text{B.2})$$

We call $\hat{f}(\mathbf{k})$ a *Fourier mode* of f . This definition does not pose any problem as $\mathbf{r} \rightarrow f(\mathbf{r})e^{-i\mathbf{k} \cdot \mathbf{r}}$ is trivially integrable for any $\mathbf{k} \in \mathbb{R}^n$. Provided \hat{f} is in $L^1(\mathbb{R}^n)$, the Fourier transform can be inverted.

Proposition B.1.1. Provided $\hat{f} \in L^1(\mathbb{R}^n)$, Eq. (B.2) can be inverted as follows:

$$f(\mathbf{r}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}. \quad (\text{B.3})$$

Now, extending this definition to functions $f \in L^2(\mathbb{R}^n)$ is not direct since $\mathbf{r} \rightarrow f(\mathbf{r})e^{-i\mathbf{k} \cdot \mathbf{r}}$ is not generally integrable. We proceed thanks to the density of $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$.

Definition B.1.2. For $f \in L^2(\mathbb{R}^n)$, and $\{f_n\}_{n \in \mathbb{N}}$ a sequence of functions of $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ converging towards f (2-norm convergence), the Fourier transform of $f \in L^2(\mathbb{R}^n)$ is defined as the limit \hat{f} of the sequence of functions $\{\hat{f}_n\}_{n \in \mathbb{N}}$.

b) Elementary properties

Fortunately, for $f \in L^2(\mathbb{R}^n)$, \hat{f} is also in $L^2(\mathbb{R}^n)$, and we thus define the corresponding Fourier transform operator $\mathcal{F} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ by $\mathcal{F}[f] = \hat{f}$ for all $f \in L^2(\mathbb{R}^n)$. Let us now review elementary properties of the operator \mathcal{F} that interest us in this thesis.

Proposition B.1.2. \mathcal{F} is a linear operator, that is for any $f, g \in L^2(\mathbb{R}^n)$ and $\alpha, \beta \in \mathbb{C}$:

$$\mathcal{F}[\alpha f + \beta g] = \alpha \mathcal{F}[f] + \beta \mathcal{F}[g]. \quad (\text{B.4})$$

It is also invertible, and we call \mathcal{F}^{-1} its inverse.

Note that thanks to this invertibility property we may define functions from their Fourier representations.

Moreover, when properly normalized, \mathcal{F} is a unitary operator, meaning that added to be linear and invertible, it preserves the inner product of $L^2(\mathbb{R}^n)$.

Proposition B.1.3. With $\tilde{\mathcal{F}} = (2\pi)^{-n/2} \mathcal{F}$, for all f and g in $L^2(\mathbb{R}^n)$:

$$\langle \tilde{\mathcal{F}}[f], \tilde{\mathcal{F}}[g] \rangle = \langle f, g \rangle. \quad (\text{B.5})$$

This relation is called Parseval's formula. Taking $f = g$, we deduce the Plancherel's formula, which traduces the energy conservation of the Fourier transform. This reads:

$$\int_{\mathbb{R}^n} |f(\mathbf{r})|^2 d\mathbf{r} = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} |\hat{f}(\mathbf{k})|^2 d\mathbf{k}. \quad (\text{B.6})$$

Another important property of the Fourier transform derives from its application to convolutions. We recall that the convolution $f \star g$ of two functions $f, g \in L^2(\mathbb{R}^n)$ is defined by:

$$f \star g(\mathbf{r}) = \int_{\mathbb{R}^n} f(\mathbf{r}') g(\mathbf{r} - \mathbf{r}') d\mathbf{r}', \quad (\text{B.7})$$

for any $\mathbf{r} \in \mathbb{R}^n$.

Proposition B.1.4. The Fourier transform of a convolution of two functions is the product of their respective Fourier transforms, that is with $f, g \in L^2(\mathbb{R}^n)$:

$$\mathcal{F}[f * g] = \mathcal{F}[f] \times \mathcal{F}[g]. \quad (\text{B.8})$$

Finally, let us mention the hermitian symmetry property of the Fourier transform when it is applied to a real-valued function.

Proposition B.1.5. With $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a real-valued function of $L^2(\mathbb{R}^n)$, we have:

$$\mathcal{F}[f](-\mathbf{k}) = \overline{\mathcal{F}[f](\mathbf{k})}. \quad (\text{B.9})$$

B.1.2 Finite discrete signals and the discrete Fourier transform

One-dimensional digitized data takes the form of finite sequences, from which one can define an equivalent of the previous Fourier transform. When these finite sequences represent equally-spaced samples, this equivalent is the *discrete Fourier transform* (DFT).

a) Definition

Definition B.1.3. The DFT of a complex or real-valued finite sequence $\{x[n]\}_{0 \leq n < N}$ is the complex-valued sequence $\{\hat{x}[k]\}_{0 \leq k < N}$ defined by:

$$\hat{x}[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-\frac{2i\pi kn}{N}\right). \quad (\text{B.10})$$

This relation can always be inverted, thus defining the inverse DFT:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{x}[k] \exp\left(\frac{2i\pi kn}{N}\right). \quad (\text{B.11})$$

The DFT can be formulated in a matrix form. With $X = (x[0], \dots, x[n])^\top$ the column vector representing $\{x[n]\}_{0 \leq n < N}$, and $\hat{X} = (\hat{x}[0], \dots, \hat{x}[n])^\top$ that representing $\{\hat{x}[k]\}_{0 \leq k < N}$, we have:

$$\hat{X} = F_N X \quad \text{with} \quad F_N = \left(\exp\left(-\frac{2i\pi mn}{N}\right) \right)_{0 \leq m, n < N}. \quad (\text{B.12})$$

F_N is called a *DFT matrix*. This matrix is invertible, and when normalized by $1/\sqrt{N}$ the resulting matrix is unitary.

We can extend this definition to multi-dimensional data.

Definition B.1.4. Let us consider x an array of size $N_1 \times \dots \times N_d$, denoting by $x[n_1, \dots, n_d]$ its elements, with $0 \leq n_1 < N_1, \dots, 0 \leq n_d < N_d$. The DFT of x is the array \hat{x} defined by:

$$\hat{x}[k_1, \dots, k_d] = \sum_{n_1=0}^{N_1-1} \dots \sum_{n_d=0}^{N_d-1} x[n_1, \dots, n_d] \exp\left(-\frac{2i\pi k_1 n_1}{N_1} - \dots - \frac{2i\pi k_d n_d}{N_d}\right), \quad (\text{B.13})$$

with $0 \leq k_1 < N_1, \dots, 0 \leq k_d < N_d$.

The inverse of the multi-dimensional DFT extends similarly with:

$$x[n_1, \dots, n_d] = \frac{1}{N_1 \times \dots \times N_d} \sum_{k_1=0}^{N_1-1} \dots \sum_{k_d=0}^{N_d-1} \hat{x}[k_1, \dots, k_d] \exp\left(\frac{2i\pi k_1 n_1}{N_1} + \dots + \frac{2i\pi k_d n_d}{N_d}\right). \quad (\text{B.14})$$

These relations can also be formulated in a matrix form by defining a multi-dimensional DFT matrix relying on the Kronecker product of one-dimensional DFT matrices.

b) Properties

The same properties introduced for the Fourier transform on $L^2(\mathbb{R}^n)$ remain valid for the DFT in a comparable form.

For example, in a multi-dimensional context, Eq. (B.6), which traduces the conservation of energy, becomes:

$$\sum_{n_1=0}^{N_1-1} \dots \sum_{n_d=0}^{N_d-1} |x[n_1, \dots, n_d]|^2 = \frac{1}{N_1 \times \dots \times N_d} \sum_{k_1=0}^{N_1-1} \dots \sum_{k_d=0}^{N_d-1} |\hat{x}[k_1, \dots, k_d]|^2. \quad (\text{B.15})$$

For Eq. (B.8) to remain valid for the DFT, the discrete convolution operation must be a *circular convolution*. In one dimension, for $x = \{x[n]\}_{0 \leq n < N}$ and $y = \{y[n]\}_{0 \leq n < N}$ two finite signals of length N , the circular convolution of x and y , also denoted by $x \star y$, is the finite sequence of length N defined as:

$$x \star y[n] = \sum_{p=0}^{N-1} x[p]y[(n-p) \bmod N]. \quad (\text{B.16})$$

This corresponds to a discrete convolution where periodic boundary conditions are assumed. For circular convolutions, we thus have for all $k \in \llbracket 0, N-1 \rrbracket$:

$$\widehat{x \star y}[k] = \hat{x}[k]\hat{y}[k]. \quad (\text{B.17})$$

c) The fast Fourier transform

Numerically, computing the DFT of a one-dimensional sequence of length N directly from the definition written in Eq. (B.10) would a priori require N multiplications and N additions for each mode. Therefore, multiplying this number by the total number of independent modes, the complexity of such a naive algorithm would be $\mathcal{O}(N^2)$. Fortunately, there exists a much more efficient algorithm to compute this DFT, called fast Fourier transform (FFT, see [230]). Indeed, the FFT has a $\mathcal{O}(N \log(N))$ complexity, which generalizes to the multi-dimensional case as $\mathcal{O}(N_1 \times \dots \times N_d \log(N_1 \times \dots \times N_d))$. Moreover, thanks to Eq. (B.17), it also allows us to compute convolutions efficiently, meaning with the same complexity as the FFT.

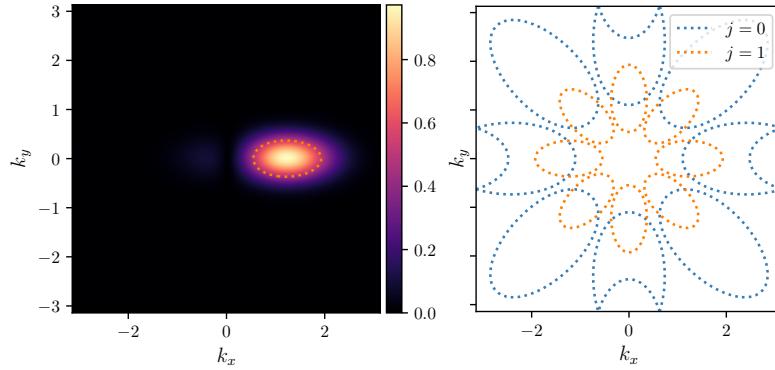


Fig. B.1 *Morlet wavelets in Fourier space.* In the left panel, we show the modulus of the Fourier transform of the Morlet wavelet $\psi_{1,0}$. We also add in dashed lines the contour of this function at half-maximum level. In the right panel, we illustrate how Morlet wavelets pave Fourier space showing similar contours for wavelets $\psi_{j,\theta}$ with $j \in \{0, 1\}$ and $\theta \in \{\frac{k\pi}{4}, k \in [0, 7]\}$.

