



**HAL**  
open science

## Generative Models of Multichannel Data from a Single Example-Application to Dust Emission

Bruno Régaldo-Saint Blancard, Constant Auclair, François Boulanger, Michael Eickenberg, François Levrier, Léo Vacher, Sixin Zhang, Erwan Allys

► **To cite this version:**

Bruno Régaldo-Saint Blancard, Constant Auclair, François Boulanger, Michael Eickenberg, François Levrier, et al.. Generative Models of Multichannel Data from a Single Example-Application to Dust Emission. *The Astrophysical Journal*, 2023, 943 (1), pp.9. 10.3847/1538-4357/aca538 . hal-04274204

**HAL Id: hal-04274204**

**<https://hal.science/hal-04274204>**

Submitted on 16 Sep 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License



# Generative Models of Multichannel Data from a Single Example—Application to Dust Emission

Bruno Régaldo-Saint Blancard<sup>1</sup>, Erwan Allys<sup>2</sup>, Constant Auclair<sup>2</sup>, François Boulanger<sup>2</sup>, Michael Eickenberg<sup>1</sup>,  
François Levrier<sup>2</sup>, Léo Vacher<sup>3</sup>, and Sixin Zhang<sup>4</sup>

<sup>1</sup>Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA; [bregaldosaintblancard@flatironinstitute.org](mailto:bregaldosaintblancard@flatironinstitute.org)

<sup>2</sup>Laboratoire de Physique de l'ENS, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, F-75005 Paris, France

<sup>3</sup>IRAP, Université de Toulouse, CNRS, CNES, UPS, Toulouse, France

<sup>4</sup>Université de Toulouse, INP, IRIT, Toulouse, France

Received 2022 August 5; revised 2022 November 21; accepted 2022 November 21; published 2023 January 19

## Abstract

The quest for primordial  $B$ -modes in the cosmic microwave background has emphasized the need for refined models of the Galactic dust foreground. Here we aim at building a realistic statistical model of the multifrequency dust emission from a single example. We introduce a generic methodology relying on microcanonical gradient descent models conditioned by an extended family of wavelet phase harmonic (WPH) statistics. To tackle the multichannel aspect of the data, we define *cross-WPH* statistics, quantifying non-Gaussian correlations between maps. Our data-driven methodology could apply to various contexts, and we have updated the software `PyWPH`, on which this work relies, accordingly. Applying this to dust emission maps built from a magnetohydrodynamics simulation, we construct and assess two generative models: (1) a  $(I, E, B)$  multi-observable input, and (2) a  $\{I_\nu\}_\nu$  multifrequency input. The samples exhibit consistent features compared to the original maps. A statistical analysis of model 1 shows that the power spectra, distributions of pixels, and Minkowski functionals are captured to a good extent. We analyze model 2 by fitting the spectral energy distribution (SED) of both the synthetic and original maps with a modified blackbody (MBB) law. The maps are equally well fitted, and a comparison of the MBB parameters shows that our model succeeds in capturing the spatial variations of the SED from the data. Besides the perspectives of this work for dust emission modeling, the introduction of cross-WPH statistics opens a new avenue to characterize non-Gaussian interactions across different maps, which we believe will be fruitful for astrophysics.

*Unified Astronomy Thesaurus concepts:* [Astrostatistics \(1882\)](#); [Interstellar dust \(836\)](#); [Cosmic microwave background radiation \(322\)](#)

## 1. Introduction

Since the Planck mission (Planck Collaboration I 2020), the observation of the cosmic microwave background (CMB) has become closely entwined with the physics of the interstellar medium (ISM). On the one hand, the thermal emission of our own Galaxy severely contaminates the CMB signal, acting as a foreground to the primordial signal. On the other hand, CMB experiments provide unique data for interstellar astrophysics. Component separation methods (see, e.g., Planck Collaboration IV 2020), which aim at recovering the accurate CMB maps, have to deal with the non-Gaussian<sup>5</sup> structure of the dust emission and the spatial variations of the spectral energy distribution (SED). In the quest for primordial  $B$ -modes (Kamionkowski & Kovetz 2016) in the CMB polarization signal, which is one of the main targets of the new generation of CMB experiments such as ACTPol (Naess et al. 2014), SPIDER (Fraisse et al. 2013), LiteBIRD (Ishino et al. 2016), the Simons Observatory (Simons Observatory Collaboration 2019), and CMB-S4 (Abazajian et al. 2019), high-precision models of the multifrequency polarized dust foreground have

become crucial (BICEP2/Keck Array & Planck Collaborations 2015).

In this paper, we address the problem of building a statistical model of multifrequency dust emission maps by introducing a generic methodology that does not rely on any prior phenomenological model (i.e., a purely *data-driven* approach). We aim at taking into account the highly non-Gaussian properties of the data and modeling the correlations between the frequency channels and between the polarization and total-intensity observables. Because, from an observational point of view, we only have a single sky, we also choose to build our model from a single multichannel input. In other words, if we call this input  $x$ , the goal is to approximate the distribution of the underlying random field  $X$  based on the single realization  $x$ . Approximating a high-dimensional distribution is always a daunting task, and with this additional constraint, this becomes even harder if not intractable. To partially alleviate this difficulty, we will thus assume (spatial) statistical homogeneity for  $X$ . Finally, we want our model to be *generative*, in the sense that drawing new samples of the model should be doable in a reasonable amount of time. This might be a crucial constraint for component separation methods requiring an important amount of simulated maps of dust emission, such as methods adopting a simulation-based inference approach as in Jeffrey et al. (2021).

Various models of the polarized emission of dust have been developed in the literature (for an extensive discussion, see Régaldo-Saint Blancard 2021, Chapter 2), some of them being packed in CMB sky simulator softwares (e.g., Delabrouille et al. 2013; Thorne et al. 2017). Among these models, we make a

<sup>5</sup> We call “non-Gaussian” any statistical feature that is not characterized by the power spectrum (or cross spectrum, when considering a pair of fields).



distinction between deterministic approaches, designed to retrieve the true emission of the sky (e.g., Planck Collaboration XI 2014; Planck Collaboration X 2016; Planck Collaboration Int. XLVIII 2016; Planck Collaboration IV 2020), and statistical approaches, focusing on its statistical properties (e.g., Vansyngel et al. 2017; Hervias-Caimapo & Huppenberger 2022). We also distinguish phenomenological approaches, making use of phenomenological priors (e.g., modified blackbody (MBB) SED), from data-driven approaches. Our approach stands in the class of statistical and data-driven models. In this category, Aylor et al. (2021) and then Thorne et al. (2021) employed a generative adversarial network (GAN) and a variational autoencoder (VAE), respectively, to model the total intensity of monofrequency dust emission maps. Krachmalnicoff & Puglisi (2021) made use of a GAN to generate subdegree angular scales in monofrequency dust polarization maps. These approaches, which involve techniques from the emerging field of deep generative modeling, rely on convolutional neural networks (CNNs) that need to be trained. Such training steps require at least hundreds if not thousands of observations, which obviously bind the models to their respective training sets. On the other hand, Allys et al. (2019) and then Regaldo-Saint Blancard et al. (2020) introduced generative models of dust emission maps in total intensity and polarization, respectively, that do not necessitate any training and can be built from a single input map. These models are approximate maximum entropy models, called *microcanonical gradient descent models*, that are conditioned by wavelet scattering transform (WST) moments (Mallat 2012; Bruna & Mallat 2013, 2019). Jeffrey et al. (2021) made use of a similar kind of model for dust polarization data but using wavelet phase harmonic (WPH) moments (Mallat et al. 2020; Allys et al. 2020; Zhang & Mallat 2021). Note that although the WST and WPH statistics share many similarities with the representations learned by CNNs (Mallat 2016; Mallat et al. 2020), they avoid some of their drawbacks, namely the lack of interpretability and the necessity of training. However, until now, none of these existing generative models have tackled the multichannel aspect of dust emission data (besides the joint modeling of the linear polarization variables; see, e.g., Regaldo-Saint Blancard et al. 2020). We propose to bridge this gap by incorporating a modeling of correlations between maps at different frequency channels, as well as correlations between total-intensity and polarization maps.

Observations of dust polarized emission at a given frequency channel  $\nu$  take the form of a triplet of two-dimensional Stokes parameter maps  $(I_\nu, Q_\nu, U_\nu)$ , with  $I_\nu$  the total intensity of the emission and  $(Q_\nu, U_\nu)$  describing the linear polarization signal. For CMB science, the polarization maps  $(Q_\nu, U_\nu)$  are usually transformed into  $(E_\nu, B_\nu)$  maps (Zaldarriaga 2001). In order to make the present work more readily usable in this cosmological context, we adopt the same polarization variables in the following.

In the continuity of the above series of works, we introduce microcanonical gradient descent models of multichannel dust emission maps  $\{(I_\nu, E_\nu, B_\nu)\}_\nu$  conditioned by a new family of WPH statistics. In particular, we introduce *cross-WPH* statistics, allowing us to combine multiple maps and characterize their non-Gaussian correlations. This paper focuses on the methodology and the validation of such models. We build two models demonstrating the capabilities of our approach: the first one is a monofrequency model based on a

triplet of simulated dust maps  $(I_{\nu_0}, E_{\nu_0}, B_{\nu_0})$ , and the second one is a multifrequency model of a set of total-intensity simulated maps  $\{I_\nu\}_\nu$ . Finally, let us emphasize that, although the focus here is on dust emission maps, our approach does not involve any prior model that is specific to dust data. As a consequence, our methodology could be applied in a very similar way to other astrophysical contexts, and more generally to other scientific contexts. All the necessary material to do so is publicly available within the Python package `PyWPH`<sup>6</sup> (Regaldo-Saint Blancard et al. 2021).

This paper is organized as follows. In Section 2, we introduce the simulated multichannel data that will be the target of our models. Then, in Section 3, we explain the underlying formalism of our approach. We introduce microcanonical gradient descent models and the family of WPH statistics that we will employ. In Section 4, we present and validate our models. Finally, Section 5 summarizes our conclusions. This paper also includes three appendices. Appendix A presents maximum entropy microcanonical models, which underlie the definition of microcanonical gradient descent models. Appendix B gives additional details on our family of WPH statistics. Appendix C properly defines the Gaussian model that we will use as a baseline in Section 4.1.

The data and codes to reproduce the models are provided in Regaldo-Saint Blancard (2022).<sup>7</sup>

## 2. Presentation of the Data

In this section, we build a set of simulated multifrequency maps  $\{(I_\nu, E_\nu, B_\nu)\}_\nu$  of dust emission by proceeding as follows:

1. In Section 2.1, we extract a statistically homogeneous 3D gas density field and magnetic field from a magnetohydrodynamics (MHD) simulation designed to reproduce typical conditions of the diffuse ISM.
2. In Section 2.2, we build Stokes maps  $I_\nu, Q_\nu,$  and  $U_\nu$  based on the previous simulation and a phenomenological model of the SED of dust grains.
3. In Section 2.3, we transform the  $\{(Q_\nu, U_\nu)\}_\nu$  maps into  $\{(E_\nu, B_\nu)\}_\nu$  maps and show the resulting data in Figure 1.

### 2.1. MHD Simulation

In order to consider a realistic gas density field  $n$  and magnetic field  $\mathbf{B}$ , we make use of an MHD simulation designed to reproduce typical conditions of the diffuse ISM. This simulation is the same as the one used in Regaldo-Saint Blancard et al. (2020). We briefly summarize its main characteristics in the following; refer to that paper for additional details.

The simulation was run in the context of Bellomi et al. (2020) to study the biphasic nature of the diffuse ISM. It employs the adaptive mesh refinement code `RAMSES` (Teyssier 2002; Fromang et al. 2006) to solve the equations of ideal MHD. The volume of the simulation is  $(50 \text{ pc})^3$ , and it is ultimately discretized on a  $512^3$  mesh with periodic boundary conditions. Heating and cooling processes of the gas are taken into account, whereas self-gravity is neglected. An isotropic Habing radiation field with  $G_0 = 1$  is applied at the boundaries of the box. An isotropic turbulent forcing is also applied,

<sup>6</sup> <https://github.com/bregaldo/pywph/>

<sup>7</sup> <https://doi.org/10.5281/zenodo.7342682>

leading to a statistical steady state after a few turnover times. In this stationary regime, the velocity dispersion is  $\sigma_v \sim 2.6 \text{ km s}^{-1}$ , and the magnetic field has a mean component  $\bar{\mathbf{B}} = B_0 \mathbf{e}_z$  with  $B_0 \sim 3.8 \mu\text{G}$  and a dispersion  $\sigma_B \sim 3.1 \mu\text{G}$ . Finally, the mean and dispersion of the gas density field  $n$  are  $\bar{n} = 1.5 \text{ cm}^{-3}$  and  $\sigma_n \sim 4.7 \text{ cm}^{-3}$ , respectively.

We extract a snapshot in this stationary regime, which thus provides the gas density field  $n$  and magnetic field  $\mathbf{B}$  we were looking for.

Note that for the quest of primordial  $B$ -modes, CMB experiments tend to focus on measurements at high Galactic latitude on degree angular scales. If this cube were to describe the structure of the Local Bubble that is responsible for most of the dust emission on Planck data at high Galactic latitudes ( $|b| > 60^\circ$ , 353 GHz; Skalidis & Pelgrims 2019; Pelgrims et al. 2020), and which extends between 200 and 300 pc away from the Sun, it would project onto the sky on a surface of about  $\sim 200 \text{ deg}^2$  at a distance of 200 pc.

## 2.2. Stokes Emission Maps

For a given frequency  $\nu$  and line of sight, and in the optically thin limit, the Stokes parameters  $I_\nu$ ,  $Q_\nu$ , and  $U_\nu$  can be expressed as follows (see, e.g., Planck Collaboration Int. XX 2015, and references therein):

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

$$Q_\nu = \int p_0 S_\nu \cos(2\phi) \cos^2 \gamma d\tau_\nu, \quad (2)$$

$$U_\nu = \int p_0 S_\nu \sin(2\phi) \cos^2 \gamma d\tau_\nu, \quad (3)$$

where  $S_\nu$  is the source function of the dust emission,  $\tau_\nu$  is the dust optical depth,  $p_0$  is an intrinsic polarization fraction parameter,  $\gamma$  is the angle that the local magnetic field makes with the plane of the sky, and  $\phi$  is the angle that the projection of the local magnetic field on the plane of the sky makes with some arbitrary reference axis in this plane. Note that the particular choice of this reference axis does not impact the ( $E_\nu$ ,  $B_\nu$ ) maps that will be derived in Section 2.3. The infinitesimal dust optical depth reads  $d\tau_\nu = \sigma_\nu n dl$ , where  $\sigma_\nu$  is the dust cross section per H at frequency  $\nu$ ,  $n$  is the hydrogen density, and  $dl$  is the infinitesimal element of length along the line of sight. All these quantities are a priori functions of the position on the line of sight.