B.2 Complements on wavelets

The wavelet transform decomposes a function into a set of *time-frequency* atoms called wavelets, thus allowing a local separation of the multiscale variability of some given data. It is at the basis of the WST and WPH statistics.

In this section, we give complementary information on the wavelets used for this thesis, and illustrate how they act as bandpass filters.

B.2.1 Wavelets used for this thesis

We give complementary details on Morlet and bump-steerable wavelets which were already introduced in Sect. III.1.2 and Sect. IV.3.1, respectively.

a) Morlet wavelets

A 2D mother Morlet wavelet ψ is generally defined by:

$$\psi(\mathbf{r}) = \alpha \left(e^{ik_0 r_x} - \beta \right) e^{-(r_x^2 + \gamma^2 r_y^2)/2\sigma^2}, \quad (\text{B.18})$$

with α and β two constants that are adjusted to ensure a zero mean and a unit L^1 norm, k_0 the wave number of the plane wave factor, σ the standard deviation of the Gaussian envelope, and γ the aspect ratio of this envelope. In practice we choose $k_0 = 3\pi/4 \text{ pixel}^{-1}$ and $\sigma = 0.8 \text{ pixel}$ as in [4]. We also choose $\gamma = 0.5$ to give an elliptical shape to the envelope, thus enhancing the angular selectivity of the wavelet.

In Fourier space, such a Morlet wavelet reads:

$$\hat{\psi}(\mathbf{k}) = 2\pi\alpha \frac{\sigma^2}{\gamma} \left(e^{-\sigma^2(k_0 - k_x)^2/2} - \beta e^{-\sigma^2 k_x^2/2} \right) e^{-\sigma^2 k_y^2/2\gamma^2}. \quad (\text{B.19})$$

Note that $\hat{\psi}(\mathbf{k}) \in \mathbb{R}$ for all \mathbf{k} . We show in Fig. B.1 (left panel) the modulus of the Fourier transform of a Morlet wavelet built from the previous mother wavelet. We also show in

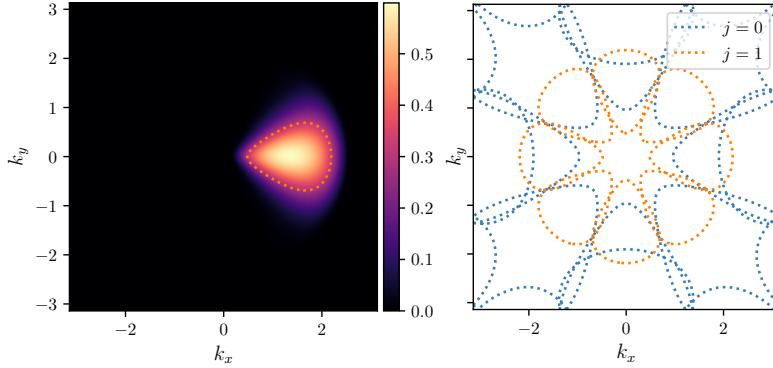


Fig. B.2 Same as Fig. B.1 but for bump-steerable wavelets.

dashed lines the contour of this function at half-maximum level. The localization of the bandpass of a Morlet wavelet makes it clearly a bandpass filter.

A bank of Morlet wavelets such as introduced in Sect. III.1.2 is designed to pave the Fourier space with the respective bandpasses of the wavelets. This is illustrated in Fig. B.1 (right panel) by showing contours at half-maximum level of the modulus of a selection of wavelets defined for $j = 0$ and $j = 1$. Note that for $j = 0$ wavelets, an aliasing effect is clearly visible when $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$. We also remark that the coverage beyond the Nyquist disc of Morlet wavelets is not optimal, which might be related to the flaws identified in the power spectrum assessment of the WST-based generative models that are discussed in Sect. IV.2.

b) Bump-steerable wavelets

We recall that a mother bump-steerable wavelet is defined in Fourier space as follows:

$$\hat{\psi}(\mathbf{k}) = \exp\left(\frac{-(k - \xi_0)^2}{\xi_0^2 - (k - \xi_0)^2}\right) \cdot 1_{[0, 2\xi_0]}(k) \times \cos^{L-1}(\arg(\mathbf{k})) \cdot 1_{[0, \pi/2]}(|\arg(\mathbf{k})|), \quad (\text{B.20})$$

with $k = \|\mathbf{k}\|$, $1_A(x)$ the indicator function that returns 1 if $x \in A$ and 0 otherwise, and $\xi_0 = 0.85\pi$ the central wavenumber of the mother wavelet. Here again, we note that $\hat{\psi}(\mathbf{k}) \in \mathbb{R}$ for all \mathbf{k} .

We show in Fig. B.2 an equivalent illustration to Fig. B.1 but here for bump-steerable wavelets. A similar aliasing effect is visible for $j = 0$ wavelets. We also remark that bump-steerable wavelets better cover spatial frequencies that go beyond the Nyquist disc.

B.2.2 Convolution with a wavelet

Thanks to the convolution theorem introduced in Sect. B.1, the convolution of a given signal x with a wavelet ψ takes a simple form in Fourier space:

$$\widehat{x \star \psi} = \hat{x} \times \hat{\psi}. \quad (\text{B.21})$$

This explains why the convolution of a given signal with a wavelet reduces to a standard bandpass filtering. Figure B.3 gives an illustration of this by comparing a given test

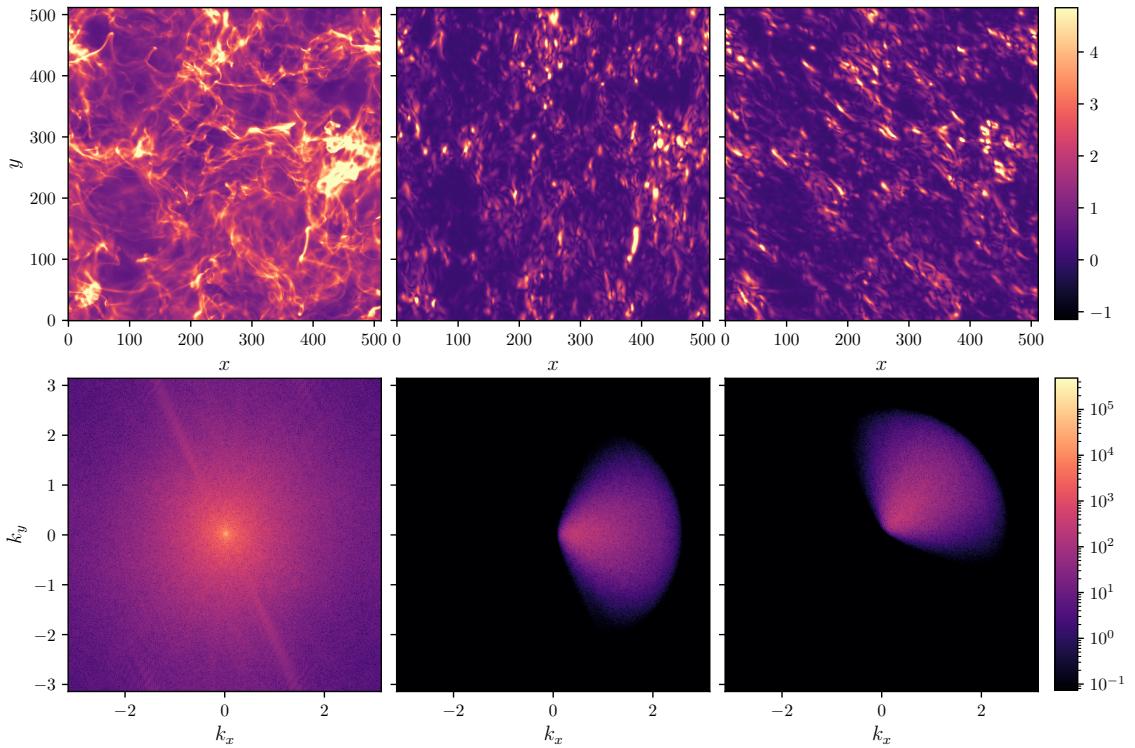


Fig. B.3 Convolution with (bump-steerable) wavelets and bandpass filtering. We show in the top row, from left to right: a test map x (left, intensity map extracted from the I_{\parallel} data set that is introduced in Chapter III), the map $|x \star \psi_{1,0}|$ (middle), and the map $|x \star \psi_{1,\pi/4}|$ (right). These are all normalized by their standard deviation to show one common color bar only. The bottom row shows the moduli of the corresponding maps in Fourier space, that is $|\mathcal{F}[x]|$ (left), $|\mathcal{F}[x \star \psi_{1,0}]|$ (middle), and $|\mathcal{F}[x \star \psi_{1,\pi/4}]|$ (right).

image x (top left) to its counterparts once it has been convolved by two bump-steerable wavelets $\psi_{1,0}$ and $\psi_{1,\pi/4}$ (top middle and right). The bottom row shows the moduli of the corresponding maps seen in Fourier space.

Appendix C

Softwares

I have been developing two public Python packages for the purposes of this thesis, called PyWST and PyWPH. Both are available on my GitHub webpage: <https://github.com/bregaldo/>.

In this appendix, we review the motivations and the main functionalities of these packages. For a technical presentation of the functionalities we refer the interested reader to the tutorials and examples provided within the GitHub repositories.

C.1 PyWST

PyWST¹ is a public Python package designed to perform statistical analyses of two-dimensional data with the Wavelet Scattering Transform (WST) and the Reduced Wavelet Scattering Transform (RWST). It relies on NumPy [231], SciPy [197] and Matplotlib [232]. It was publicly released in [168].