We make the following simplifying assumptions. We assume that the polarization fraction parameter  $p_0$ , which is related to cross-section parameters of the grains and their degree of alignment with the magnetic field (for more details, see Planck Collaboration Int. XX 2015), is uniform (for a discussion, see Reissl et al. 2020). Similarly to Planck Collaboration Int. XX (2015), we choose a typical value of  $p_0 = 0.2$ . We choose the source function of the grains  $S_\nu$  to be that of a blackbody radiation  $B_\nu(T)$ , thus depending on the dust temperature  $T$ . We assume that the frequency dependence of the dust cross section follows a power law  $\sigma_\nu = \sigma_0 (\nu/\nu_0)^\beta$ , where  $\sigma_0$  is the dust cross section at the reference frequency  $\nu_0$  and  $\beta$  is the spectral index. This assumption is quite usual in the literature (see, e.g., Planck Collaboration XI 2014, and references therein). We assume a uniform  $\sigma_0$ , whose precise value only affects the global normalization of the maps and has

no impact on the rest of this paper. Finally, we arbitrarily choose the  $z$ -axis of the simulation as the line of sight. Note that this axis corresponds to the direction of the mean magnetic field of the simulation, so that we expect statistical isotropy in the resulting 2D maps (Regaldo-Saint Blancard et al. 2020).

The fields  $T$  and  $\beta$  are defined with respect to the local values of the density field  $n$  using the following simple phenomenological model. The voxels of  $T$  and  $\beta$  are both exactly Gaussian distributed in a way that is consistent with the two following observational facts: (1) temperature is usually lower (higher) in higher-density (lower-density) regions, and (2) MBB fits of observational maps of the dust emission show that the fitted  $T$  and  $\beta$  parameters tend to be anticorrelated (Planck Collaboration XI 2014). Means and standard deviations of  $T$  and  $\beta$  are chosen to be [15 K, 1 K] and [1.5, 0.2], respectively. Formally, calling  $\mathcal{R}(n(\mathbf{r}))$  the rank of a particular value  $n(\mathbf{r})$  of the simulated density field, we thus define  $T(\mathbf{r})$  and  $\beta(\mathbf{r})$  as follows:

$$\beta(\mathbf{r}) = F_{\mathcal{N}(1.5, 0.2^2)}^{-1} \left( \frac{\mathcal{R}(n(\mathbf{r}))}{1 + N^3} \right), \quad (4)$$

$$T(\mathbf{r}) = F_{\mathcal{N}(15 \text{ K}, (1 \text{ K})^2)}^{-1} \left( 1 - \frac{\mathcal{R}(n(\mathbf{r}))}{1 + N^3} \right), \quad (5)$$

where  $F_{\mathcal{N}(m, \sigma^2)}^{-1}$  is the inverse cumulative distribution function of a Gaussian distribution of mean  $m$  and standard deviation  $\sigma$ , and  $N = 512$  is the resolution of the simulation. Note that although this 3D phenomenological model of the emission of dust grains takes inspiration from MBB analyses of dust emission maps, there is a priori no reason for the SED of the projected maps  $I_\nu$ ,  $Q_\nu$ , and  $U_\nu$  to be amenable to an MBB (see, e.g., Chluba et al. 2017; McBride et al. 2022).

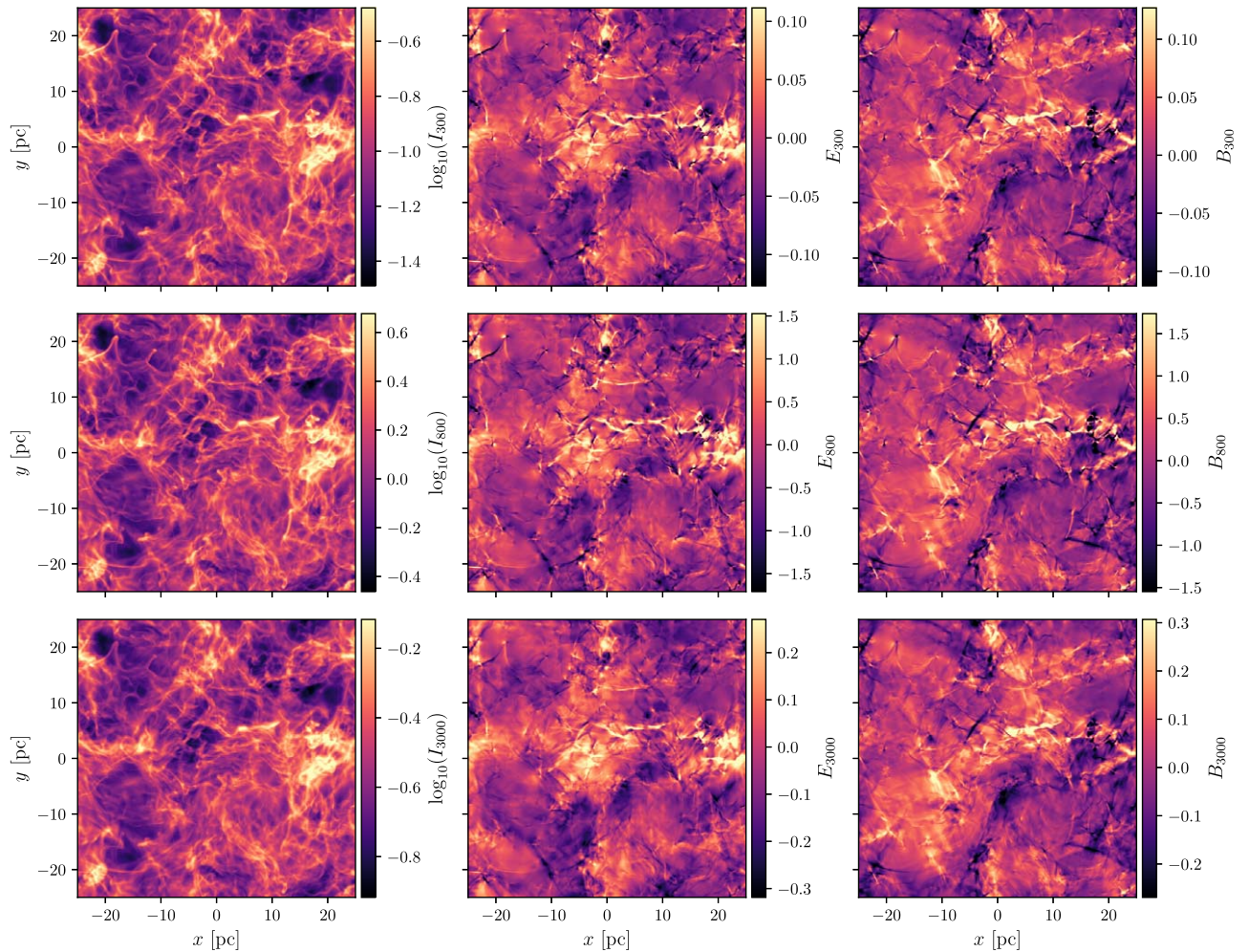
## 2.3. (E, B) Transform and Resulting Maps

In a cosmological context, linear polarization is not often studied straight from the  $Q$  and  $U$  observables, but rather in terms of  $E$ - and  $B$ -modes (Zaldarriaga 2001).  $E$ - and  $B$ -modes are linear transforms of  $Q$  and  $U$  defined as follows in Fourier space and in the flat-sky approximation (see, e.g., Kamionkowski & Kovetz 2016):

$$\begin{pmatrix} \hat{E}(\mathbf{k}) \\ \hat{B}(\mathbf{k}) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos 2\alpha_k & \sin 2\alpha_k \\ -\sin 2\alpha_k & \cos 2\alpha_k \end{pmatrix} \begin{pmatrix} \hat{Q}(\mathbf{k}) \\ \hat{U}(\mathbf{k}) \end{pmatrix}, \quad (6)$$

where  $\alpha_k$  is the angle that  $\mathbf{k}$  makes with the  $x$ -axis of the maps. The motivation for this transformation is twofold. First, for CMB science, these variables disentangle scalar primordial fluctuations from vectorial and tensorial ones (Kamionkowski et al. 1997). Indeed, vectorial and tensorial fluctuations in the primordial plasma before the decoupling would give rise to a nonzero  $B$ -mode signal in the CMB, while scalar fluctuations cannot source any  $B$ -mode signal. Second, contrary to the complex  $Q + iU$  variable, which is a spin-2 variable,  $E$  and  $B$  variables are scalar and pseudo-scalar<sup>8</sup> variables, respectively, which means in particular that these variables are invariant under any rotation of the initial basis chosen for the

<sup>8</sup> Contrary to scalar quantities, the sign of pseudo-scalar quantities is changed under a parity inversion.



**Figure 1.**  $\log_{10}(I_\nu)$ ,  $E_\nu$ , and  $B_\nu$  (from left to right) simulated maps of dust emission at three different frequency channels: 300, 800, and 3000 GHz (from top to bottom).

measurement of the polarization signal. This explains why the choice of the axis with respect to which the  $\phi$  angle introduced in Equations (2) and (3) is defined has no impact on these variables.

Following the procedure explained in this section, we thus compute  $(I_\nu, E_\nu, B_\nu)$  maps for five different frequency channels:  $\nu = 300, 500, 800, 1500, 3000$  GHz. This choice of channels is inspired by the analysis of Planck Collaboration XI (2014) on Planck and IRAS data over the 353–3000 GHz range. For this range of frequencies, the effects of  $\beta$  and  $T$  are expected to be disentangled. As an example, we display in Figure 1 the resulting maps  $I_\nu$ ,  $E_\nu$ , and  $B_\nu$  (from left to right) for the three following channels: 300, 800, and 3000 GHz. The filamentary structure of these maps underlines the non-Gaussianity of the data. Moreover, at a given frequency, coherent structures visible on the total-intensity map  $I_\nu$  have obvious counterparts in the polarization maps  $E_\nu$  and  $B_\nu$ . Statistical correlations from one observable to another are thus expected. Across the frequency axis, maps of the same observable seem to be almost structurally identical. Except for the important variation of the standard deviation of the maps and the variation of the means for total-intensity maps, the impact of frequency on the general aspect of the maps appears to be very subtle. To better see these subtle variations, we show in Figure 5 (top middle and top

right) examples of total-intensity ratio maps between consecutive frequency channels.

### 3. Formalism

This section introduces the formalism underlying the definition of our generative models, namely microcanonical gradient descent models conditioned by WPH statistics. Samples of microcanonical gradient descent models are drawn by iteratively deforming an initial Gaussian sample into a signal with some prescribed statistics. Here we choose these prescribed statistics to be WPH statistics, as these have been proven efficient in characterizing coherent structures in a variety of non-Gaussian random fields (Mallat et al. 2020; Allys et al. 2020; Jeffrey et al. 2021; Régaldo-Saint Blancard et al. 2021; Zhang & Mallat 2021; Brochard et al. 2022).

In Section 3.1, we briefly define microcanonical gradient descent models. Then, in Section 3.2, we introduce the families of moments on which the models of this paper rely. This concise introduction is complemented by Appendix B, which contains all the technical details needed to reproduce this work.

#### 3.1. Microcanonical Gradient Descent Models

Microcanonical gradient descent models were introduced in Bruna & Mallat (2019) as approximations of maximum

entropy microcanonical models.<sup>9</sup> With  $x$  the realization of the random field  $X$  that we want to approximate and  $\phi(x)$  the vector of statistics that is supposed to characterize the statistical properties of  $X$ , these models are defined by transporting an initial Gaussian distribution through gradient descent over

$$\mathcal{L}(y) = |\phi(y) - \phi(x)|^2, \quad (7)$$

where  $|\cdot|$  is the Euclidean norm. The gradient descent algorithm defines at each iteration  $k$  a mapping  $f_k(y) = y - \kappa_k \nabla \mathcal{L}(y)$ , with  $\nabla \mathcal{L}$  the gradient of  $\mathcal{L}$  and  $\kappa_k$  the gradient step at iteration  $k$ . For a given number of iterations  $n$ , and with  $y_0$  a sample drawn from our initial Gaussian distribution, the resulting sample  $y_n$  of our model reads  $y_n = f_n \circ f_{n-1} \circ \dots \circ f_1(y_0)$ . In practice, the number of iterations is empirically adapted to reach an approximate convergence.

### 3.2. Wavelet Phase Harmonic Statistics

Originally introduced in Mallat et al. (2020), the WPH statistics have been proven successful in characterizing complex coherent structures arising from a variety of two-dimensional non-Gaussian physical fields. In Zhang & Mallat (2021), they have been applied to turbulent vorticity fields, in Allys et al. (2020) to density fields of the large-scale structure of the universe, and in Regaldo-Saint Blancard et al. (2021) and Jeffrey et al. (2021) to Planck observations and simulated maps of dust polarization data. In all of these fields, structures stem from highly nonlinear physics. One of the main assets of the WPH statistics is to characterize the resulting non-Gaussianity through an efficient quantification of interactions between scales.

The WPH statistics rely on a set of bandpass and low-pass filters, allowing us to locally decompose the spectral content of the input maps onto a tiling of Fourier space. These filters are wavelets and Gaussian filters, respectively. In this section, we introduce *auto-WPH moments*, defined as covariances of nonlinear transformations of the wavelet transform of a given random field  $X$ . The nonlinear transformation is a pointwise operation called *phase harmonic*, which is defined below. Then, we extend the definition of such moments to the case of a pair of random fields ( $X, Y$ ), introducing *cross-WPH moments* designed to quantify interactions between scales across  $X$  and  $Y$ . Finally, we introduce a new family of *auto/cross-scaling moments*, defined in a similar fashion to the auto/cross-WPH moments but involving Gaussian filters instead of wavelets. These are designed to better constrain the large-scale behavior of the input fields, as well as their (joint) one-point distribution. The WPH statistics of a map  $x$ , or a pair of maps ( $x, y$ ), refer to the estimates of the auto/cross-WPH moments and auto/cross-scaling moments.

#### 3.2.1. Wavelet Transform

A wavelet is a spatially localized waveform with a zero mean that acts as a bandpass filter. From an initial wavelet  $\psi$ , called the *mother wavelet*, we build a bank of wavelets  $\{\psi_{j,\theta}\}_{j,\theta}$  by dilation and rotation of  $\psi$ , where  $j$  is an index of dilation and  $\theta$  is an angle of rotation. Formally, we have

$$\psi_{j,\theta}(\mathbf{r}) = 2^{-2j} \psi(2^{-j} \mathbf{R}_\theta^{-1} \mathbf{r}). \quad (8)$$

The numbers of dilations and rotations considered are  $J$  and  $L$ , respectively, so that  $0 \leq j \leq J-1$  and  $\theta \in \{k\pi/L, 0 \leq k \leq L-1\}$ . Consequently, our bank of wavelets is made of  $J \times L$  wavelets. In the following, we make use of bump-steerable wavelets. These are complex-valued wavelets defined in Appendix B.1.

The wavelet transform of  $X$  is finally defined as the set of bandpass-filtered maps  $\{X * \psi_{j,\theta}\}_{j,\theta}$ , where  $*$  denotes the convolution operation. These convolutions correspond to local bandpass filterings of  $X$  at spatial frequencies centered on modes of the form  $\mathbf{k}_{j,\theta} = k_0 2^{-j} \mathbf{u}_\theta$ , with  $\mathbf{u}_\theta = \cos(\theta) \mathbf{e}_x + \sin(\theta) \mathbf{e}_y$  and  $k_0$  the central frequency of the mother wavelet  $\psi$ .

In this study, we work with  $512 \times 512$  maps and choose  $J=8$  and  $L=4$ . We show in Figure 8 one wavelet from our bank.

#### 3.2.2. WPH Moments

*Auto-WPH moments.*—The auto-WPH moments of  $X$  are covariances of the phase harmonics of the wavelet transform of  $X$ , i.e., these are defined as

$$C_{\lambda,p,\lambda',p'}(\boldsymbol{\tau}) = \text{Cov}([X * \psi_\lambda(\mathbf{r})]^p, [X * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]^{p'}), \quad (9)$$

with  $\lambda$  and  $\lambda'$  referring to two oriented scales ( $j, \theta$ ) and ( $j', \theta'$ ) and the square brackets denoting the pointwise phase harmonic operator  $z \mapsto [z]^p = |z|^p \cdot e^{ip \arg(z)}$ .<sup>10,11</sup> When applied to a complex  $z$ , the phase harmonic operator preserves the modulus of  $z$  but multiplies its phase by a factor  $p$ . Note that, for  $p=0$ , this operation simply consists in taking the modulus of  $z$ , and for  $p=1$  it is the identity. This operator plays a key role in capturing efficiently interactions between different scales in  $X$ . For  $\lambda = \lambda'$  and  $p = p' = 1$ , the corresponding moments are averages of the power spectrum over the bandpass of  $\psi_\lambda$ , so that this class of moments does include the power-spectrum information. We refer to Appendix B.2 for additional details on these moments.

*Cross-WPH moments.*—The previous moments can be extended to the characterization of interactions between the scales of two different fields  $X$  and  $Y$ . We define such cross-WPH moments as follows:

$$C_{\lambda,p,\lambda',p'}^\times(\boldsymbol{\tau}) = \text{Cov}([X * \psi_\lambda(\mathbf{r})]^p, [Y * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]^{p'}). \quad (10)$$

Just like before, for  $\lambda = \lambda'$  and  $p = p' = 1$ , the corresponding moments are averages of the cross spectrum over the bandpass of  $\psi_\lambda$ , so that the cross-spectrum information is also captured by this class of moments. Note that, in the field of texture synthesis, a similar class of cross moments has been used to characterize color channel interactions in RGB images (Vacher & Briand 2021; Brochard et al. 2022).

*Discretization of  $\tau$ .*—The previous moments all depend on the relative shift  $\boldsymbol{\tau}$  between  $X * \psi_\lambda$  and  $X * \psi_{\lambda'}$  (or  $Y * \psi_{\lambda'}$ ). Inspired by Brochard et al. (2022), we discretize this variable as

<sup>9</sup> We refer the reader to Appendix A for formal definitions of maximum entropy microcanonical models.

<sup>10</sup> We recall that  $\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))^*)$ .

<sup>11</sup> These moments do not depend on the  $\mathbf{r}$  variable because of the homogeneity of  $X$ .

follows:

$$\tau_{n,\alpha} = \begin{cases} 2^n \mathbf{u}_\alpha & \text{for } n \geq 1, \\ \mathbf{0} & \text{for } n = 0, \end{cases} \quad (11)$$

with  $0 \leq n \leq \Delta_n - 1$  and  $\alpha \in \{\alpha_k = \frac{k\pi}{A}, 0 \leq k \leq 2A - 1\}$ , where  $\Delta_n$  and  $A$  are two integers playing similar roles to  $J$  and  $L$  but for  $n$  and  $\alpha$  variables instead of  $j$  and  $\theta$ , respectively. In this study, we choose  $\Delta_n = 5$  and  $A = 4$ .<sup>12</sup>

*Choice of a subset of moments.*—Estimating the auto/cross-WPH moments for every possible value of  $\lambda$ ,  $\lambda'$ ,  $p$ ,  $p'$ , and  $\tau$  is not an option for several reasons. Besides the fact that this would be computationally expensive, some moments may vanish or be redundant by construction. Moreover, the number of moments should be kept sufficiently low compared to the dimension of the data to allow for statistical diversity in the resulting generative model. Indeed, when based on a single realization  $x$ , if the number of statistical constraints is too high, samples of such models will tend to reproduce the specific features of  $x$  instead of being representative of the variability of  $X$  (for a discussion, see Brochard et al. 2022).

Therefore, we need to choose a reduced subset of moments that will characterize the relevant statistical properties of our data. This is a crucial and difficult modeling step. Although this subset should be dependent on the nature of the data, here our choice is mostly inspired by the literature on this subject (Allys et al. 2020; Régaldo-Saint Blancard et al. 2021; Brochard et al. 2022), intending for it to be relevant for a reasonable variety of non-Gaussian random fields. We refer to Appendix B.3 for a detailed presentation of this subset.

For our choice of  $J$ ,  $L$ ,  $\Delta_n$ , and  $A$  values, the resulting number of auto-WPH coefficients is 6940, which amounts to 2.6% of the number of pixels of a  $512 \times 512$  image. However, most of these coefficients are complex-valued numbers, so that the effective dimension of the description is  $\sim 1.3 \times 10^4$ , leading to a ratio of  $\sim 5\%$ . On the other hand, for a pair of maps, we consider 1264 cross-WPH coefficients, with, again, most of them being complex-valued numbers.

### 3.2.3. Scaling Moments

To better constrain the large scales, as well as the (joint) one-point distribution of the input fields, we introduce a new class of auto/cross-scaling moments relying on a family of isotropic Gaussian filters  $\{\varphi_j\}$ . These filters are built by dilating an initial Gaussian function  $\varphi$  (defined in Appendix B.1) similarly to what is done in Section 3.2.1 for wavelets. Assuming zero-mean  $X$  and  $Y$ , our auto/cross-scaling moments are defined as follows:

$$L_{j,p,p'} = \text{Cov}([X * \varphi_j(\mathbf{r})]^p, [X * \varphi_j(\mathbf{r})]^{p'}), \quad (12)$$

$$L_{j,p,p'}^\times = \text{Cov}([X * \varphi_j(\mathbf{r})]^p, [Y * \varphi_j(\mathbf{r})]^{p'}). \quad (13)$$

In this work, we choose  $j \in \{-1, 0, 1, 2\}$  and consider the following set of values for  $(p, p')$ :  $\{(0, 0), (1, 1), (0, 1)\}$  for auto moments, and  $\{(0, 0), (1, 1), (0, 1), (1, 0)\}$  for cross moments.<sup>13</sup> These moments complement our description with a

<sup>12</sup> Since we deal with fields with periodic boundary conditions, in order to avoid redundancy in the coefficients,  $\Delta_n$  should verify  $\Delta_n < \log_2(N/2) + 1$ , where  $N$  is the number of pixels along the smallest axis of our images. Our choice meets this criterion.

<sup>13</sup> Note that this choice of  $(p, p')$  values for the cross moments makes the resulting set of moments invariant under the exchange of  $X$  and  $Y$ .

very small number of coefficients: 12 and 16 in the auto and cross cases, respectively.

### 3.2.4. Estimation and Computation

We estimate the previous moments from a given map  $x$  or couple of maps  $(x, y)$  as explained in Appendix B.5. Note that the resulting statistical coefficients include a normalization designed to facilitate the gradient descent involved during the sampling of our model (see Section 3.1). We denote by  $\phi_{\text{auto}}(x)$  and  $\phi_{\text{cross}}(x, y)$  the corresponding vector of auto- and cross-WPH statistics, respectively. Computations all employ the GPU-accelerated Python package `PyWPH` (Régaldo-Saint Blancard et al. 2021).

## 4. Models and Their Validation

This section presents the core results of this paper: the construction and validation of two distinct generative models built from the simulated data introduced in Section 2. These are as follows:

1. A monofrequency model based on the joint observation of  $(I_{300}, E_{300}, B_{300})$ , introduced in Section 4.1.
2. A multifrequency model based on the joint observation of  $(I_{300}, I_{500}, I_{800}, I_{1500}, I_{3000})$ , introduced in Section 4.2.

For each of these models, we perform a visual and quantitative assessment of their realism.

The data and codes to reproduce the models are provided in Régaldo-Saint Blancard (2022).<sup>14</sup>

### 4.1. Monofrequency Model

#### 4.1.1. Description of the Model

We build a generative model based on the joint observation of  $(I_{300}, E_{300}, B_{300})$  as follows (referred to as “the WPH model” hereafter). We define a microcanonical model of  $\mathbf{x} = (\log(I_{300}), E_{300}, B_{300})$  conditioned by the following descriptive statistics:

$$\begin{aligned} \phi(\mathbf{x}) = & \phi_{\text{auto}}(\log(I_{300})) \oplus \phi_{\text{auto}}(E_{300}) \oplus \phi_{\text{auto}}(B_{300}) \\ & \oplus \phi_{\text{cross}}(\log(I_{300}), E_{300}) \oplus \phi_{\text{cross}}(\log(I_{300}), B_{300}), \end{aligned} \quad (14)$$

where  $\oplus$  denotes vector concatenation. We choose to model the logarithm of the total-intensity map  $I_{300}$  for two main reasons: (1) the logarithm tends to Gaussianize the data (analytical statistical models of dust total-intensity maps are usually defined as lognormal fields; see, e.g., Levrier et al. 2018) and thus to simplify the modeling, and (2) pixels of  $I$  maps take positive values, so that working with the logarithm avoids imposing positive values on the synthetic  $I$  maps during the sampling procedure. Also note that  $\phi_{\text{cross}}$  is designed to be *pseudo*-symmetric (see Appendix B.4), so that  $\phi_{\text{cross}}(\log(I_{300}), E_{300})$  and  $\phi_{\text{cross}}(E_{300}, \log(I_{300}))$  are equally informative in characterizing the couplings between  $\log(I_{300})$  and  $E_{300}$  (and obviously this remains true for  $(\log(I_{300}), B_{300})$ ). Our choice to ignore the couplings between  $E_{300}$  and  $B_{300}$  is motivated by observational measurements indicating a null  $EB$  cross spectrum (Planck Collaboration XI 2020). However, note that these measurements do not say

<sup>14</sup> <https://doi.org/10.5281/zenodo.7342682>

anything on higher-order cross statistics, so that this modeling choice remains partly arbitrary.

The effective dimension of  $\phi(\mathbf{x})$  is  $\sim 4.4 \times 10^4$ , which amounts to less than 6% of the total number of pixels of  $\mathbf{x}$ . Samples are drawn by minimizing the objective function defined in Equation (7), starting from a triplet of realizations of three independent Gaussian random fields having the same power spectrum as the empirical power spectrum of  $\log(I_{300})$ ,  $E_{300}$ , and  $B_{300}$ , respectively.<sup>15</sup> To perform this optimization, we use the L-BFGS algorithm that is implemented in SciPy (Byrd et al. 1995; Zhu et al. 1997; Virtanen et al. 2020).<sup>16</sup> This necessitates the computation of the gradients of the objective with respect to the pixels of the optimized maps at every iteration. We compute such gradients using automatic differentiation as implemented in PyTorch (Paszke et al. 2019). The sampling takes  $\sim 6$  minutes for 200 iterations on an NVIDIA A100-SXM4 GPU with 40 GB of memory.

In the following, we also make use of a Gaussian equivalent of this previous model as a baseline (referred to as “the Gaussian model”). It is a microcanonical model conditioned by the auto- and cross-WPH statistics that only estimate the power- and cross-spectrum information, respectively. As before, we ignore the couplings between  $E_{300}$  and  $B_{300}$ . We refer to Appendix C for additional details on this model and to Figure 10 for visual examples of its samples. The samples are drawn as above.

#### 4.1.2. Visual Assessment

We show in Figure 2 the original simulated maps ( $\log(I_{300})$ ,  $E_{300}$ ,  $B_{300}$ ) (first row, from left to right) next to three different samples (also referred to as “syntheses”) of the WPH model (second to fourth rows).

The synthetic  $\log(I_{300})$  maps statistically reproduce the main features of the original map. Given the important variability of structures across samples and the difficulty in distinguishing original from synthetic maps at first glance, this model seems to provide a relevant approximation of the underlying probability distribution of the original data. We emphasize that to achieve similar results, deep generative models usually require thousands of observations, whereas here our model is only based on a single example. In Aylor et al. (2021) and Thorne et al. (2021), slightly more than 1000 total-intensity maps were used to train a GAN and a VAE, respectively. While the visual quality of the GAN syntheses is roughly equivalent to ours, this is not the case for that of the VAE model, which notably fails to reproduce small-scale patterns of the maps.

For the polarization maps  $E_{300}$  and  $B_{300}$ , there is also qualitative visual agreement between the synthetic and original maps, although close scrutiny reveals subtle artifacts compared to the  $I$  case. The diversity of structures across samples is still satisfactory in this case.

Finally, spatial correlations between the total-intensity maps and the polarization maps are also well reproduced. Circles in black dashed lines show examples of such correlations for

both the original and synthetic data. Taking into account these correlations is a significant improvement over both previous microcanonical models (Allys et al. 2019; Régaldo-Saint Blancard et al. 2020; Jeffrey et al. 2021) and deep generative models of dust emission maps (Aylor et al. 2021; Krachmalnicoff & Puglisi 2021; Thorne et al. 2021).