The initial motivation for designing PyWST was to provide a simple, flexible and self-consistent Python framework to compute, handle, and plot WST and RWST coefficients for astrophysical purposes. It was not designed for gradient-based optimizations, consequently no framework allowing automatic differentiation is employed (see Chapter IV). For the generation of statistical syntheses such as introduced in Chapter IV, we rather make use of a modified version of Kymatio [198] and its PyTorch backend [192].

PyWST allows to compute WST and RWST coefficients from real or complex-valued images such as introduced in Sect. III.1.2 and Sect. III.2.4, respectively. Coefficients can be computed either locally or globally. The package is built in a modular way so that changing wavelets or defining its own RWST model is made easily possible. It also implements convenient plotting methods permitting to easily display and compare the coefficients.

A pedagogical introduction to using PyWST taking the form of a Jupyter notebook is available [here](#).

¹<https://github.com/bregaldo/pywst/>

C.2 PyWPH

PyWPH² is a Python package designed for the computation and handling of the Wavelet Phase Harmonics (WPH) statistics such as defined in Sect. IV.3.1. It mainly relies on PyTorch so that calculations may be GPU-accelerated thanks to PyTorch native CUDA support [192]. It also employs to a lesser extent NumPy [231]. It was publicly released in [200].

The need for such a package was motivated by the statistical denoising algorithm introduced in Chapter V. When this work was conducted, there were, at my knowledge, only two publicly available codes in Python, which were released in [6] and [199], respectively. I made use of this last one, progressively modifying it until the development of PyWPH, which was initiated by the will to optimize and simplify the computations of WPH statistics.

PyWST provides a framework to compute and handle WPH coefficients from real or complex-valued images such as defined in Sect. IV.3.1. The calculations are implemented with the PyTorch framework, allowing GPU acceleration with CUDA, and gradient-based optimizations requiring automatic differentiation. The implementation remains flexible so that changing wavelets or choosing a custom subset of WPH or scaling moments is made easily possible. Code examples are provided, notably showing how to generate statistical syntheses as introduced in Sect. IV.3, and how to perform statistical denoisings as introduced in Sect. V.2. We point out that PyWPH is still in active development at the time of writing.

²<https://github.com/bregaldo/pywph/>

Bibliography

- [1] BICEP2 COLLABORATION; “Detection of B-Mode Polarization at Degree Angular Scales by BICEP2”; *Physical Review Letters* **112**, p. 241101 (2014). 1403.3985. vii
- [2] BICEP2/KECK ARRAY AND PLANCK COLLABORATIONS; “Joint Analysis of BICEP2/Keck Array and Planck Data”; *Physical Review Letters* **114**, p. 101301 (2015). 1502.00612. vii, 8, 12, 29, 99, 124
- [3] S. MALLAT; “Group Invariant Scattering”; *Communications on Pure and Applied Mathematics* **65**, pp. 1331–1398 (2012). viii, 47, 50, 54, 55
- [4] J. BRUNA & S. MALLAT; “Invariant scattering convolution networks”; *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, pp. 1872–1886 (2013). viii, 47, 50, 51, 52, 54, 55, 61, 149
- [5] S. MALLAT, S. ZHANG & G. ROCHELLE; “Phase harmonic correlations and convolutional neural networks”; *Information and Inference: A Journal of the IMA* **9**, pp. 721–747 (2019). 1810.12136. viii, 47, 88, 89, 90
- [6] S. ZHANG & S. MALLAT; “Maximum entropy models from phase harmonic covariances”; *Applied and Computational Harmonic Analysis* **53**, pp. 199–230 (2021). 1911.10017. viii, 47, 88, 90, 94, 95, 107, 154
- [7] V. M. SLIPHER; “Spectrographic Observations of Nebulae”; *Popular Astronomy* **23**, pp. 21–24 (1915). 2
- [8] V. M. SLIPHER; “Nebulæ”; *Proceedings of the American Philosophical Society* **56**, pp. 403–409 (1917). 2
- [9] A. EINSTEIN; “Zur Elektrodynamik bewegter Körper”; *Annalen der Physik* **322**, pp. 891–921 (1905). 2
- [10] A. EINSTEIN; “Die Grundlage der allgemeinen Relativitätstheorie”; *Annalen der Physik* **354**, pp. 769–822 (1916). 2
- [11] A. FRIEDMANN; “Über die Krümmung des Raumes”; *Zeitschrift für Physik* **10**, pp. 377–386 (1922). 2
- [12] G. LEMAÎTRE; “Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques”; *Annales de la Société Scientifique de Bruxelles* **47**, pp. 49–59 (1927). 3

- [13] PLANCK COLLABORATION I; “*Planck* 2018 results. I. Overview, and the cosmological legacy of *Planck*”; *Astronomy & Astrophysics* **641**, p. A1 (2020). 1807.06205. 3, 4, 5, 6, 7, 11, 12, 13
- [14] E. HUBBLE; “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”; *Proceedings of the National Academy of Science* **15**, pp. 168–173 (1929). 3
- [15] A. G. RIESS, A. V. FILIPPIENKO, P. CHALLIS *et al.*; “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”; *The Astronomical Journal* **116**, pp. 1009–1038 (1998). 3
- [16] S. PERLMUTTER, G. ALDERING, G. GOLDHABER *et al.*; “Measurements of Ω and Λ from 42 High-Redshift Supernovae”; *The Astrophysical Journal* **517**, pp. 565–586 (1999). 3
- [17] P. PETER & J. UZAN; *Cosmologie primordiale, 2^{ème} édition*; Collection Echelles (Belin) (2012). ISBN 9782701162447. 4, 5, 7, 8
- [18] F. BERNARDEAU; *Cosmologie : des fondements théoriques aux observations*; Savoirs actuels (EDP Sciences CNRS éd.) (2007). ISBN 9782868839541. 4, 8
- [19] D. J. FIXSEN; “The Temperature of the Cosmic Microwave Background”; *The Astrophysical Journal* **707**, pp. 916–920 (2009). 0911.1955. 5, 6
- [20] PLANCK COLLABORATION VI; “*Planck* 2018 results. VI. Cosmological parameters”; *Astronomy & Astrophysics* **641**, p. A6 (2020). 1807.06209. 5, 12
- [21] R. A. ALPHER, H. BETHE & G. GAMOW; “The Origin of Chemical Elements”; *Physical Review* **73**, pp. 803–804 (1948). 5
- [22] G. GAMOW; “The Evolution of the Universe”; *Nature* **162**, pp. 680–682 (1948). 5
- [23] R. A. ALPHER & R. HERMAN; “Evolution of the Universe”; *Nature* **162**, pp. 774–775 (1948). 5
- [24] P. NOTERDAEME, P. PETITJEAN, R. SRIANAND, C. LEDOUX & S. LÓPEZ; “The evolution of the cosmic microwave background temperature. Measurements of T_{CMB} at high redshift from carbon monoxide excitation”; *Astronomy & Astrophysics* **526**, L7 (2011). 1012.3164. 5
- [25] A. A. PENZIAS & R. W. WILSON; “A measurement of excess antenna temperature at 4080 Mc/s”; *The Astrophysical Journal* **142**, pp. 419–421 (1965). 6
- [26] R. H. DICKE, P. J. E. PEEBLES, P. G. ROLL & D. T. WILKINSON; “Cosmic Black-Body Radiation”; *The Astrophysical Journal* **142**, pp. 414–419 (1965). 6
- [27] J. R. BOND, L. KOFMAN & D. POGOSYAN; “How filaments of galaxies are woven into the cosmic web”; *Nature* **380**, pp. 603–606 (1996). astro-ph/9512141. 6

- [28] R. VAN DE WEYGAERT & J. R. BOND; “Clusters and the Theory of the Cosmic Web”; in “A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure,” , edited by M. PLIONIS, D. HUGHES & O. LÓPEZ-CRUZ; pp. 335–408 (Springer Netherlands, Dordrecht) (2008); ISBN 9781402069413. 6
- [29] PLANCK COLLABORATION IV; “*Planck* 2018 results. IV. Diffuse component separation”; *Astronomy & Astrophysics* **641**, p. A4 (2020). 1807.06208. 7, 12, 13, 15, 16, 37, 115, 120
- [30] PLANCK COLLABORATION V; “*Planck* 2018 results. V. Power spectra and likelihoods”; *Astronomy & Astrophysics* **641**, p. A5 (2020). 1907.12875. 7, 8, 12
- [31] G. F. SMOOT, C. L. BENNETT, A. KOGUT *et al.*; “Structure in the COBE Differential Microwave Radiometer First-Year Maps”; *The Astrophysical Journal Letters* **396**, p. L1 (1992). 6, 11
- [32] C. L. BENNETT, M. HALPERN, G. HINSHAW *et al.*; “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results”; *The Astrophysical Journal Supplement Series* **148**, pp. 1–27 (2003). astro-ph/0302207. 6, 11
- [33] PLANCK COLLABORATION VII; “*Planck* 2018 results. VII. Isotropy and statistics”; *Astronomy & Astrophysics* **641**, p. A7 (2020). 1906.02552. 6, 12, 33, 129
- [34] A. KOSOWSKY; “Introduction to microwave background polarization”; *New Astronomy Reviews* **43**, pp. 157–168 (1999). astro-ph/9904102. 7
- [35] M. KAMIONKOWSKI & E. D. KOVETZ; “The Quest for B Modes from Inflationary Gravitational Waves”; *Annual Review of Astronomy and Astrophysics* **54**, pp. 227–269 (2016). 1510.06042. 7, 10, 11
- [36] M. KAMIONKOWSKI, A. KOSOWSKY & A. STEBBINS; “A Probe of Primordial Gravity Waves and Vorticity”; *Physical Review Letters* **78**, pp. 2058–2061 (1997). astro-ph/9609132. 8, 10
- [37] M. ZALDARRIAGA & U. SELJAK; “All-sky analysis of polarization in the microwave background”; *Physical Review D* **55**, pp. 1830–1840 (1997). astro-ph/9609170. 8, 10
- [38] U. SELJAK & M. ZALDARRIAGA; “Signature of Gravity Waves in the Polarization of the Microwave Background”; *Physical Review Letters* **78**, pp. 2054–2057 (1997). 8
- [39] PLANCK COLLABORATION X; “*Planck* 2018 results. X. Constraints on inflation”; *Astronomy & Astrophysics* **641**, p. A10 (2020). 1807.06211. 8, 12
- [40] KECK ARRAY AND BICEP2 COLLABORATIONS; “Constraints on Primordial Gravitational Waves Using *Planck*, WMAP, and New BICEP2/*Keck* Observations through the 2015 Season”; *Physical Review Letters* **121**, p. 221301 (2018). 1810.05216. 8

- [41] S. NAESS, M. HASSELFIELD, J. MCMAHON *et al.*; “The Atacama Cosmology Telescope: CMB polarization at $200 < l < 9000$ ”; *Journal of Cosmology and Astroparticle Physics* **2014**, 007 (2014). [1405.5524](#). 8
- [42] A. A. FRAISSE, P. A. R. ADE, M. AMIRI *et al.*; “SPIDER: probing the early Universe with a suborbital polarimeter”; *Journal of Cosmology and Astroparticle Physics* **2013**, 047 (2013). [1106.3087](#). 8
- [43] H. ISHINO, Y. AKIBA, K. ARNOLD *et al.*; “LiteBIRD: lite satellite for the study of B-mode polarization and inflation from cosmic microwave background radiation detection”; in “Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave,” , volume 9904, edited by H. A. MACEWEN, G. G. FAZIO *et al.*; pp. 354 – 361; International Society for Optics and Photonics (SPIE) (2016). 8
- [44] B. M. SUTIN, M. ALVAREZ, N. BATTAGLIA *et al.*; “PICO - the probe of inflation and cosmic origins”; in “Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave,” , volume 10698, edited by M. LYSTRUP, H. A. MACEWEN *et al.*; pp. 1323 – 1333; International Society for Optics and Photonics (SPIE) (2018). [1808.01368](#). 8
- [45] SIMONS OBSERVATORY COLLABORATION; “The Simons Observatory: science goals and forecasts”; *Journal of Cosmology and Astroparticle Physics* **2019**, 056 (2019). [1808.07445](#). 8
- [46] K. ABAZAJIAN, G. ADDISON, P. ADSHEAD *et al.*; “CMB-S4 Science Case, Reference Design, and Project Plan”; (2019). [1907.04473](#). 8
- [47] PLANCK COLLABORATION XXII; “*Planck* 2013 results. XXII. Constraints on inflation”; *Astronomy & Astrophysics* **571**, p. A22 (2014). [1303.5082](#). 8
- [48] A. LINDE; “Inflationary Cosmology after Planck”; in “Post-Planck Cosmology,” , *Lecture notes of the Les Houches Summer School*, volume 100, edited by L. F. CUGLIANDOLO, C. DEFFAYET *et al.*; chapter 6, pp. 231–316 (Oxford University Press) (2015); ISBN 9780198728856. 8
- [49] M. ZALDARRIAGA; “Nature of the $E - B$ decomposition of CMB polarization”; *Physical Review D* **64**, p. 103001 (2001). [astro-ph/0106174](#). 10
- [50] D. L. CLEMENTS; “An introduction to the *Planck* mission”; *Contemporary Physics* **58**, pp. 331–348 (2017). [1707.09220](#). 11
- [51] C. L. BENNETT, A. J. BANDY, K. M. GORSKI *et al.*; “Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results”; *The Astrophysical Journal Letters* **464**, p. L1 (1996). [astro-ph/9601067](#). 11
- [52] C. L. BENNETT, D. LARSON, J. L. WEILAND *et al.*; “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results”; *The Astrophysical Journal Supplement Series* **208**, 20 (2013). [1212.5225](#). 11
- [53] PLANCK COLLABORATION; “The Scientific Programme of Planck”; ESA publication ESA-SCI(2005)/01 (2005). [astro-ph/0604069](#). 11

- [54] PLANCK COLLABORATION II; “*Planck* 2018 results. II. Low Frequency Instrument data processing”; *Astronomy & Astrophysics* **641**, p. A2 (2020). 1807.06206. 11
- [55] PLANCK COLLABORATION III; “*Planck* 2018 results. III. High Frequency Instrument data processing”; *Astronomy & Astrophysics* **641**, p. A3 (2020). 1807.06207. 11, 99, 111, 115, 122
- [56] PLANCK COLLABORATION VIII; “*Planck* 2018 results. VIII. Gravitational lensing”; *Astronomy & Astrophysics* **641**, p. A8 (2020). 1807.06210. 12
- [57] PLANCK COLLABORATION XI; “*Planck* 2018 results. XI. Polarized dust foregrounds”; *Astronomy & Astrophysics* **641**, p. A11 (2020). 1801.04945. 12, 23, 24, 37, 75
- [58] PLANCK COLLABORATION XII; “*Planck* 2018 results. XII. Galactic astrophysics using polarized dust emission”; *Astronomy & Astrophysics* **641**, p. A12 (2020). 1807.06212. 12, 21, 23, 70, 120
- [59] PLANCK COLLABORATION IX; “*Planck* 2018 results. IX. Constraints on primordial non-Gaussianity”; *Astronomy & Astrophysics* **641**, p. A9 (2020). 1905.05697. 12, 31
- [60] PLANCK COLLABORATION XII; “*Planck* 2013 results. XII. Diffuse component separation”; *Astronomy & Astrophysics* **571**, p. A12 (2014). 1303.5072. 13
- [61] PLANCK COLLABORATION X; “*Planck* 2015 results. X. Diffuse component separation: Foreground maps”; *Astronomy & Astrophysics* **594**, p. A10 (2016). 1502.01588. 13, 14, 37
- [62] J. DELABROUILLE & J.-F. CARDOSO; “Diffuse Source Separation in CMB Observations”; in “*Data Analysis in Cosmology*,” , edited by V. J. MARTINEZ, E. SAAR, E. M. GONZALES & M. J. PONS-BORDERIA; pp. 159–205 (Springer Berlin Heidelberg, Berlin, Heidelberg) (2009); ISBN 9783540447672. 13
- [63] H. K. ERIKSEN, I. J. O’DWYER, J. B. JEWELL *et al.*; “Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling”; *The Astrophysical Journal Supplement Series* **155**, pp. 227–241 (2004). astro-ph/0407028. 14
- [64] H. K. ERIKSEN, J. B. JEWELL, C. DICKINSON *et al.*; “Joint Bayesian Component Separation and CMB Power Spectrum Estimation”; *The Astrophysical Journal* **676**, pp. 10–32 (2008). 0709.1058. 14
- [65] A. GELMAN, J. B. CARLIN, H. S. STERN *et al.*; *Bayesian Data Analysis*; 3rd edition (Chapman and Hall/CRC) (2013). ISBN 9781439840955. <http://www.stat.columbia.edu/~gelman/book/>. 15
- [66] M. REMAZEILLES, J. DELABROUILLE & J.-F. CARDOSO; “Foreground component separation with generalized Internal Linear Combination”; *Monthly Notices of the Royal Astronomical Society* **418**, pp. 467–476 (2011). 1103.1166. 15

- [67] PLANCK COLLABORATION INT. XLVIII; “*Planck* intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies”; *Astronomy & Astrophysics* **596**, p. A109 (2016). 1605.09387. 15, 23, 37, 117
- [68] F. J. NARCOWICH, P. PETRUSHEV & J. D. WARD; “Localized Tight Frames on Spheres”; *SIAM Journal on Mathematical Analysis* **38**, pp. 574–594 (2006). 16, 127
- [69] F. GUILLOUX, G. FAË & J.-F. CARDOSO; “Practical wavelet design on the sphere”; *Applied and Computational Harmonic Analysis* **26**, pp. 143–160 (2009). 0706.2598. 16, 127
- [70] J.-F. CARDOSO, M. LE JEUNE, J. DELABROUILLE *et al.*; “Component Separation With Flexible Models—Application to Multichannel Astrophysical Observations”; *IEEE Journal of Selected Topics in Signal Processing* **2**, pp. 735–746 (2008). 0803.1814. 16
- [71] M. A. PETROFF, G. E. ADDISON, C. L. BENNETT & J. L. WEILAND; “Full-sky Cosmic Microwave Background Foreground Cleaning Using Machine Learning”; *The Astrophysical Journal* **903**, 104 (2020). 2004.11507. 17
- [72] S. WAGNER-CARENA, M. HOPKINS, A. DIAZ RIVERO & C. DVORKIN; “A novel CMB component separation method: hierarchical generalized morphological component analysis”; *Monthly Notices of the Royal Astronomical Society* **494**, pp. 1507–1529 (2020). 1910.08077. 17
- [73] K. AYLOR, M. HAQ, L. KNOX *et al.*; “Cleaning our own dust: simulating and separating galactic dust foregrounds with neural networks”; *Monthly Notices of the Royal Astronomical Society* **500**, pp. 3889–3897 (2021). 1909.06467. 17, 33, 44, 45
- [74] M. REMAZEILLES, A. ROTTI & J. CHLUBA; “Peeling off foregrounds with the constrained moment ILC method to unveil primordial CMB B modes”; *Monthly Notices of the Royal Astronomical Society* **503**, pp. 2478–2498 (2021). 2006.08628. 17
- [75] B. T. DRAINE; *Physics of the interstellar and intergalactic medium* (Princeton University Press) (2011). ISBN 9780691122144. 17, 18, 19, 57
- [76] K. FERRIÈRE; “Interstellar magnetic fields: from Galactic scales to the edge of the heliosphere”; *Journal of Physics: Conference Series* **577**, p. 012008 (2015). 18
- [77] M. HAVERKORN; “Magnetic Fields in the Milky Way”; in “Magnetic Fields in Diffuse Media,” , edited by A. LAZARIAN, E. M. DE GOUVEIA DAL PINO & C. MELOLI; pp. 483–506 (Springer Berlin Heidelberg, Berlin, Heidelberg) (2015); ISBN 9783662446256; 1406.0283. 18
- [78] E. FERMI; “On the Origin of the Cosmic Radiation”; *Physical Review* **75**, pp. 1169–1174 (1949). 18
- [79] J. S. HALL; “Observations of the Polarized Light from Stars”; *Science* **109**, pp. 166–167 (1949). 18, 20