#### 4.1.3. Quantitative Assessment

We choose to quantitatively assess the realism of the WPH model by means of a statistical comparison. We build 10 independent samples of both the WPH and the Gaussian model, following the same procedure as before. In Figure 3, we compare the empirical power spectra, distribution of pixels values, and the Minkowski functionals derived from the samples of the WPH model, those of the Gaussian model, and the original maps, for each of the observables  $I$ ,  $E$ , and  $B$  taken separately. For the WPH Gaussian models, we show mean statistics across the 10 independent samples, and when displayed, the error bars correspond to the standard deviations across these samples. In Figure 4, we compare the joint distributions of pixel values between all pairs of observables in the original, WPH, and Gaussian cases. Contrary to Figure 3, in the WPH and Gaussian cases, we show statistics estimated from a single sample of the model. The results are discussed below.

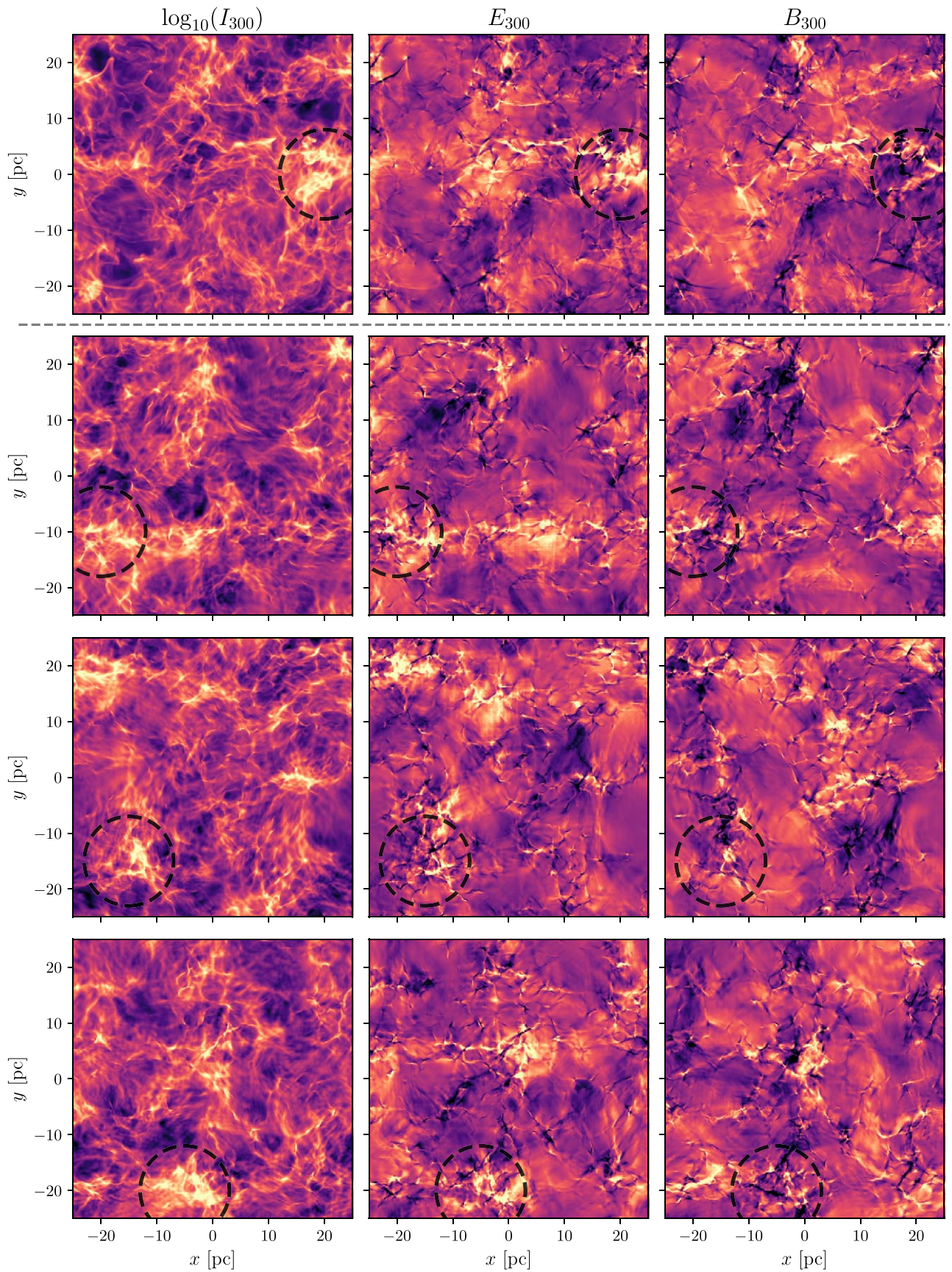
*Power spectrum.*—We first recall that our models directly include power-spectrum constraints (see Appendix B.2 for additional details), so that the power spectra of the synthetic maps are expected to match those of the original maps very well. We show the power-spectrum analysis in the top row of Figure 3. While the power spectra of the  $E$  and  $B$  maps are indeed very well reproduced, the agreement is slightly worse for the  $I$  maps at small scales. The discrepancies take the form of a very subtle excess of power in the synthetic maps. As explained previously, we constrain the WPH statistics of  $\log(I)$  instead of  $I$ , so that these kinds of discrepancies are not surprising. We have checked that the power spectrum of  $\log(I)$  is very well reproduced, as expected. These discrepancies thus show that the WPH model does not perfectly capture the power spectrum of the exponential of the constrained data. Similarly, we observe comparable discrepancies between the power spectrum of the Gaussian data and that of the original data. Again, the Gaussian model applies to  $\log I$  and not  $I$ , meaning that  $I$  is modeled by a log-Gaussian model. Such discrepancies thus underlie the limit of log-Gaussian models for dust intensity maps.

*Distribution of pixels.*—We show the distributions of pixel values in the second row of Figure 3. The bulks of these distributions agree between the syntheses and the original maps to a very good extent (three orders of magnitude on the  $y$ -axis). Discrepancies appear in the tails of the distributions. These are subtle for the polarization observables  $E$  and  $B$  but more significant for the  $I$  maps. Note that Aylor et al. (2021) pointed out similar difficulties in capturing the tails of these distributions with their GAN model. The quality of the constraints on these statistics within the WPH model highly depends on our choice of scaling moments. It is likely that we could improve these results by taking into account more  $j$  values in the definition of our subset of scaling moments (see Section 3.2.3). Compared to the Gaussian model, the WPH model does much better, showing its ability to capture non-Gaussian properties of the data.

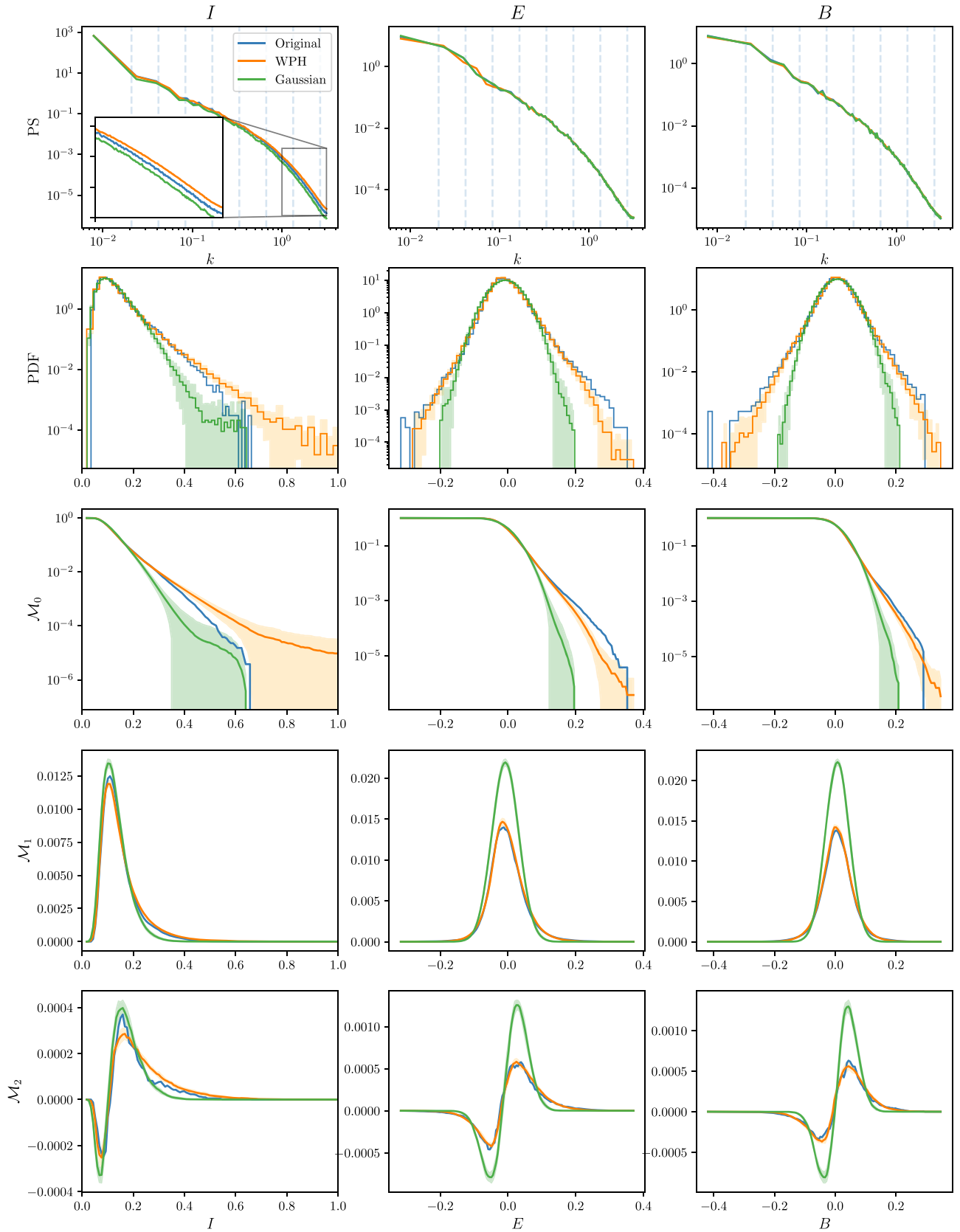
<sup>15</sup> Note that the power spectrum of the maps is constrained during the optimization; thus, we could have also started from independent realizations of Gaussian white noises without impacting the results of this paper. Here, this choice is only motivated by the numerical efficiency of the optimization.

<sup>16</sup> The L-BFGS algorithm is a quasi-Newton method, which is not, properly speaking, a gradient descent method. However, this algorithm has been shown experimentally to be more suitable for solving this optimization problem compared to standard gradient descent algorithms.

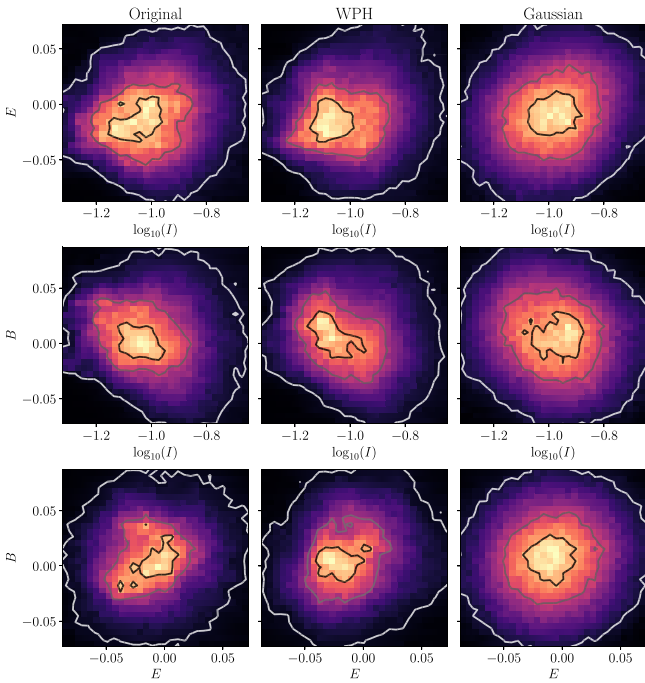




**Figure 2.** Joint ( $\log_{10}(I_{300})$ ,  $E_{300}$ ,  $B_{300}$ ) syntheses (from second to last rows) sampled from the generative model described in Section 4.1, and which only relies on the statistics of the joint map shown in the top row (same joint map as the top row of Figure 1). Dashed circles show examples of spatial correlations across the observables  $I$ ,  $E$ , and  $B$ .



**Figure 3.** Quantitative validation of the monofrequency model described in Section 4.1. We compare the empirical power spectra (first row), distributions of pixel values (second row), and Minkowski functionals (third to fifth rows) computed separately for the  $I$ ,  $E$ , and  $B$  maps (from left to right) of the original, WPH, and Gaussian data. For the WPH and Gaussian models, we show mean statistics across 10 independent samples, and the errors are the standard deviations across these samples. Vertical dashed lines on the power-spectrum plots mark the central frequencies of the wavelets used for this analysis.



**Figure 4.** Quantitative validation of the monofrequency model described in Section 4.1. We compare the joint distributions of the pixel values of  $(\log(I), E)$ ,  $(\log(I), B)$ , and  $(E, B)$  (from top to bottom) in the original, WPH, and Gaussian maps (from left to right). We add, for each of these plots, contours at  $0.5\sigma$ ,  $1\sigma$ , and  $2\sigma$ . For the WPH and Gaussian models, statistics are estimated from a single sample.