- [80] W. A. HILTNER; “Polarization of Light from Distant Stars by Interstellar Medium”; *Science* **109**, p. 165 (1949). 18, 20
- [81] M. HAVERKORN, J. C. BROWN, B. M. GAENSLER & N. M. MCCLURE-GRIFFITHS; “The Outer Scale of Turbulence in the Magnetoionized Galactic Interstellar Medium”; *The Astrophysical Journal* **680**, pp. 362–370 (2008). 0802.2740. 18
- [82] P. HENNEBELLE & E. FALGARONE; “Turbulent molecular clouds”; *The Astronomy and Astrophysics Review* **20**, 55 (2012). 1211.0637. 19
- [83] L. F. RICHARDSON & P. LYNCH; *Weather Prediction by Numerical Process*; Cambridge Mathematical Library; 2nd edition (Cambridge University Press) (2007). 19
- [84] U. FRISCH; *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press) (1995). 19, 57
- [85] E. E. BARNARD; “On a nebulous groundwork in the constellation Taurus.” *The Astrophysical Journal* **25**, pp. 218–225 (1907). 19
- [86] E. E. BARNARD; “On a great nebulous region and the question of absorbing matter in space and the transparency of the nebulae.” *The Astrophysical Journal* **31**, pp. 8–14 (1910). 19
- [87] W. STEIN; “Infrared Radiation from Interstellar Grains”; *The Astrophysical Journal* **144**, p. 318 (1966). 19, 20
- [88] V. WAKELAM, E. BRON, S. CAZAUX *et al.*; “H₂ formation on interstellar dust grains: The viewpoints of theory, experiments, models and observations”; *Molecular Astrophysics* **9**, pp. 1–36 (2017). 1711.10568. 19
- [89] B. T. DRAINE & A. A. FRAISSE; “Polarized Far-Infrared and Submillimeter Emission from Interstellar Dust”; *The Astrophysical Journal* **696**, pp. 1–11 (2009). 0809.2094. 20
- [90] L. DAVIS & J. L. GREENSTEIN; “The Polarization of Starlight by Interstellar Dust Particles in a Galactic Magnetic Field”; *Physical Review* **75**, pp. 1605–1605 (1949). 20
- [91] J. DAVIS, Leverett & J. L. GREENSTEIN; “The Polarization of Starlight by Aligned Dust Grains.” *The Astrophysical Journal* **114**, p. 206 (1951). 20
- [92] B. G. ANDERSSON, A. LAZARIAN & J. E. VAILLANCOURT; “Interstellar Dust Grain Alignment”; *Annual Review of Astronomy and Astrophysics* **53**, pp. 501–539 (2015). 20
- [93] S. J. BARNETT; “Magnetization by Rotation”; *Physical Review* **6**, pp. 239–270 (1915). 20
- [94] S. RESSL, V. GUILLET, R. BRAUER *et al.*; “A systematic study of radiative torque grain alignment in the diffuse interstellar medium”; *Astronomy & Astrophysics* **640**, A118 (2020). 2002.11792. 20, 25, 39, 41

- [95] PLANCK COLLABORATION I; “*Planck* 2015 results. I. Overview of products and results”; *Astronomy & Astrophysics* **594**, p. A1 (2016). 1502.01582. 22
- [96] PLANCK COLLABORATION INT. XIX; “*Planck* intermediate results. XIX. An overview of the polarized thermal emission from Galactic dust”; *Astronomy & Astrophysics* **576**, p. A104 (2015). 1405.0871. 22, 23
- [97] S. PLASZCZYNSKI, L. MONTIER, F. LEVRIER & M. TRISTRAM; “A novel estimator of the polarization amplitude from normally distributed Stokes parameters”; *Monthly Notices of the Royal Astronomical Society* **439**, pp. 4048–4056 (2014). 1312.0437. 22
- [98] L. MONTIER, S. PLASZCZYNSKI, F. LEVRIER *et al.*; “Polarization measurement analysis. I. Impact of the full covariance matrix on polarization fraction and angle measurements”; *Astronomy & Astrophysics* **574**, A135 (2015). 1406.6536. 22
- [99] L. MONTIER, S. PLASZCZYNSKI, F. LEVRIER *et al.*; “Polarization measurement analysis. II. Best estimators of polarization fraction and angle”; *Astronomy & Astrophysics* **574**, A136 (2015). 1407.0178. 22
- [100] B. S. HENSLEY & P. BULL; “Mitigating Complex Dust Foregrounds in Future Cosmic Microwave Background Polarization Experiments”; *The Astrophysical Journal* **853**, 127 (2018). 1709.07897. 23, 37
- [101] A. MANGILLI, J. AUMONT, A. ROTTI *et al.*; “Dust moments: towards a new modeling of the galactic dust emission for CMB B-modes analysis”; *Astronomy & Astrophysics* **647**, A52 (2021). 1912.09567. 23, 37
- [102] W. R. COULTON & D. N. SPERGEL; “The bispectrum of polarized galactic foregrounds”; *Journal of Cosmology and Astroparticle Physics* **2019**, 056 (2019). 1901.04515. 23
- [103] PLANCK COLLABORATION INT. XX; “*Planck* intermediate results. XX. Comparison of polarized thermal emission from Galactic dust with simulations of MHD turbulence”; *Astronomy & Astrophysics* **576**, p. A105 (2015). 1405.0872. 23, 26, 39, 57, 70
- [104] PLANCK COLLABORATION INT. XXXII; “*Planck* intermediate results. XXXII. The relative orientation between the magnetic field and structures traced by interstellar dust”; *Astronomy & Astrophysics* **586**, p. A135 (2016). 1409.6728. 23
- [105] PLANCK COLLABORATION INT. XXXV; “*Planck* intermediate results. XXXV. Probing the role of the magnetic field in the formation of structure in molecular clouds”; *Astronomy & Astrophysics* **586**, p. A138 (2016). 1502.04123. 23
- [106] S. E. CLARK, J. E. G. PEEK & M. E. PUTMAN; “Magnetically Aligned H I Fibers and the Rolling Hough Transform”; *The Astrophysical Journal* **789**, 82 (2014). 1312.1338. 23

- [107] S. E. CLARK, J. C. HILL, J. E. G. PEEK *et al.*; “Neutral Hydrogen Structures Trace Dust Polarization Angle: Implications for Cosmic Microwave Background Foregrounds”; *Physical Review Letters* **115**, 241302 (2015). 1508.07005. 23, 24
- [108] PLANCK COLLABORATION INT. XXX; “*Planck* intermediate results. XXX. The angular power spectrum of polarized dust emission at intermediate and high Galactic latitudes”; *Astronomy & Astrophysics* **586**, p. A133 (2016). 1409.5738. 23
- [109] PLANCK COLLABORATION INT. XXXVIII; “*Planck* intermediate results. XXXVIII. *E*- and *B*-modes of dust polarization from the magnetized filamentary structure of the interstellar medium”; *Astronomy & Astrophysics* **586**, p. A141 (2016). 1505.02779. 24
- [110] S. E. CLARK, C.-G. KIM, J. C. HILL & B. S. HENSLEY; “The Origin of Parity Violation in Polarized Dust Emission and Implications for Cosmic Birefringence”; arXiv e-prints arXiv:2105.00120 (2021). 2105.00120. 24
- [111] S. GALTIER; *Introduction to Modern Magnetohydrodynamics* (Cambridge University Press) (2016). ISBN 9781316665961. 25
- [112] R. TEYSSIER; “Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES”; *Astronomy & Astrophysics* **385**, pp. 337–364 (2002). astro-ph/0111367. 25, 57
- [113] S. FROMANG, P. HENNEBELLE & R. TEYSSIER; “A high order Godunov scheme with constrained transport and adaptive mesh refinement for astrophysical magnetohydrodynamics”; *Astronomy & Astrophysics* **457**, pp. 371–384 (2006). astro-ph/0607230. 25, 57
- [114] J. M. STONE, T. A. GARDINER, P. TEUBEN *et al.*; “Athena: A New Code for Astrophysical MHD”; *The Astrophysical Journal Supplement Series* **178**, pp. 137–177 (2008). 0804.0402. 25
- [115] C.-G. KIM & E. C. OSTRIKER; “Three-phase Interstellar Medium in Galaxies Resolving Evolution with Star Formation and Supernova Feedback (TIGRESS): Algorithms, Fiducial Model, and Convergence”; *The Astrophysical Journal* **846**, 133 (2017). 1612.03918. 25
- [116] G. MOMFERRATOS, P. LESAFFRE, E. FALGARONE & G. PINEAU DES FORÊTS; “Turbulent energy dissipation and intermittency in ambipolar diffusion magnetohydrodynamics”; *Monthly Notices of the Royal Astronomical Society* **443**, pp. 86–101 (2014). 1502.03624. 25
- [117] E. BELLOMI, B. GODARD, P. HENNEBELLE *et al.*; “3D chemical structure of diffuse turbulent ISM. I. Statistics of the HI-to-H₂ transition”; *Astronomy & Astrophysics* **643**, A36 (2020). 2009.05466. 25, 57
- [118] T. RICHARD, P. LESAFFRE, E. FALGARONE & A. LEHMANN; “Probing the nature of dissipation in compressible MHD turbulence”; (submitted). 25

- [119] W. SCHMIDT, W. HILLEBRANDT & J. C. NIEMEYER; “Numerical dissipation and the bottleneck effect in simulations of compressible isotropic turbulence”; *Computers & Fluids* **35**, pp. 353–371 (2006). [astro-ph/0407616](#). 25, 57
- [120] S. RESSL, S. WOLF & R. BRAUER; “Radiative transfer with POLARIS. I. Analysis of magnetic fields through synthetic dust continuum polarization measurements”; *Astronomy & Astrophysics* **593**, A87 (2016). [1604.05305](#). 26
- [121] K. M. GÓRSKI, E. HIVON, A. J. BANDY *et al.*; “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere”; *The Astrophysical Journal* **622**, pp. 759–771 (2005). [astro-ph/0409513](#). 26, 115
- [122] B. BURKHART; “Diagnosing Turbulence in the Neutral and Molecular Interstellar Medium of Galaxies”; (2021). [2106.02239](#). 27
- [123] J. DELABROUILLE, M. BETOULE, J. B. MELIN *et al.*; “The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths”; *Astronomy & Astrophysics* **553**, A96 (2013). [1207.3675](#). 26, 115
- [124] B. THORNE, J. DUNKLEY, D. ALONSO & S. NÆSS; “The Python Sky Model: software for simulating the Galactic microwave sky”; *Monthly Notices of the Royal Astronomical Society* **469**, pp. 2821–2833 (2017). [1608.02841](#). 30
- [125] M. A. MIVILLE-DESCHÊNES, P. G. MARTIN, A. ABERGEL *et al.*; “Herschel-SPIRE observations of the Polaris flare: Structure of the diffuse interstellar medium at the sub-parsec scale”; *Astronomy & Astrophysics* **518**, L104 (2010). [1005.2746](#). 33, 51
- [126] B. BURKHART, D. FALCETA-GONÇALVES, G. KOWAL & A. LAZARIAN; “Density Studies of MHD Interstellar Turbulence: Statistical Moments, Correlations and Bispectrum”; *The Astrophysical Journal* **693**, pp. 250–266 (2009). [0811.0822](#). 33
- [127] S. K. N. PORTILLO, Z. SLEPIAN, B. BURKHART *et al.*; “Developing the 3-point Correlation Function for the Turbulent Interstellar Medium”; *The Astrophysical Journal* **862**, 119 (2018). [1711.09907](#). 33, 34
- [128] P. HILY-BLANT, E. FALGARONE & J. PETY; “Dissipative structures of diffuse molecular gas. III. Small-scale intermittency of intense velocity-shears”; *Astronomy & Astrophysics* **481**, pp. 367–380 (2008). [0802.0758](#). 33
- [129] P. HILY-BLANT & E. FALGARONE; “Intermittency of interstellar turbulence: parsec-scale coherent structure of intense, velocity shear”; *Astronomy & Astrophysics* **500**, pp. L29–L32 (2009). [0905.0368](#). 33
- [130] PLANCK COLLABORATION INT. XXIII; “*Planck* intermediate results. XXIII. Galactic plane emission components derived from *Planck* with ancillary data”; *Astronomy & Astrophysics* **580**, p. A13 (2015). [1406.5093](#). 33
- [131] S. CODIS, C. PICHON, D. POGOSYAN *et al.*; “Non-Gaussian Minkowski functionals and extrema counts in redshift space”; *Monthly Notices of the Royal Astronomical Society* **435**, pp. 531–564 (2013). [1305.7402](#). 33

- [132] C. PARRONI, V. F. CARDONE, R. MAOLI & R. SCARAMELLA; “Going deep with Minkowski functionals of convergence maps”; *Astronomy & Astrophysics* **633**, A71 (2020). 1911.06243. 33, 35
- [133] N. KRACHMALNICOFF & G. PUGLISI; “ForSE: A GAN-based Algorithm for Extending CMB Foreground Models to Subdegree Angular Scales”; *The Astrophysical Journal* **911**, 42 (2021). 2011.02221. 33, 44
- [134] I. MAKARENKO, P. BUSHBY, A. FLETCHER *et al.*; “Topological data analysis and diagnostics of compressible magnetohydrodynamic turbulence”; *Journal of Plasma Physics* **84**, 735840403 (2018). 1804.04688. 35
- [135] P. PRANAV, R. VAN DE WEYGAERT, G. VEGTER *et al.*; “Topology and geometry of Gaussian random fields I: on Betti numbers, Euler characteristic, and Minkowski functionals”; *Monthly Notices of the Royal Astronomical Society* **485**, pp. 4167–4208 (2019). 1812.07310. 35
- [136] T. SOUSBIE; “The persistent cosmic web and its filamentary structure - I. Theory and implementation”; *Monthly Notices of the Royal Astronomical Society* **414**, pp. 350–383 (2011). 1009.4015. 35
- [137] T. SOUSBIE, C. PICHON & H. KAWAHARA; “The persistent cosmic web and its filamentary structure - II. Illustrations”; *Monthly Notices of the Royal Astronomical Society* **414**, pp. 384–403 (2011). 1009.4014. 35
- [138] PLANCK COLLABORATION XI; “*Planck* 2013 results. XI. All-sky model of thermal dust emission”; *Astronomy & Astrophysics* **571**, p. A11 (2014). 1312.1300. 36, 37, 38
- [139] Y.-K. CHIANG & B. MÉNARD; “Extragalactic Imprints in Galactic Dust Maps”; *The Astrophysical Journal* **870**, 120 (2019). 1808.03294. 37
- [140] D. P. FINKBEINER, M. DAVIS & D. J. SCHLEGEL; “Extrapolation of Galactic Dust Emission at 100 Microns to Cosmic Microwave Background Radiation Frequencies Using FIRAS”; *The Astrophysical Journal* **524**, pp. 867–886 (1999). astro-ph/9905128. 37
- [141] A. M. MEISNER & D. P. FINKBEINER; “Modeling Thermal Dust Emission with Two Components: Application to the Planck High Frequency Instrument Maps”; *The Astrophysical Journal* **798**, 88 (2015). 1410.7523. 37
- [142] J. CHLUBA, J. C. HILL & M. H. ABITBOL; “Rethinking CMB foregrounds: systematic extension of foreground parametrizations”; *Monthly Notices of the Royal Astronomical Society* **472**, pp. 1195–1213 (2017). 1701.00274. 37
- [143] PLANCK COLLABORATION INT. XXIX; “*Planck* intermediate results. XXIX. All-sky dust modelling with *Planck*, IRAS, and WISE observations”; *Astronomy & Astrophysics* **586**, p. A132 (2016). 1409.2495. 37

- [144] B. T. DRAINE & A. LI; “Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era”; *The Astrophysical Journal* **657**, pp. 810–837 (2007). [astro-ph/0608003](#). 39
- [145] V. GUILLET, L. FANCIULLO, L. VERSTRAETE, F. BOULANGER, A. P. JONES, M. A. MIVILLE-DESCHÈNES, N. YSARD, F. LEVRIER & M. ALVES; “Dust models compatible with Planck intensity and polarization data in translucent lines of sight”; *Astronomy & Astrophysics* **610**, A16 (2018). [1710.04598](#). 39
- [146] PLANCK COLLABORATION INT. XLIV; “*Planck* intermediate results. XLIV. The structure of the Galactic magnetic field from dust polarization maps of the southern Galactic cap”; *Astronomy & Astrophysics* **596**, p. A105 (2016). [1604.01029](#). 40
- [147] F. LEVRIER; “The dynamical and magnetized interstellar medium”; Mémoire d’habilitation à diriger des recherches (2016). 40
- [148] F. VANSYNGEL, F. BOULANGER, T. GHOSH *et al.*; “Statistical simulations of the dust foreground to cosmic microwave background polarization”; *Astronomy & Astrophysics* **603**, A62 (2017). [1611.02577](#). 40, 46
- [149] T. GHOSH, F. BOULANGER, P. G. MARTIN *et al.*; “Modelling and simulation of large-scale polarized dust emission over the southern Galactic cap using the GASS HI data”; *Astronomy & Astrophysics* **601**, A71 (2017). [1611.02418](#). 40, 41
- [150] D. ADAK, T. GHOSH, F. BOULANGER *et al.*; “Dust polarization modelling at large scale over the northern Galactic cap using EBHIS and Planck data”; *Astronomy & Astrophysics* **640**, A100 (2020). [1906.07445](#). 40, 41
- [151] S. E. CLARK; “A New Probe of Line-of-sight Magnetic Field Tangling”; *The Astrophysical Journal Letters* **857**, L10 (2018). [1802.00011](#). 41
- [152] S. E. CLARK & B. S. HENSLEY; “Mapping the Magnetic Interstellar Medium in Three Dimensions over the Full Sky with Neutral Hydrogen”; *The Astrophysical Journal* **887**, 136 (2019). [1909.11673](#). 41
- [153] F. LEVRIER, J. NEVEU, E. FALGARONE *et al.*; “Statistics of the polarized sub-millimetre emission maps from thermal dust in the turbulent, magnetized, diffuse ISM”; *Astronomy & Astrophysics* **614**, A124 (2018). [1802.08725](#). 41, 46
- [154] A. G. KRITSUK, R. FLAUGER & S. D. USTYUGOV; “Dust-Polarization Maps for Local Interstellar Turbulence”; *Physical Review Letters* **121**, 021104 (2018). [1711.11108](#). 41, 46
- [155] C.-G. KIM, S. K. CHOI & R. FLAUGER; “Dust Polarization Maps from TIGRESS: E/B Power Asymmetry and TE Correlation”; *The Astrophysical Journal* **880**, 106 (2019). [1901.07079](#). 41, 46
- [156] Y. LECUN, L. BOTTOU, Y. BENGIO & P. HAFFNER; “Gradient-based learning applied to document recognition”; *Proceedings of the IEEE* **86**, pp. 2278–2324 (1998). 42, 43