*Minkowski functionals.*—Minkowski functionals are often used to characterize the morphological aspects of smooth random fields. In cosmology, these have been used in various contexts, such as the investigation of potential non-Gaussianity and anisotropy in the CMB (see, e.g., Planck Collaboration Int. XXIII 2015; Planck Collaboration VII 2020), the characterization of the large-scale structure (e.g., Codis et al. 2013), or that of weak-lensing data (e.g., Parroni et al. 2020). In the ISM community, these are, however, much less popular statistics, although it has already been applied in the context of dust modeling (Aylor et al. 2021; Burkhart 2021; Krachmalnicoff & Puglisi 2021). In two dimensions, there are three Minkowski functionals  $\mathcal{M}_0$ ,  $\mathcal{M}_1$ , and  $\mathcal{M}_2$ , which are defined as follows. For a given map  $x$  and  $\alpha \in \mathbb{R}$ , we define the excursion set  $\Gamma_\alpha = \{r \mid x(r) \geq \alpha\}$ , which simply corresponds to the region where the field is greater than a given threshold. The Minkowski functionals associated with  $x$  are

$$\mathcal{M}_0(\alpha) = \frac{1}{A} \int_{\Gamma_\alpha} da, \quad (15)$$

$$\mathcal{M}_1(\alpha) = \frac{1}{2\pi A} \int_{\partial\Gamma_\alpha} dl, \quad (16)$$

$$\mathcal{M}_2(\alpha) = \frac{1}{2\pi^2 A} \int_{\partial\Gamma_\alpha} \kappa dl, \quad (17)$$

where  $A$  is the total area of the field;  $\partial\Gamma_\alpha$  is the boundary of  $\Gamma_\alpha$ ;  $da$  and  $dl$  are the surface and line elements associated with  $\Gamma_\alpha$  and  $\partial\Gamma_\alpha$ , respectively; and  $\kappa$  is the curvature of  $\partial\Gamma_\alpha$ . The functionals  $\mathcal{M}_0$ ,  $\mathcal{M}_1$ , and  $\mathcal{M}_2$  are called the area, perimeter, and genus (i.e., number of “holes”) characteristics, respectively.

We show in the third, fourth, and fifth rows of Figure 3 the Minkowski functional statistics for our data. We compute these

**Table 1**

Mutual Information (in Shannon Units) Associated with the Joint Distributions of Pixel Values of  $(\log_{10}(I), E)$ ,  $(\log_{10}(I), B)$ , and  $(E, B)$ , Estimated for the Original, WPH, and Gaussian Data

	Original	WPH	Gaussian
$(\log_{10}(I), E)$	0.092	$0.095 \pm 0.006$	$0.033 \pm 0.002$
$(\log_{10}(I), B)$	0.122	$0.101 \pm 0.006$	$0.020 \pm 0.003$
$(E, B)$	0.059	$0.054 \pm 0.008$	$0.014 \pm 0.002$

**Note.** In the WPH and Gaussian cases, the values correspond to the means and standard deviations of the mutual information across 10 independent samples.

statistics using the Python package `QuantImpy` (Boelens & Tchepeli 2021). The area  $\mathcal{M}_0$  just gives another perspective of the previous distributions of pixels, as it directly relates to the cumulative distribution of pixel values. Contrary to this, the perimeter  $\mathcal{M}_1$  and genus characteristics  $\mathcal{M}_2$  provide new insights on our models. We see that for the polarization observables  $E$  and  $B$ , these statistics are very well reproduced by the WPH model. In comparison, the Gaussian model performs poorly. Comparable results on polarization data were obtained in Krachmalnicoff & Puglisi (2021) using a GAN (although in a slightly different context). Being able to reproduce such results with a model based on a single observation and without training is one of the successes of our approach. On the other hand, for  $I$ , although the WPH model seems to do slightly better than the Gaussian model for  $\mathcal{M}_1$ , this is less clear for  $\mathcal{M}_2$ . This underlies the higher difficulty of the WPH model to perfectly capture the complexity of  $I$  maps. Note that we would get similar results for  $\log(I)$  on  $\mathcal{M}_1$  and  $\mathcal{M}_2$  since applying a pointwise monotonous function to our maps would not impact the level sets involved in the derivation of these statistics.

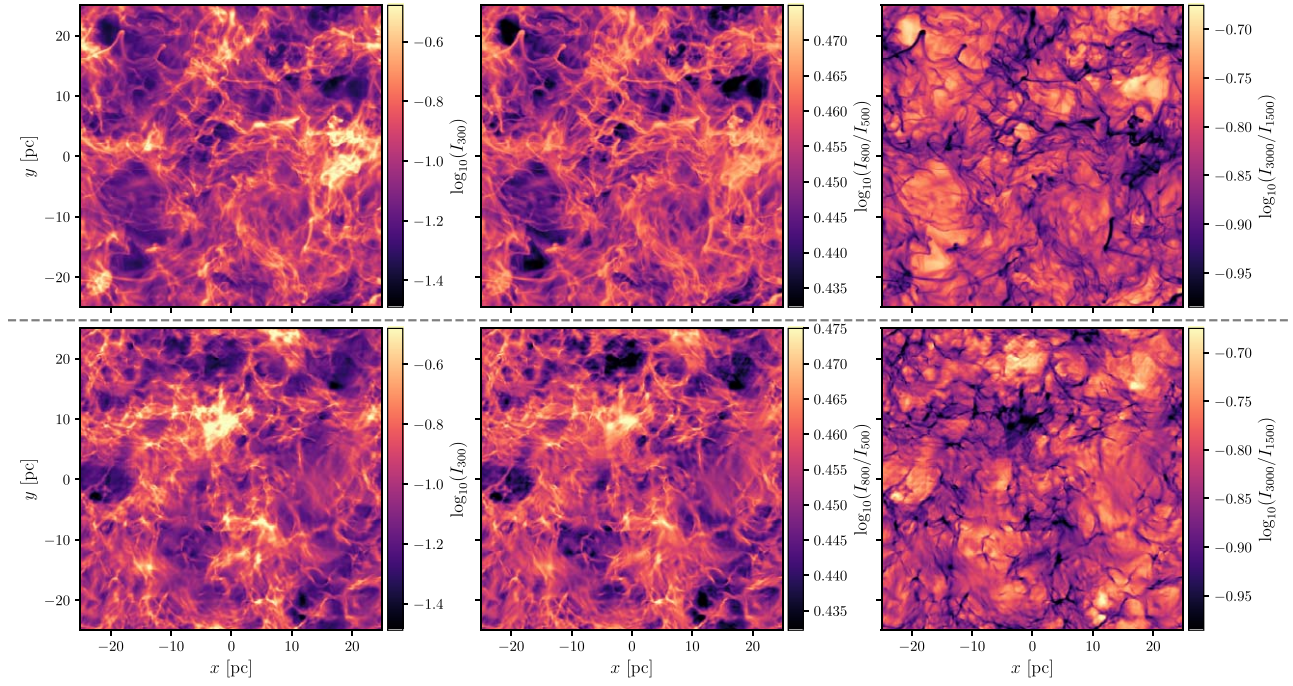
*Joint distribution of pixels.*—Figure 4 shows the joint distributions of pixel values of  $(\log(I), E)$ ,  $(\log(I), B)$ , and  $(E, B)$  (from top to bottom row) for our data. These statistics are, again, mostly constrained by the scaling moments we have introduced in the WPH model. The agreement of these distributions between the WPH model and the original data is satisfactory, and it is significantly better than the Gaussian case. Note that no cross constraint between  $E$  and  $B$  was imposed, so that it is not surprising to find a slightly worse agreement of the contours. In order to make sure that this agreement is not a mere consequence of the agreement of the marginal distributions of pixel values, we have estimated the mutual information associated with each of these distributions.<sup>17</sup> The results of this analysis are shown in Table 1. These show a significant dependence between all pairs of observables on the original data, as well as a significantly better capture of these dependencies in the WPH model compared to the Gaussian model.

## 4.2. Multifrequency Model

### 4.2.1. Description of the Model

We now define a generative model of the multifrequency simulated maps  $(I_{300}, I_{500}, I_{800}, I_{1500}, I_{3000})$  as follows. For the

<sup>17</sup> The mutual information of two random variables  $X$  and  $Y$  quantifies the mutual dependence between  $X$  and  $Y$  and is defined as the Kullback–Leibler divergence of the joint distribution  $p_{X,Y}$  from the product of the marginal distributions  $p_X p_Y$ .



**Figure 5.**  $\log_{10}(I_{300})$ ,  $\log_{10}(I_{800}/I_{500})$ , and  $\log_{10}(I_{3000}/I_{1500})$  synthetic maps (bottom row, from left to right) sampled from the generative model described in Section 4.2, and which is based on the maps shown in the top row.

same reasons as before, we prefer to work with the logarithm of these intensity maps. Since the maps from one frequency channel to another look very similar (modulo a global scaling or shift of the mean), we choose to impose statistical constraints on the differences of the logarithms of these maps (which are also the logarithms of the ratios of these maps) between consecutive frequency channels. One of the maps also needs to be constrained in absolute to serve as a reference. Consequently, with  $\nu_1, \dots, \nu_5 = 300, \dots, 3000$  GHz, we define a micro-canonical model of  $\mathbf{x} = (\log(I_{\nu_i})) \oplus (\log(I_{\nu_{i+1}}/I_{\nu_i}))_{1 \leq i \leq 4}$  conditioned by

$$\phi(\mathbf{x}) = (\phi_{\text{auto}}(x_i))_{1 \leq i \leq 5} \oplus (\phi_{\text{cross}}(x_i, x_j))_{1 \leq i < j \leq 5}. \quad (18)$$

Contrary to the monofrequency case, we consider cross constraints between all the possible pairs of maps. Although this significantly increases the dimensionality of the statistical description, as well as the computational cost of the sampling procedure, this gives the best visual and quantitative results in what follows. The effective dimension of  $\phi(\mathbf{x})$  is  $\sim 8.9 \times 10^4$ , which amounts to  $\sim 7\%$  of the total number of pixels of  $\mathbf{x}$ . The sampling procedure takes  $\sim 17$  minutes for 200 iterations on an NVIDIA A100-SXM4 GPU with 40 GB of memory. We make use of a single sample of this model in the following.

#### 4.2.2. Visual Assessment

We show in Figure 5 a selection of maps derived from the resulting multifrequency synthesis (bottom row) below the corresponding original maps (top row). We show the  $\log(I_{300})$  maps and the ratio maps  $\log(I_{800}/I_{500})$  and  $\log(I_{3000}/I_{1500})$  (from left to right). Visually, the synthetic maps are very satisfactory, with a realistic filamentary structure. Moreover, spatial correlations from one map to another are consistent with those visible on the original maps. This shows that our statistical description seems to properly capture interactions between frequency channels.

#### 4.2.3. Quantitative Assessment

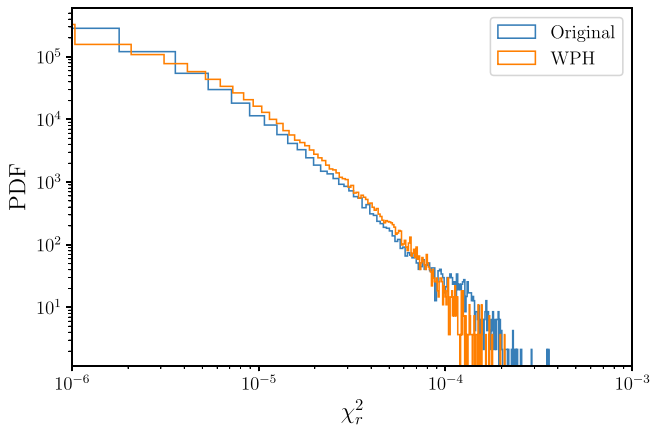
We quantitatively assess the realism of such multifrequency models by focusing on how well both the original and synthetic multifrequency maps can be fitted by an MBB model and comparing their respective parameters. MBB models are ubiquitous in the interstellar dust literature and constitute the standard means to model and parameterize the frequency dependence of emission maps (see, e.g., Planck Collaboration XI 2014). These are directly related to the model of the emission and absorption properties of dust grains that was employed in Section 2 to build our original simulated maps. Indeed, an MBB model  $\tilde{I}_\nu(\mathbf{r})$  is defined as

$$\tilde{I}_\nu(\mathbf{r}) = \tau(\mathbf{r}) \nu^{\beta(\mathbf{r})} B_\nu(T(\mathbf{r})), \quad (19)$$

where, in addition to the spectral index map  $\beta(\mathbf{r})$  and temperature map  $T(\mathbf{r})$ , we also introduce the optical depth map  $\tau(\mathbf{r})$ .

For each pixel of both the original and synthetic multifrequency data, we fit the parameters of this model to the data by performing a nonlinear least-squares minimization. To do that, we employ the SciPy implementation of the Levenberg–Marquardt algorithm (Virtanen et al. 2020). We perform the minimization on logarithmic values and initialize the parameters with  $T_0 = 20$  K,  $\beta_0 = 1.5$ , and  $\tau_0 = I_{300}(\mathbf{r}) / (300^{\beta_0} B_\nu(T_0))$ . We show in Figure 6 the resulting distribution of  $\chi_r^2$  values. The  $\chi_r^2$  values are nowhere larger than  $\sim 3 \times 10^{-4}$ , showing that the fit performs very well on both the simulated and synthetic data.<sup>18</sup> This result is not particularly surprising for the original data (albeit nontrivial), given the fact that we had employed an MBB-like law of dust grain emission at the voxel level. However, no prior information on this MBB law was given in our model, which

<sup>18</sup> Since we perform the fit on logarithmic values,  $\chi_r^2$  roughly correspond to the mean square relative error.



**Figure 6.** Distributions of  $\chi_r^2$  values associated with the pixelwise MBB fits on the original and synthetic multifrequency  $I$  maps.

shows that our statistical description has been somehow able to capture this SED. Moreover, the agreement of the  $\chi_r^2$  distributions is a stronger result, as it shows that the MBB is as well suited to model the SED of the synthetic maps as it is for the original maps, which is the ideal behavior of such a model.

We show in Figure 7(a) the resulting parameter maps  $\tau$ ,  $T$ , and  $\beta$  for both the original (top row) and synthetic data (bottom row), as well as their distributions in panel (b). Let us first remark that the anticorrelation between  $\beta$  and  $T$  and the anticorrelation between  $n$  and  $T$  that were instilled at the voxel level when building the simulated data in Section 2.3 are clearly reflected in the parameter maps of this (projected) simulated data. Indeed,  $T$  clearly tends to be higher (lower) for low (high) values of  $\tau$ , and similarly,  $T$  and  $\beta$  appear to be anticorrelated. The synthetic parameter maps exhibit the same properties. Moreover, the consistency of the structures between the synthetic and original parameter maps shows that our model is able to capture the spatial variations of this frequency dependence in a very satisfactory way. Comparisons of the distributions of these parameter maps strengthen this conclusion.