- [157] Y. LECUN, Y. BENGIO & G. HINTON; “Deep Learning”; *Nature* **521**, pp. 436–444 (2015). 42
- [158] Y. LECUN, B. BOSER, J. S. DENKER *et al.*; “Backpropagation Applied to Handwritten Zip Code Recognition”; *Neural Computation* **1**, p. 541–551 (1989). 42
- [159] I. GOODFELLOW, Y. BENGIO & A. COURVILLE; *Deep Learning* (The MIT Press) (2016). <http://www.deeplearningbook.org>. 42, 43
- [160] I. J. GOODFELLOW, J. POUGET-ABADIE, M. MIRZA *et al.*; “Generative Adversarial Networks”; (2014). [1406.2661](https://arxiv.org/abs/1406.2661). 44
- [161] D. P. KINGMA & M. WELLING; “Auto-Encoding Variational Bayes”; (2014). [1312.6114](https://arxiv.org/abs/1312.6114). 44
- [162] B. THORNE, L. KNOX & K. PRABHU; “A generative model of galactic dust emission using variational autoencoders”; *Monthly Notices of the Royal Astronomical Society* **504**, pp. 2603–2613 (2021). [2101.11181](https://arxiv.org/abs/2101.11181). 44, 45, 46
- [163] A. RADFORD, L. METZ & S. CHINTALA; “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”; (2016). [1511.06434](https://arxiv.org/abs/1511.06434). 44
- [164] J.-B. DURRIVE, P. LESAFFRE & K. FERRIÈRE; “Magnetic fields from multiplicative chaos”; *Monthly Notices of the Royal Astronomical Society* **496**, pp. 3015–3034 (2020). [2005.11972](https://arxiv.org/abs/2005.11972). 46
- [165] J. B. DURRIVE, P. LESAFFRE, T. GHOSH & B. REGALDO-SAINT BLANCARD; “A method to statistically characterize turbulent data with physically motivated parameters, illustrated on a centroid velocity map”; (submitted). [2101.07205](https://arxiv.org/abs/2101.07205). 46
- [166] S. MALLAT; “Understanding deep convolutional networks”; *Philosophical Transactions of the Royal Society of London Series A* **374**, p. 20150203 (2016). [1601.04920](https://arxiv.org/abs/1601.04920). 47, 50, 88
- [167] E. ALLYS, F. LEVRIER, S. ZHANG *et al.*; “The RWST, a comprehensive statistical description of the non-Gaussian structures in the ISM”; *Astronomy & Astrophysics* **629**, A115 (2019). [1905.01372](https://arxiv.org/abs/1905.01372). 49, 51, 54, 63, 65, 66, 70, 71, 74, 82
- [168] B. REGALDO-SAINT BLANCARD, F. LEVRIER, E. ALLYS *et al.*; “Statistical description of dust polarized emission from the diffuse interstellar medium - A RWST approach”; *Astronomy & Astrophysics* **642**, p. A217 (2020). [2007.08242](https://arxiv.org/abs/2007.08242). 49, 51, 82, 83, 153
- [169] S. MALLAT; “Recursive interferometric representations”; in “2010 18th European Signal Processing Conference,” pp. 716–720 (2010). 50
- [170] A. KRIZHEVSKY, I. SUTSKEVER & G. E. HINTON; “ImageNet Classification with Deep Convolutional Neural Networks”; in “*Advances in Neural Information Processing Systems*,” , volume 25, edited by F. PEREIRA, C. J. C. BURGES *et al.* (Curran Associates, Inc.) (2012). 50

- [171] J. ANDÉN & S. MALLAT; “Deep Scattering Spectrum”; *IEEE Transactions on Signal Processing* **62**, pp. 4114–4128 (2014). 1304.6763. 50
- [172] L. SIFRE & S. MALLAT; “Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination”; in “2013 IEEE Conference on Computer Vision and Pattern Recognition,” pp. 1233–1240 (2013). 50, 55
- [173] M. EICKENBERG, G. EXARCHAKIS, M. HIRN & S. MALLAT; “Solid Harmonic Wavelet Scattering: Predicting Quantum Molecular Energy from Invariant Descriptors of 3D Electronic Densities”; in “Advances in Neural Information Processing Systems,” , volume 30, edited by I. GUYON, U. V. LUXBURG *et al.* (Curran Associates, Inc.) (2017). 50
- [174] J. BRUNA, S. MALLAT, E. BACRY & J.-F. MUZY; “Intermittent process analysis with scattering moments”; *The Annals of Statistics* **43**, pp. 323–351 (2015). 1311.4104. 50, 54, 71
- [175] S. CHENG, Y.-S. TING, B. MÉNARD & J. BRUNA; “A new approach to observational cosmology using the scattering transform”; *Monthly Notices of the Royal Astronomical Society* **499**, pp. 5902–5914 (2020). 2006.08561. 51
- [176] S. CHENG & B. MÉNARD; “Weak lensing scattering transform: dark energy and neutrino mass sensitivity”; *Monthly Notices of the Royal Astronomical Society* **507**, pp. 1012–1020 (2021). 2103.09247. 51
- [177] A. K. SAYDJARI, S. K. N. PORTILLO, Z. SLEPIAN *et al.*; “Classification of Magnetohydrodynamic Simulations Using Wavelet Scattering Transforms”; *The Astrophysical Journal* **910**, 122 (2021). 2010.11963. 51
- [178] A. KHALIL, G. JONCAS, F. NEKKA *et al.*; “Morphological Analysis of H I Features. II. Wavelet-based Multifractal Formalism”; *The Astrophysical Journal Supplement Series* **165**, pp. 512–550 (2006). 51
- [179] J. F. ROBITAILLE, G. JONCAS & M. A. MIVILLE-DESCHÈNES; “Multiscale analysis of Galactic dust emission using complex wavelet transforms - I. Separation of Gaussian and non-Gaussian fluctuations in Herschel observations”; *Monthly Notices of the Royal Astronomical Society* **440**, pp. 2726–2741 (2014). 51
- [180] M. FARGE; “Wavelet Transforms And Their Applications To Turbulence”; *Annual Review of Fluid Mechanics* **24**, pp. 395–458 (1992). 1508.05650. 52
- [181] I. WALDSPURGER, A. d’ASPREMONT & S. MALLAT; “Phase Recovery, MaxCut and Complex Semidefinite Programming”; (2013). 1206.0102. 54
- [182] O. IFFRIG & P. HENNEBELLE; “Structure distribution and turbulence in self-consistently supernova-driven ISM of multiphase magnetized galactic discs”; *Astronomy & Astrophysics* **604**, A70 (2017). 1703.10421. 57

- [183] C. FEDERRATH, J. ROMAN-DUVAL, R. S. KLESSEN, W. SCHMIDT & M. M. MAC LOW; “Comparing the statistics of interstellar turbulence in simulations and observations. Solenoidal versus compressive turbulence forcing”; *Astronomy & Astrophysics* **512**, A81 (2010). 0905.1060. 57
- [184] M. HAVERKORN, P. KATGERT & A. G. DE BRUYN; “Structure in the polarized Galactic synchrotron emission, in particular “depolarization canals””; *Astronomy & Astrophysics* **427**, pp. 549–559 (2004). astro-ph/0408115. 58
- [185] A. OPPENHEIM & J. LIM; “The importance of phase in signals”; *Proceedings of the IEEE* **69**, pp. 529–541 (1981). 59
- [186] F. LEVRIER, E. FALGARONE & F. VIALLEFOND; “Fourier phase analysis in radio-interferometry”; *Astronomy & Astrophysics* **456**, pp. 205–214 (2006). astro-ph/0606222. 59
- [187] B. D. WANDEL; “Gaussian Random Fields in Cosmostatistics”; in “Astrostatistical Challenges for the New Astronomy,” , edited by J. M. HILBE; pp. 87–105 (Springer New York) (2013); ISBN 9781461435082. 59, 143
- [188] L. M. FISSEL, P. A. R. ADE, F. E. ANGILÈ *et al.*; “Balloon-Borne Submillimeter Polarimetry of the Vela C Molecular Cloud: Systematic Dependence of Polarization Fraction on Column Density and Local Polarization-Angle Dispersion”; *The Astrophysical Journal* **824**, p. 134 (2016). 1509.05298. 70
- [189] P. HENNEBELLE & S.-i. INUTSUKA; “The role of magnetic field in molecular cloud formation and evolution”; *Frontiers in Astronomy and Space Sciences* **6**, 5 (2019). 1902.00798. 71
- [190] J. BRUNA & S. MALLAT; “Multiscale sparse microcanonical models”; *Mathematical Statistics and Learning* **1**, pp. 257–315 (2019). 1801.02013. 77, 78, 79, 98
- [191] B. JULESZ; “Textons, the elements of texture perception, and their interactions”; *Nature* **290**, pp. 91–97 (1981). 77
- [192] A. PASZKE, S. GROSS, F. MASSA *et al.*; “PyTorch: An Imperative Style, High-Performance Deep Learning Library”; in “Advances in Neural Information Processing Systems,” , volume 32, edited by H. WALLACH, H. LAROCHELLE *et al.* (Curran Associates, Inc.) (2019). 78, 81, 82, 153, 154
- [193] D. LEVIN & Y. PERES; *Markov Chains and Mixing Times*; 2nd edition (American Mathematical Society) (2017). ISBN 9781470429621. 79
- [194] R. FLETCHER; *Practical Methods of Optimization*; 2nd edition (John Wiley & Sons) (1987). ISBN 9781118723203. 80
- [195] R. BYRD, P. LU, J. NOCEDAL & C. ZHU; “A Limited Memory Algorithm for Bound Constrained Optimization”; *SIAM Journal on Scientific Computing* **16**, pp. 1190–1208 (1995). 80, 82, 109

- [196] C. ZHU, R. H. BYRD, P. LU & J. NOCEDAL; “Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization”; *ACM Transactions on Mathematical Software* **23**, p. 550–560 (1997). 80, 82
- [197] P. VIRTANEN, R. GOMMERS, T. E. OLIPHANT *et al.*; “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”; *Nature Methods* **17**, pp. 261–272 (2020). 80, 153
- [198] M. ANDREUX, T. ANGLES, G. EXARCHAKIS *et al.*; “Kymatio: Scattering Transforms in Python”; *Journal of Machine Learning Research* **21**, pp. 1–6 (2020). <http://jmlr.org/papers/v21/19-047.html>; 1812.11214. 82, 153
- [199] E. ALLYS, T. MARCHAND, J.-F. CARDOSO *et al.*; “New interpretable statistics for large-scale structure analysis and generation”; *Physical Review D* **102**, p. 103506 (2020). 2006.06298. 88, 89, 90, 92, 94, 95, 97, 98, 105, 110, 154
- [200] REGALDO-SAINT BLANCARD, B., ALLYS, E., BOULANGER, F. *et al.*; “A new approach for the statistical denoising of *Planck* interstellar dust polarization data”; *A&A* **649**, p. L18 (2021). 2102.03160. 89, 100, 109, 118, 154
- [201] N. JEFFREY, F. BOULANGER, B. D. WANDELT *et al.*; “Single frequency CMB B -mode inference with realistic foregrounds from a single training image”; *Monthly Notices of the Royal Astronomical Society: Letters* (2021). 2111.01138. 97, 128
- [202] M. TRISTRAM, J. F. MACÍAS-PÉREZ, C. RENAULT & D. SANTOS; “XSPECT, estimation of the angular power spectrum by computing cross-power spectra with analytical error bars”; *Monthly Notices of the Royal Astronomical Society* **358**, pp. 833–842 (2005). astro-ph/0405575. 103
- [203] A. ZONCA, L. SINGER, D. LENZ *et al.*; “healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python”; *Journal of Open Source Software* **4**, p. 1298 (2019). 115
- [204] A. BUADES, B. COLL & J. M. MOREL; “Image Denoising Methods. A New Nonlocal Principle”; *SIAM Review* **52**, pp. 113–147 (2010). 117
- [205] K. DABOV, A. FOI, V. KATKOVNIK & K. EGIAZARIAN; “Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering”; *IEEE Transactions on Image Processing* **16**, pp. 2080–2095 (2007). 117
- [206] K. ZHANG, W. ZUO, Y. CHEN *et al.*; “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising”; *IEEE Transactions on Image Processing* **26**, pp. 3142–3155 (2017). 1608.03981. 117
- [207] N. WIENER; *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications* (The MIT Press) (1949). ISBN 9780262257190. 118
- [208] S. ZAROUBI, Y. HOFFMAN, K. B. FISHER & O. LAHAV; “Wiener Reconstruction of the Large-Scale Structure”; *The Astrophysical Journal* **449**, p. 446 (1995). astro-ph/9410080. 118