Note that, in real observational conditions, departures from the MBB model are expected in the dust signal. Formalisms to deal with them have been studied in the literature for both total-intensity (see, e.g., Chluba et al. 2017; Hensley & Bull 2018; Mangilli et al. 2021) and polarization data (Ichiki et al. 2019; McBride et al. 2022; Vacher et al. 2023), and we expect such more refined models to play a central role for high-precision dust foreground modeling in the context of primordial  $B$ -mode detection. In this context, we believe that learning such models directly from the data, as shown here, constitutes a valuable complementary approach to the accurate modeling of this frequency dependence.

## 5. Conclusion

We have introduced a generic methodology using WPH statistics to build generative models of multichannel non-Gaussian data. Our approach is purely data driven, and the models can be derived from a single example of simulated or observational data. Models are microcanonical gradient descent models conditioned by the set of WPH statistics of the target data. We have reviewed the main ideas of the underlying

formalism and introduced an extended family of WPH statistics enhancing previous works.

In order to characterize the couplings between two different channels, we have introduced the cross-WPH moments. These are key to this paper. They correspond to direct extensions of the (auto-)WPH moments previously introduced in the literature. They include the cross-spectrum information but also quantify non-Gaussian interactions across channels. We also have introduced a new family of moments, called auto/cross-*scaling moments*, that is designed to probe the largest scales and better constrain the single map or joint one-point distribution. The Python package `PyWPH`,<sup>19</sup> previously released in Regaldo-Saint Blancard et al. (2021), has been updated to easily build such multichannel generative models from two-dimensional data.

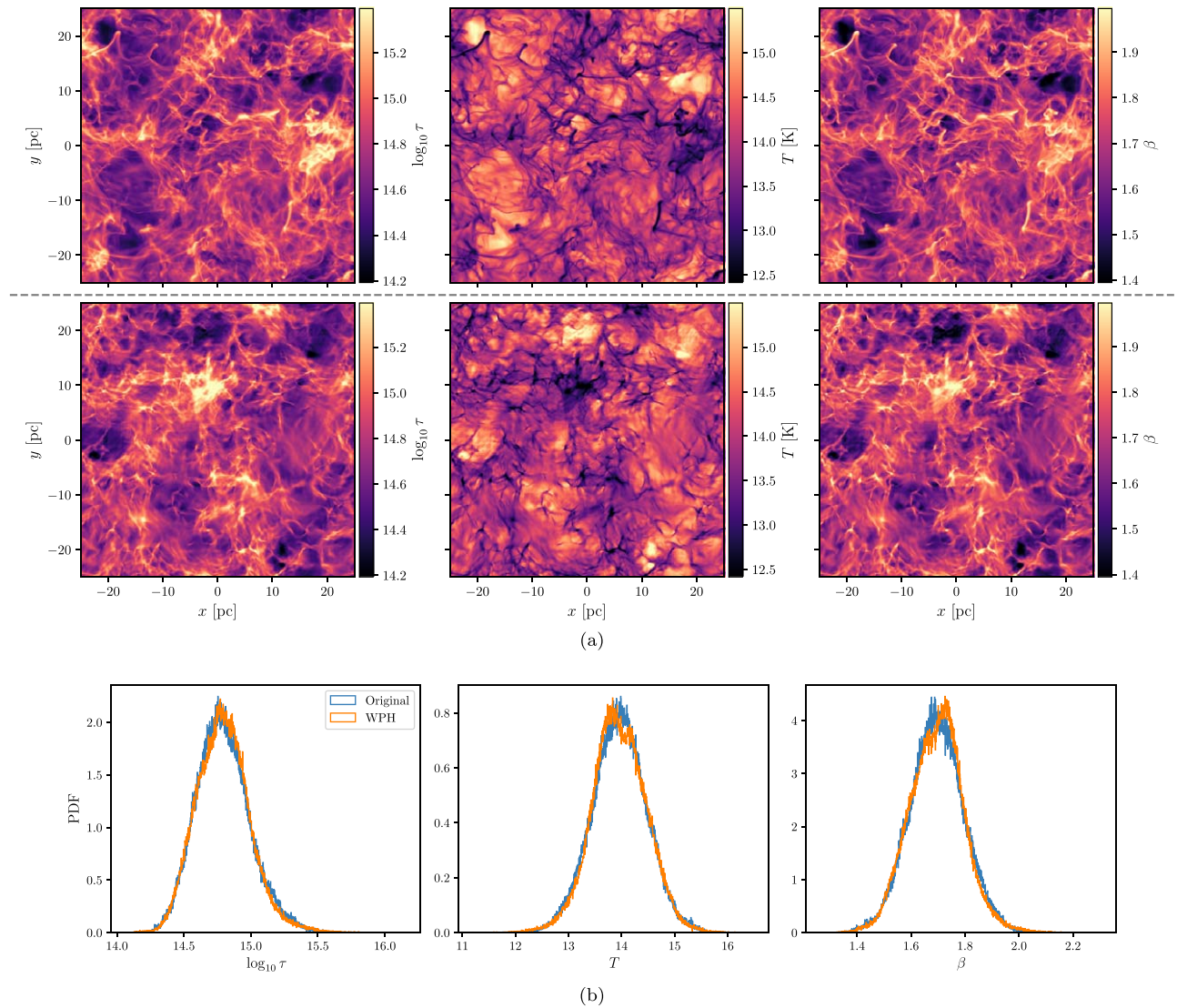
Here we have applied this methodology to multichannel simulated maps of the interstellar dust emission. The simulated maps were built from an MHD simulation assuming a simple phenomenological model of the multifrequency emission of dust grains. We have constructed and studied two separate generative models:

1. A model of a monofrequency  $(I, E, B)$  input taking into account  $I-E$  and  $I-B$  correlations. Samples of this model were shown to exhibit features statistically consistent with those of the original maps, including correlated structures across  $I$ ,  $E$ , and  $B$ . A quantitative assessment demonstrated that the empirical power spectra, distributions of pixel values, and Minkowski functionals of the  $I$ ,  $E$ , and  $B$  synthetic maps match the corresponding statistics of the original maps to a good extent. However, slight discrepancies of these statistics were found for the  $I$  maps. Finally, an analysis of the joint distributions of pixel values has shown that our model captures the dependencies between the observables to a good extent.
2. A model of a multifrequency  $(I_\nu)_\nu$  input, for five frequency channels, and taking into account the strong correlations of the maps across the frequency axis. Samples of the model exhibit features and correlations consistent with those of the original maps. We have quantitatively assessed this model by fitting the SED of our data with an MBB law. This law fits very well and consistently both the synthetic maps and the original maps. The comparison of the parameter maps has shown a very good agreement, underlining the success of our approach in capturing the spatial variations of the SED from the data.

The perspectives of this work are numerous.

First, we underline that the statistical validation of our models could be improved. Indeed, a more rigorous approach would be to compare the statistics of the samples of our model to those of a large number of independent samples of the random field  $X$  we wish to approximate. We have chosen here to build our model from a single observation  $x$ , implicitly assuming that the statistics of  $x$  are representative of those of  $X$  (which relates to an “ergodic” assumption; see Bruna & Mallat 2019), and consequently, the validation of our models was performed relatively to the statistics of  $x$  only. We could go beyond this assumption by studying the variability of the

<sup>19</sup> <https://github.com/bregaldo/pywph/>



**Figure 7.** (a) MBB parameter maps  $\log_{10}(\tau)$ ,  $T$ , and  $\beta$  (from left to right) resulting from the pixelwise fits of an MBB model to both the original (top row) and synthetic (bottom row) multifrequency  $I$  maps. (b) Resulting distributions of these parameters.

statistics of a collection of samples  $\{x_i\}$  of  $X$  in our analysis and validating our model with respect to this variability. However, let us mention that in our case the data rely on a costly MHD simulation, which prevents the construction of an arbitrarily large number of samples (for a discussion, see Regaldo-Saint Blancard et al. 2020).

We will address the modeling of multifrequency polarization maps in a follow-up paper. In the case of multifrequency  $I$  maps, we have constructed a microcanonical model based on the ratio maps between consecutive channels. Proceeding similarly for  $E$  and  $B$  is not an option since these variables can exhibit null values. Moreover, the nature of the linear polarization observable is vectorial, which might require the introduction of a complex variable  $E + iB$ , or  $Q + iU$  as in Regaldo-Saint Blancard et al. (2020, 2021).<sup>20</sup> These modeling choices should take inspiration from the literature on analytical models of the dust polarization SED (Ichiki et al. 2019; McBride et al. 2022; Vacher et al. 2023).

<sup>20</sup> Note that this is not an obstacle to our methodology and that the computation of WPH statistics on complex-valued maps is supported by PyWPH.

From a more general perspective, the cross-WPH statistics open a new way to analyze and combine multichannel data, allowing us to efficiently describe and model non-Gaussian correlations across different maps. Moreover, given a WPH model of the interactions between two channels  $A$  and  $B$ , and combined with the formalism of microcanonical models, these statistics should allow for statistically relevant predictions of  $B$  maps based on the observation of  $A$  maps. Such “conditioned” models will be explored in further works.

In the context of dust modeling for CMB analysis, such generative models are expected to improve forward models of the CMB sky. In Jeffrey et al. (2021), a WPH model of dust polarization maps played a crucial role for CMB foreground marginalization in a likelihood-free inference framework. In this work, the inference was performed on monofrequency maps; the present paper now paves the way to an extension to the multifrequency case.

Finally, this work is expected to improve statistical denoising methods of dust polarization maps, such as introduced in Regaldo-Saint Blancard et al. (2021). Taking into account the multichannel aspect of the data should provide

more accurate estimations of the statistics of the noise-free emission since observations at different frequency bands are usually affected by independent noise processes. Note that very recently, and in parallel to this work, a significant step has been taken in this direction by Delouis et al. (2022) employing WST statistics. In that paper, the authors extended the WST to the sphere and introduced cross-WST statistics to characterize correlations across observables in order to statistically denoise Planck all-sky maps of the dust emission.

We thank Michael O’Brien and Blakesley Burkhart for their help in investigating a bispectrum validation of the models of this paper. F.B. acknowledges support from the Agence Nationale de la Recherche (project BxB: ANR-17-CE31-0022).

*Software:* NumPy (Harris et al. 2020), PyTorch (Paszke et al. 2019), PyWPH (Regaldo-Saint Blancard et al. 2021), QuantimPy (Boelens & Tchelepi 2021), SciPy (Virtanen et al. 2020).

## Appendix A

### Maximum Entropy Microcanonical Models

In this appendix, we define maximum entropy microcanonical models, which underlie the microcanonical gradient descent models introduced in Section 3.1. This presentation is based on Bruna & Mallat (2019), and we refer to this paper for additional details.

Microcanonical models 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 the largest entropy (in the sense of information theory).

Let us consider a random field  $X$ , with  $x$  one of its realizations. We want to approximate the distribution of  $X$  based on this single realization  $x$ . In practice, we make some statistical measurements on  $x$  that define a vector of statistics  $\phi(x)$ . We believe  $\phi(x)$  to be sufficiently “exhaustive” to describe the statistical properties of  $X$ .<sup>21</sup> Typically, the realization  $x$  lives in  $\mathbb{R}^m$ , while  $\phi(x)$  lives in  $\mathbb{R}^n$  with  $n < m$ . We introduce *microcanonical sets* as ensembles of vectors of  $\mathbb{R}^m$  whose statistics are “sufficiently close” to those of  $x$ . Formally, for a given  $\epsilon > 0$ , we define the microcanonical set  $\Omega_\epsilon$  as

$$\Omega_\epsilon = \{y \in \mathbb{R}^m \mid |\phi(y) - \phi(x)| \leq \epsilon\}, \quad (\text{A1})$$

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

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

$$H(\mu) = -\int f_\mu(y) \log f_\mu(y) dy, \quad (\text{A2})$$

where  $f_\mu$  is the probability density function associated with  $\mu$ . Assuming that the function  $\phi$  allows  $\Omega_\epsilon$  to be compact, this maximum entropy distribution  $\mu_\epsilon$  is simply the uniform distribution on  $\Omega_\epsilon$  and is defined by its uniform density:

$$f_{\mu_\epsilon}(y) = \frac{1_{\Omega_\epsilon}(y)}{\int_{\Omega_\epsilon} dy}. \quad (\text{A3})$$

<sup>21</sup> For the statistics of a realization  $x$  to be representative of the statistics of  $X$ , we need to make an additional assumption of ergodicity (see Bruna & Mallat 2019).

The relevance of this kind of model directly depends on our choice of statistical measurements, represented by the function  $\phi$ , as well as on the value of  $\epsilon$ , which is a proxy of the volume of the microcanonical set  $\Omega_\epsilon$ . Ideally, we want to choose  $\phi$  and  $\epsilon$  so that typical samples of  $X$  are contained in  $\Omega_\epsilon$  and, conversely, typical samples of  $\mu_\epsilon$  are representative of those of  $X$ .

Even if we manage to define relevant  $\phi$  and  $\epsilon$  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  $\mu_\epsilon$ . Usual strategies make use of Markov Chain Monte Carlo algorithms; however, these algorithms reach computational limits when the dimension of the samples  $m$  increases.<sup>22</sup> For the applications of this paper, these are not an option. Microcanonical gradient descent models as defined in Section 3.1 allow us to circumvent this sampling problem, although these are no longer of maximum entropy in general.

## Appendix B

### Additional Details on the WPH Statistics

#### B.1. Filters

*Bump-steerable mother wavelet.*—The mother bump-steerable wavelet is defined in Fourier space as follows:

$$\hat{\psi}(\mathbf{k}) = \exp\left(\frac{-(k - k_0)^2}{k_0^2 - (k - k_0)^2}\right) \cdot 1_{[0, 2k_0]}(k) \\ \times \cos^{L-1}(\arg(\mathbf{k})) \cdot 1_{[0, \pi/2]}(|\arg(\mathbf{k})|), \quad (\text{B1})$$

with  $k = |\mathbf{k}|$ ,  $1_A(x)$  the indicator function that returns 1 if  $x \in A$  and 0 otherwise, and  $k_0 = 0.85\pi$  the central wavenumber of the mother wavelet. In this paper, we work with  $512 \times 512$  maps and  $L = 4$ . We show in Figure 8 one example wavelet from the resulting bank in both physical and Fourier space.

*Gaussian filter.*—The Gaussian filters used in this paper are dilated versions of a Gaussian function  $\varphi$ , which is defined in Fourier space by the following:

$$\hat{\varphi}(\mathbf{k}) = \exp\left(-\frac{\|\mathbf{k}\|^2}{2\sigma^2}\right), \quad (\text{B2})$$

with  $\sigma = 0.496 \times 2^{-0.55} k_0$  (following Zhang & Mallat 2021).

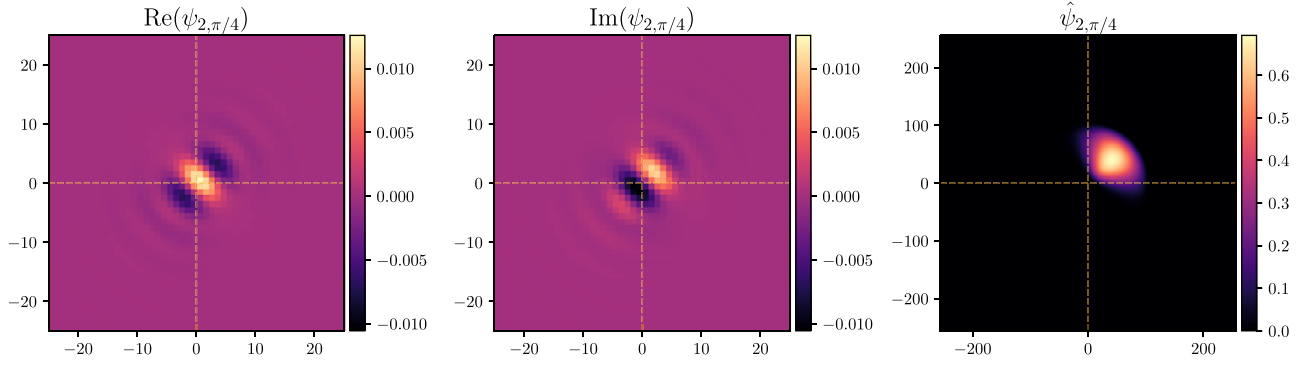
#### B.2. Properties of the WPH Moments

The auto-WPH moments are able to capture interactions between different scales of  $X$  thanks to the phase harmonic operator. Indeed, the covariance between  $X * \psi_\lambda$  and  $X * \psi_{\lambda'}$  vanishes when the wavelets  $\psi_\lambda$  and  $\psi_{\lambda'}$  have nonintersecting bandpasses, and it is otherwise a function of the power spectrum of  $X$  and of the bandpasses of the wavelets (Allys et al. 2020; Zhang & Mallat 2021). This is a consequence of the following relation:

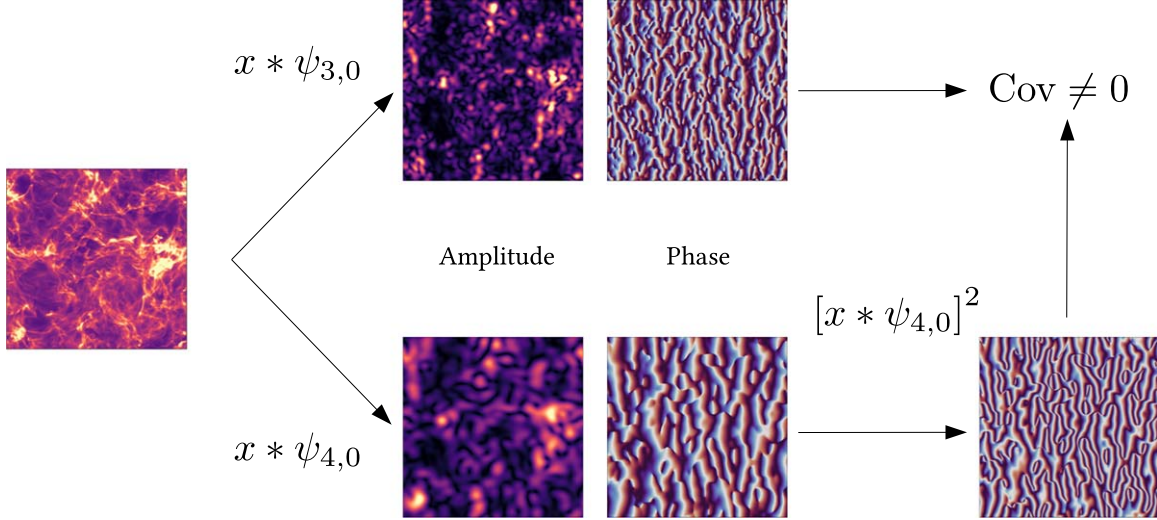
$$C_{\lambda, \lambda', 1}(\boldsymbol{\tau}) = \int S_X(\mathbf{k}) \hat{\psi}_\lambda(\mathbf{k}) \hat{\psi}_{\lambda'}^*(\mathbf{k}) e^{-i\mathbf{k} \cdot \boldsymbol{\tau}} d\mathbf{k}, \quad (\text{B3})$$

with  $S_X$  the power spectrum of  $X$  (for a proof, see Zhang & Mallat 2021). With proper  $p$  and  $p'$  values, the phase harmonic operator can make  $[X * \psi_\lambda]^p$  and  $[X * \psi_{\lambda'}]^{p'}$  comparable in

<sup>22</sup> In general, Markov chain mixing time depends on the exponential of  $m$  (Levin & Peres 2017).



**Figure 8.** Bump-steerable wavelet  $\psi_{2,\pi/4}$  shown in physical space (real part on the left and imaginary part in the middle) and Fourier space (right).



**Figure 9.** Comparison of the amplitude and phase maps of  $x * \psi_{2,0}$  and  $x * \psi_{4,0}$ , where  $x$  is the map shown on the left. The action of the phase harmonic operator is shown on the phase map of  $x * \psi_{4,0}$ . Figure inspired by Allys et al. (2020).

the sense that they share common spatial frequencies, allowing an extraction of high-order information through their covariance. Conveniently, this operator is a Lipschitz continuous operator, that is we have for all  $z, z' \in \mathbb{C}^2$ ,  $||[z]^p - [z']^p|| \leq \max(|p|, 1)|z - z'|$  (Mallat et al. 2020). 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 (Zhang & Mallat 2021).

To illustrate the importance of the phase harmonic operator to measure phase alignment between scales, we show in Figure 9 how the amplitude and phase maps of  $x * \psi_\lambda$  and  $x * \psi_{\lambda'}$  compare for two different oriented scales  $\lambda$  and  $\lambda'$ , with  $x$  the  $I_{300}$  map built in Section 2.3. We choose  $\lambda = (3, 0)$  and  $\lambda' = (4, 0)$ , so that  $\psi_\lambda$  and  $\psi_{\lambda'}$  probe similar orientations but different scales, with  $\psi_{\lambda'}$  probing scales that are twice as large as those probed by  $\psi_\lambda$ . The amplitude maps show local variations of the signal  $x$  filtered at different scales, with naturally coarser variations in  $|x * \psi_{\lambda'}|$  than in  $|x * \psi_\lambda|$ . 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  $\arg(x * \psi_\lambda)$  is approximately twice that of  $\arg(x * \psi_{\lambda'})$ . The phase harmonic operator with  $k=2$  transforms  $\arg(x * \psi_{\lambda'})$  into a phase map (bottom right map) that is much more coherent with  $\arg(x * \psi_\lambda)$ . Consequently, the sample

covariance of these maps does not vanish and quantifies the phase alignment between  $x * \psi_\lambda$  and  $x * \psi_{\lambda'}$ .

### B.3. Subset of Auto-WPH Moments

Allys et al. (2020) identified a relevant set of auto-WPH moments to build models of simulated data of the large-scale structure of the universe. In the present work, we define WPH statistics that are directly inspired by this work, although they are made slightly more optimal by taking out coefficients that were redundant. We will consider five categories of moments defined as follows:

1. The  $S^{(1,1)}$  moments, of the form  $C_{\lambda,1,\lambda,1}(\tau) = \text{Cov}(X * \psi_\lambda(\mathbf{r}), X * \psi_\lambda(\mathbf{r} + \tau))$ , at every  $\tau_{n,\alpha}$ . They measure weighted averages of the power spectrum over the bandpass of  $\psi_\lambda$  (see Equation (B3)).
2. The  $S^{(0,0)}$  moments, of the form  $C_{\lambda,0,\lambda,0}(\tau) = \text{Cov}(|X * \psi_\lambda(\mathbf{r})|, |X * \psi_\lambda(\mathbf{r} + \tau)|)$ , at every  $\tau_{n,\alpha}$ . They capture information related to the sparsity of the data in the bandpass of  $\psi_\lambda$ .
3. The  $S^{(0,1)}$  moments, of the form  $C_{\lambda,0,\lambda,1}(\tau) = \text{Cov}(|X * \psi_\lambda(\mathbf{r})|, X * \psi_\lambda(\mathbf{r} + \tau))$ , at  $\tau=0$  only. They capture information related to the couplings between the scales included in the same bandpass.
4. The  $C^{(0,1)}$  moments, of the form  $C_{\lambda,0,\lambda',1}(\tau) = \text{Cov}(|X * \psi_\lambda(\mathbf{r})|, X * \psi_{\lambda'}(\mathbf{r} + \tau))$ , considering  $0 \leq j < j' \leq J - 1$ , at every  $\tau_{n,\alpha}$  when  $\theta = \theta'$  and at  $\tau=0$  only



**Table 2**  
Number of Statistical Coefficients per Class of Moments for the Parameters Used in This Work

	$S^{(1,1)}$	$S^{(0,0)}$	$S^{(0,1)}$	$S^{(1,0)}$	$C^{(0,1)}$	$C^{(1,0)}$	$C^{\text{phase}}$	$C^{\text{phase,inv}}$	$L$	Total	Ratio (%)
Auto	544	544	32	N/A	4032	N/A	1776	N/A	12	6940	~2.6
Cross	32	32	32	32	448	448	112	112	16	1264	~0.24

**Note.** The last column gives the ratio of the total number of coefficients to the number of pixels in one (or two, for the cross case)  $512 \times 512$  image(s).

when  $\theta \neq \theta'$ . They capture information related to the correlation between local levels of oscillation for the scales in the bandpasses associated with  $\psi_\lambda$  and  $\psi_{\lambda'}$ .

- The  $C^{\text{phase}}$  moments, of the form  $C_{\lambda,1,\lambda,p'}(\boldsymbol{\tau}) = \text{Cov}[X * \psi_\lambda(\mathbf{r}), [X * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]^{p'}]$  with  $p' = 2^{j'-j} > 1$ , considering  $0 \leq j < j' \leq J-1$  and  $\theta = \theta'$ , at every  $\boldsymbol{\tau}_{n,\alpha}$ . They capture information related to the statistical phase alignment of oscillations between the scales in the bandpasses associated with  $\psi_\lambda$  and  $\psi_{\lambda'}$ .

*Symmetries in the subset of moments.*—We identify in the following 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  $C_{\lambda,1,\lambda,1}(\boldsymbol{\tau}_{n,\alpha})$  with  $\lambda = (j, \theta)$ , and we 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 the case of bump-steerable wavelets), we identify the following list of symmetries:

$$S^{(1,1)}(j, \theta + \pi, n, \alpha) = S^{(1,1)}(j, \theta, n, \alpha), \quad (\text{B4})$$

$$S^{(1,1)}(j, \theta, n, \alpha + \pi) = \overline{S^{(1,1)}(j, \theta, n, \alpha)}, \quad (\text{B5})$$

$$S^{(0,0)}(j, \theta + \pi, n, \alpha) = S^{(0,0)}(j, \theta, n, \alpha), \quad (\text{B6})$$

$$S^{(0,0)}(j, \theta, n, \alpha + \pi) = \overline{S^{(0,0)}(j, \theta, n, \alpha)}, \quad (\text{B7})$$

$$S^{(0,1)}(j, \theta + \pi) = \overline{S^{(0,1)}(j, \theta)}, \quad (\text{B8})$$

$$C^{(0,1)}(j_1, \theta_1 + \pi, j_2, \theta_2, n, \alpha) = C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha), \quad (\text{B9})$$

$$C^{(0,1)}(j_1, \theta_1, j_2, \theta_2 + \pi, n, \alpha) = \overline{C^{(0,1)}(j_1, \theta_1, j_2, \theta_2, n, \alpha)}, \quad (\text{B10})$$

$$\begin{aligned} C^{\text{phase}}(j_1, \theta + \pi, j_2, \theta + \pi, n, \alpha) \\ = \overline{C^{\text{phase}}(j_1, \theta, j_2, \theta, n, \alpha)}. \end{aligned} \quad (\text{B11})$$

Proofs of these relations essentially stem from the facts that  $\text{Cov}(X, Y) = \overline{\text{Cov}(Y, X)}$  and that  $[X * \psi_{j,\theta+\pi}]^p = \overline{[X * \psi_{j,\theta}]^p}$  for any  $p \in \mathbb{R}$ . The latter relation is a consequence of the commutativity of the complex conjugation with the phase harmonic and convolution operations.

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, it is sufficient to consider wavelets with  $\theta \in [0, \pi)$ . Moreover, to avoid further redundancy, for the  $S^{(1,1)}$  and  $S^{(0,0)}$  moments, it is sufficient to consider  $\boldsymbol{\tau}_{n,\alpha}$  vectors with  $\alpha \in [0, \pi)$  only. We show in Table 2 the resulting number of auto coefficients per class of moments for the parameters used in this work.