- [209] B. D. WANDELT, D. L. LARSON & A. LAKSHMINARAYANAN; “Global, exact cosmic microwave background data analysis using Gibbs sampling”; *Physical Review D* **70**, 083511 (2004). [astro-ph/0310080](#). 118
- [210] F. ELSNER & B. D. WANDELT; “Efficient Wiener filtering without preconditioning”; *Astronomy & Astrophysics* **549**, A111 (2013). [1210.4931](#). 118, 119
- [211] J. JASCHE & G. LAVAUX; “Matrix-free large-scale Bayesian inference in cosmology”; *Monthly Notices of the Royal Astronomical Society* **447**, pp. 1204–1212 (2015). [1402.1763](#). 118
- [212] J. ALSING, A. HEAVENS & A. H. JAFFE; “Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference”; *Monthly Notices of the Royal Astronomical Society* **466**, pp. 3272–3292 (2017). [1607.00008](#). 118
- [213] N. JEFFREY, F. B. ABDALLA, O. LAHAV *et al.*; “Improving weak lensing mass map reconstructions using Gaussian and sparsity priors: application to DES SV”; *Monthly Notices of the Royal Astronomical Society* **479**, pp. 2871–2888 (2018). [1801.08945](#). 118
- [214] N. JEFFREY, A. F. HEAVENS & P. D. FORTIO; “Fast sampling from Wiener posteriors for image data with dataflow engines”; *Astronomy and Computing* **25**, pp. 230–237 (2018). [1810.02821](#). 118
- [215] D. KODI RAMAHAN, G. LAVAUX & B. D. WANDELT; “Wiener filtering and pure \mathcal{E}/\mathcal{B} decomposition of CMB maps with anisotropic correlated noise”; *Monthly Notices of the Royal Astronomical Society* **490**, pp. 947–961 (2019). [1906.10704](#). 118
- [216] C. AUCLAIR *et al.* (in prep.). 125, 128
- [217] S. NAESS, S. AIOLA, J. E. AUSTERMANN *et al.*; “The Atacama Cosmology Telescope: arcminute-resolution maps of 18 000 square degrees of the microwave sky from ACT 2008–2018 data combined with Planck”; *Journal of Cosmology and Astroparticle Physics* **2020**, 046 (2020). [2007.07290](#). 127
- [218] J. M. BARDEEN, J. R. BOND, N. KAISER & A. S. SZALAY; “The Statistics of Peaks of Gaussian Random Fields”; *The Astrophysical Journal* **304**, p. 15 (1986). 129
- [219] B. T. DRAINE & A. LI; “Infrared Emission from Interstellar Dust. I. Stochastic Heating of Small Grains”; *The Astrophysical Journal* **551**, pp. 807–824 (2001). [astro-ph/0011318](#). 129
- [220] S. MILLER & D. CHILDERS; *Probability and Random Processes, 2nd Edition* (Academic Press) (2012). ISBN 9780123869814. 129, 134, 135
- [221] A. LECLAIRE; *Champs à phase aléatoire et champs gaussiens pour la mesure de netteté d'images et la synthèse rapide de textures*; Ph.D. thesis (2015). 134
- [222] P. STOICA & R. MOSES; *Spectral Analysis of Signals* (Pearson Prentice Hall) (2005). ISBN 9780131139565. 135

- [223] C. FASSINO, G. PISTONE & M. P. ROGANTIN; “Computing the Moments of the Complex Gaussian: Full and Sparse Covariance Matrix”; *Mathematics* **7** (2019). 139
- [224] S. COHEN & J. ISTAS; *Fractional Fields and Applications*; Mathématiques et Applications 73 (Springer-Verlag Berlin Heidelberg) (2013). ISBN 9783642367397. 139
- [225] D. P. KROESE & Z. I. BOTEV; “Spatial Process Simulation”; in “Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms,” , edited by V. SCHMIDT; pp. 369–404 (Springer International Publishing, Cham) (2015); ISBN 9783319100647. 139, 140
- [226] C. R. DIETRICH & G. N. NEWSAM; “Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix”; *SIAM Journal on Scientific Computing* **18**, pp. 1088–1107 (1997). 141
- [227] D. B. PERCIVAL; “Exact simulation of complex-valued Gaussian stationary processes via circulant embedding”; *Signal Processing* **86**, pp. 1470–1476 (2006). 141
- [228] A. M. SYKULSKI & D. B. PERCIVAL; “Exact simulation of noncircular or improper complex-valued stationary Gaussian processes using circulant embedding”; in “2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP),” pp. 1–6 (2016). 141
- [229] B. GALERNE, Y. GOUSSEAU & J.-M. MOREL; “Random Phase Textures: Theory and Synthesis”; *IEEE Transactions on Image Processing* **20**, pp. 257–267 (2011). 142
- [230] S. MALLAT; *A Wavelet Tour of Signal Processing: The Sparse Way*; 3rd edition (Academic Press) (2009). ISBN 9780123743701. 145, 148
- [231] C. R. HARRIS, K. J. MILLMAN, S. J. VAN DER WALT *et al.*; “Array programming with NumPy”; *Nature* **585**, pp. 357–362 (2020). 153, 154
- [232] J. D. HUNTER; “Matplotlib: A 2D graphics environment”; *Computing in Science & Engineering* **9**, pp. 90–95 (2007). 153

RÉSUMÉ

L'émission thermique de la poussière interstellaire est le principal avant-plan de la polarisation du fond diffus cosmologique (FDC) au-delà de 100 GHz. Pour cette raison, la quête de modes B dans le FDC, associés aux ondes gravitationnelles générées durant l'ère inflationnaire de l'Univers primordial, est étroitement liée à la physique du milieu interstellaire (MIS), des grains de poussières qu'il contient, et du champ magnétique qui le traverse. La complexité de cette physique fait de la caractérisation statistique du MIS magnétisé diffus un défi majeur. Pour tenir compte des statistiques non-Gaussiennes de la distribution spatiale de cette émission polarisée des poussières interstellaires, nous avons besoin de descripteurs statistiques permettant de quantifier les couplages entre échelles. Cette thèse vise donc à définir un modèle statistique de cette émission. J'emploie la *wavelet scattering transform* (WST) et les *wavelet phase harmonics* (WPH) pour obtenir des représentations multi-échelles de cartes de polarisation. La dépendance angulaire des coefficients WST peut être modélisée avec la *reduced wavelet scattering transform* (RWST), un modèle angulaire introduit dans des travaux antérieurs pour des cartes en intensité totale. La RWST fournit une description statistique des cartes de polarisation, en quantifiant leurs propriétés multi-échelles en termes de contributions isotropes et anisotropes, donc interprétables géométriquement et potentiellement physiquement. La (R)WST, et de façon similaire la WPH, permettent de définir des modèles statistiques génératifs reposant sur ces coefficients, à partir desquels de nouvelles réalisations aléatoires, statistiquement similaires aux cartes originales, peuvent être construites. Lorsque le bruit devient important dans les observations, ces statistiques sont fortement contaminées. Pour surmonter cette difficulté, j'introduis une méthode de débruitage statistique fondée sur les statistiques WPH, visant à retrouver les propriétés statistiques non Gaussiennes de l'émission non-bruitee. J'ai également développé deux logiciels pour les besoins de cette thèse, appelés PyWST et PyWPH, qui prennent la forme de paquets Python rendus publics.

MOTS CLÉS

Milieu interstellaire, poussière, statistiques, fond diffus cosmologique, champs magnétiques.

ABSTRACT

The thermal emission of interstellar dust is the main foreground to cosmic microwave background (CMB) polarization above 100 GHz. For this reason, the quest for primordial B -modes in the CMB, which are expected to arise from gravitational waves produced during the inflation era in the very early Universe, is closely related to the physics of the interstellar medium (ISM), of its dust grains, and of its magnetic field. The complexity of this physics makes the statistical characterization of the diffuse magnetized ISM a major challenge. To account for the non-Gaussian statistics of the spatial distribution of the polarized emission of interstellar dust, we need statistical descriptors that quantify couplings across scales. This thesis precisely aims to define a statistical model of this emission. I employ the wavelet scattering transform (WST) and the wavelet phase harmonics (WPH) to derive multiscale representations of polarization maps. The angular dependence of the WST coefficients can be fitted with the reduced wavelet scattering transform (RWST), an angular model introduced in previous works related to total intensity maps. The RWST provides a statistical description of polarization maps, quantifying their multiscale properties in terms of isotropic and anisotropic contributions, which can be interpreted geometrically, and potentially related to the physics of the medium. The (R)WST, and similarly the WPH, allow me to define generative statistical models, from which new random realizations statistically similar to the original maps can be drawn. When noise becomes prominent in the observations, these statistics are strongly contaminated. To overcome this difficulty, I devise a statistical denoising method based on WPH statistics, aiming at retrieving the non-Gaussian statistical properties of the noise-free emission. Additionally, I have developed two softwares for the purposes of this thesis, called PyWST and PyWPH, which take the form of public Python packages.

KEYWORDS

Interstellar medium, dust, statistics, cosmic microwave background, magnetic fields.