#### B.4. Subset of Cross-WPH Moments

Just like before, we focus on a specific subset of cross-WPH moments. We make sure that this subset is nonredundant and *pseudo*-symmetric under the exchange of  $X$  and  $Y$  (modulo complex conjugation). These moments are the following:

- The  $S_\times^{(1,1)}$  moments, of the form  $C_{\lambda,1,\lambda,1}^\times(\boldsymbol{\tau}) = \text{Cov}[X * \psi_\lambda(\mathbf{r}), Y * \psi_\lambda(\mathbf{r} + \boldsymbol{\tau})]$ . Note that they measure weighted averages of the cross spectrum over the bandpass of  $\psi_\lambda$ .
- The  $S_\times^{(0,0)}$  moments, of the form  $C_{\lambda,0,\lambda,0}^\times(\boldsymbol{\tau}) = \text{Cov}[|X * \psi_\lambda(\mathbf{r})|, |Y * \psi_\lambda(\mathbf{r} + \boldsymbol{\tau})|]$ .
- The  $S_\times^{(0,1)}$  moments, of the form  $C_{\lambda,0,\lambda,1}^\times(\boldsymbol{\tau}) = \text{Cov}[|X * \psi_\lambda(\mathbf{r})|, Y * \psi_\lambda(\mathbf{r} + \boldsymbol{\tau})]$ .
- The  $S_\times^{(1,0)}$  moments, of the form  $C_{\lambda,1,\lambda,0}^\times(\boldsymbol{\tau}) = \text{Cov}[X * \psi_\lambda(\mathbf{r}), |Y * \psi_\lambda(\mathbf{r} + \boldsymbol{\tau})|]$ .
- The  $C_\times^{(0,1)}$  moments, of the form  $C_{\lambda,0,\lambda,1}^\times(\boldsymbol{\tau}) = \text{Cov}[|X * \psi_\lambda(\mathbf{r})|, Y * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]$ , considering  $0 \leq j < j' \leq J-1$ .
- The  $C_\times^{(1,0)}$  moments, of the form  $C_{\lambda,0,\lambda,1}^\times(\boldsymbol{\tau}) = \text{Cov}[X * \psi_\lambda(\mathbf{r}), |Y * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})|]$ , considering  $0 \leq j' < j \leq J-1$ .
- The  $C_\times^{\text{phase}}$  moments, of the form  $C_{\lambda,1,\lambda,p'}^\times(\boldsymbol{\tau}) = \text{Cov}[X * \psi_\lambda(\mathbf{r}), [Y * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]^{p'}]$  with  $p' = 2^{j'-j} > 1$ , considering  $0 \leq j < j' \leq J-1$  and  $\theta = \theta'$ .
- The  $C_\times^{\text{inv,phase}}$  moments, of the form  $C_{\lambda,p,\lambda,1}^\times(\boldsymbol{\tau}) = \text{Cov}[|X * \psi_\lambda(\mathbf{r})|^p, Y * \psi_{\lambda'}(\mathbf{r} + \boldsymbol{\tau})]$  with  $p = 2^{j'-j} > 1$ , considering  $0 \leq j' < j \leq J-1$  and  $\theta = \theta'$ .

Moreover, for this work, inspired by Brochard et al. (2022), we only consider cross-WPH moments with  $\boldsymbol{\tau} = \mathbf{0}$ . We show in Table 2 the resulting number of cross coefficients per class of moments for the parameters used in this work.

#### B.5. Normalized Estimates

In practice, our statistical coefficients are all normalized similarly to Zhang & Mallat (2021) and Allys et al. (2020). This normalization has been shown to speed up the minimization involved during the sampling process of microcanonical gradient descent models (Zhang & Mallat 2021).

The normalization of the auto-WPH and scaling coefficients depends on the target map  $x_0$  involved in this minimization. Denoting by  $\tilde{C}_{\lambda,p,\lambda,p'}(\boldsymbol{\tau})$  and  $\tilde{L}_{j,p,p'}$  the normalized estimates of  $C_{\lambda,p,\lambda,p'}(\boldsymbol{\tau})$  and  $L_{j,p,p'}$ , respectively, we choose

$$\tilde{C}_{\lambda,p,\lambda,p'}(\boldsymbol{\tau}) = \frac{\langle x^{(\lambda,p)}(\mathbf{r}) \overline{x^{(\lambda,p')}(\mathbf{r} + \boldsymbol{\tau})} \rangle}{\sqrt{\langle |x_0^{(\lambda,p)}|^2 \rangle \langle |x_0^{(\lambda,p')}|^2 \rangle}}, \quad (\text{B12})$$

$$\tilde{L}_{j,p,p'} = \frac{\langle x^{(j,p)} x^{(j,p')} \rangle}{\sqrt{\langle |x_0^{(j,p)}|^2 \rangle \langle |x_0^{(j,p')}|^2 \rangle}}, \quad (\text{B13})$$

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

For cross coefficients, we define similarly

$$\tilde{C}_{\lambda,p,\lambda',p'}^\times(\boldsymbol{\tau}) = \frac{\langle x^{(\lambda,p)}(\mathbf{r}) \overline{y^{(\lambda',p')}}(\mathbf{r} + \boldsymbol{\tau}) \rangle}{\sqrt{\langle |x_0^{(\lambda,p)}|^2 \rangle \langle |y_0^{(\lambda',p')}} \rangle}}, \quad (\text{B14})$$

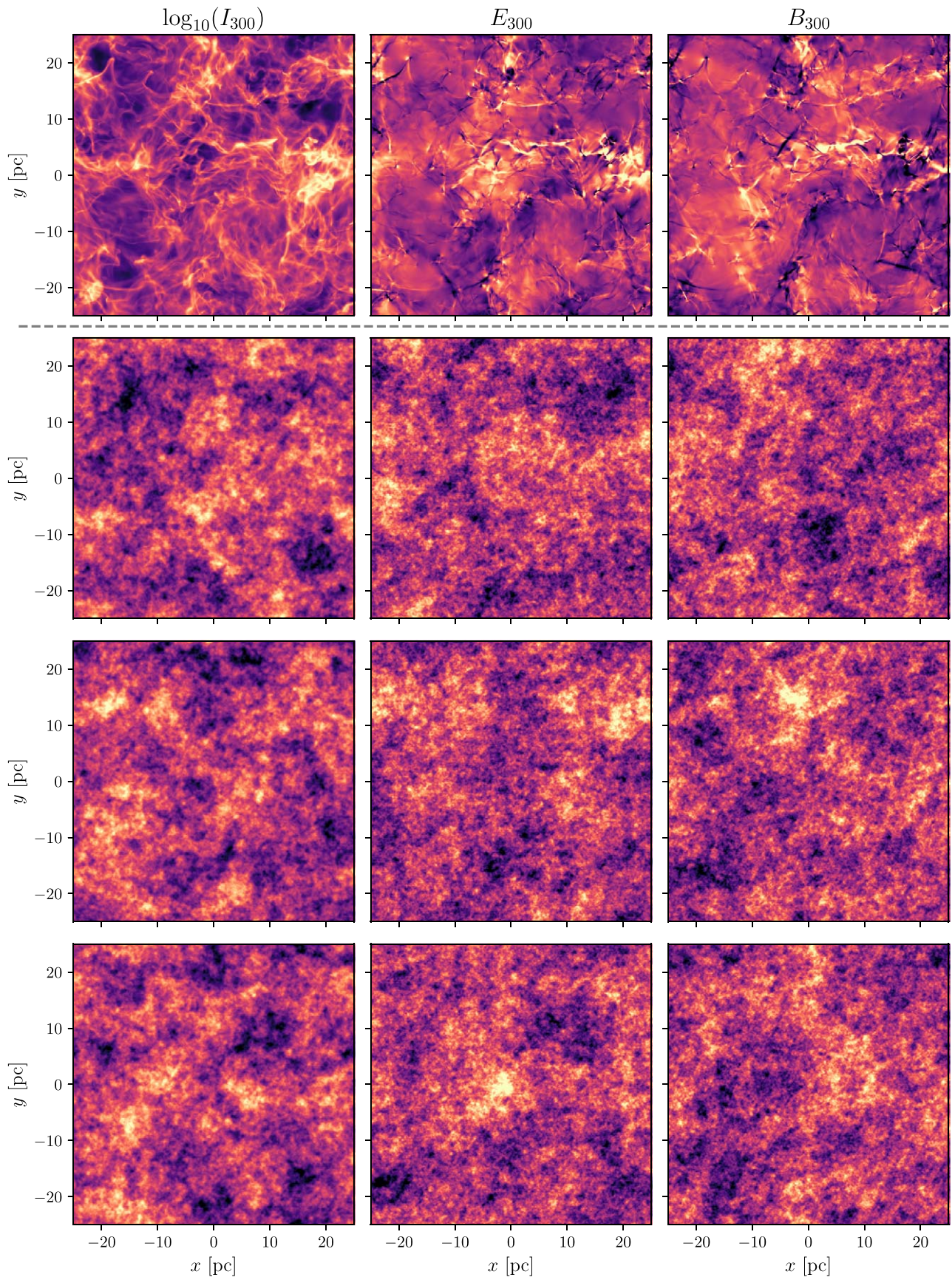
$$\tilde{L}_{j,p,p'}^\times = \frac{\langle \mathbf{x}^{(j,p)} \mathbf{y}^{(j,p')} \rangle}{\sqrt{\langle |x_0^{(j,p)}|^2 \rangle \langle |y_0^{(j,p')}|^2 \rangle}}. \quad (\text{B15})$$

## Appendix C Gaussian Model

The Gaussian model of the data  $\mathbf{x} = (\log(I_{300}), E_{300}, B_{300})$ , which is used as a baseline in Section 4.1, is defined as follows. It is a microcanonical model conditioned by

$$\begin{aligned} \phi^G(\mathbf{x}) = & \phi_{\text{auto}}^G(\log(I_{300})) \oplus \phi_{\text{auto}}^G(E_{300}) \oplus \phi_{\text{auto}}^G(B_{300}) \\ & \oplus \phi_{\text{cross}}^G(\log(I_{300}), E_{300}) \oplus \phi_{\text{cross}}^G(\log(I_{300}), B_{300}), \end{aligned} \quad (\text{C1})$$

where  $\phi_{\text{auto}}^G(\cdot)$  and  $\phi_{\text{cross}}^G(\cdot, \cdot)$  only include normalized estimates of the  $S^{(1,1)}$  and  $S_\times^{(1,1)}$  moments as introduced in Appendices B.3 and B.4, respectively. Such microcanonical models are rigorously Gaussian (Bruna & Mallat 2019). We show in Figure 10 three different samples of this model (from second to last row) next to the original data  $\mathbf{x}$  (first row).



**Figure 10.** Same as Figure 2, but for the Gaussian model.

## ORCID iDs

Bruno Régaldo-Saint Blancard  <https://orcid.org/0000-0003-0055-0953>

François Boulanger  <https://orcid.org/0000-0003-1097-6042>

François Levrier  <https://orcid.org/0000-0002-3065-9944>

Léo Vacher  <https://orcid.org/0000-0001-9551-1417>

## References

- Abazajian, K., Addison, G., Adshead, P., et al. 2019, arXiv:1907.04473
- Allys, E., Levrier, F., Zhang, S., et al. 2019, *A&A*, **629**, A115
- Allys, E., Marchand, T., Cardoso, J. F., et al. 2020, *PhRvD*, **102**, 103506
- Aylor, K., Haq, M., Knox, L., Hezaveh, Y., & Perreault-Levasseur, L. 2021, *MNRAS*, **500**, 3889
- Bellomi, E., Godard, B., Hennebelle, P., et al. 2020, *A&A*, **643**, A36
- BICEP2/Keck Array, & Planck Collaborations 2015, *PhRvL*, **114**, 101301
- Boelens, A. M. P., & Tchepeli, H. A. 2021, *SoftX*, **16**, 100823
- Brochard, A., Zhang, S., & Mallat, S. 2022, arXiv:2203.07902
- Bruna, J., & Mallat, S. 2019, arXiv:1801.02013
- Bruna, J., & Mallat, S. 2013, *ITPAM*, **35**, 1872
- Burkhart, B. 2021, *PASP*, **133**, 102001
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. 1995, *SIJSC*, **16**, 1190
- Chluba, J., Hill, J. C., & Abitbol, M. H. 2017, *MNRAS*, **472**, 1195
- Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., & Matsubara, T. 2013, *MNRAS*, **435**, 531
- Delabrouille, J., Betoule, M., Melin, J. B., et al. 2013, *A&A*, **553**, A96
- Delouis, J. M., Allys, E., Gauvrit, E., & Boulanger, F. 2022, *A&A*, **668**, A122
- Fraisse, A. A., Ade, P. A. R., Amiri, M., et al. 2013, *JCAP*, **2013**, 047
- Fromang, S., Hennebelle, P., & Teyssier, R. 2006, *A&A*, **457**, 371
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hensley, B. S., & Bull, P. 2018, *ApJ*, **853**, 127
- Hervias-Caimapo, C., & Huffenberger, K. M. 2022, *ApJ*, **928**, 65
- Ichiki, K., Kanai, H., Katayama, N., & Komatsu, E. 2019, *PTEP*, **2019**, 033E01
- Ishino, H., Akiba, Y., Arnold, K., et al. 2016, *Proc. SPIE*, **9904**, 361
- Jeffrey, N., Boulanger, F., Wandelt, B. D., et al. 2021, *MNRAS: Letters*, **510**, L1
- Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, *PhRvL*, **78**, 2058
- Kamionkowski, M., & Kovetz, E. D. 2016, *ARA&A*, **54**, 227
- Krachmalnicoff, N., & Puglisi, G. 2021, *ApJ*, **911**, 42
- Levin, D., & Peres, Y. 2017, *Markov Chains and Mixing Times* (2nd ed.; Providence, RI: American Mathematical Society)
- Levrier, F., Neveu, J., Falgarone, E., et al. 2018, *A&A*, **614**, A124
- Mallat, S. 2012, *Commun. Pure Appl. Math.*, **65**, 1331
- Mallat, S. 2016, *RSPTA*, **374**, 20150203
- Mallat, S., Zhang, S., & Rochette, G. 2020, *Inf. Inference*, **9**, 721
- Mangilli, A., Aumont, J., Rotti, A., et al. 2021, *A&A*, **647**, A52
- McBride, L., Bull, P., & Hensley, B. S. 2022, *MNRAS*, in press
- Naess, S., Hasselfield, M., McMahon, J., et al. 2014, *JCAP*, **2014**, 007
- Parroni, C., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, *A&A*, **633**, A71
- Paszke, A., Gross, S., Massa, F., et al. 2019, in 33rd Conf. on Neural Information Processing Systems (NeurIPS 2019), ed. H. Wallach & H. Larochelle (Red Hook, NY: Curran Associates, Inc.), <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>
- Pelgrims, V., Ferrière, K., Boulanger, F., Lallement, R., & Montier, L. 2020, *A&A*, **636**, A17
- Planck Collaboration I 2020, *A&A*, **641**, A1
- Planck Collaboration IV 2020, *A&A*, **641**, A4
- Planck Collaboration VII 2020, *A&A*, **641**, A7
- Planck Collaboration X 2016, *A&A*, **594**, A10
- Planck Collaboration XI 2014, *A&A*, **571**, A11
- Planck Collaboration XI 2020, *A&A*, **641**, A11
- Planck Collaboration Int. XX 2015, *A&A*, **576**, A105
- Planck Collaboration Int. XXIII 2015, *A&A*, **580**, A13
- Planck Collaboration Int. XLVIII 2016, *A&A*, **596**, A109
- Régaldo-Saint Blancard, B. 2021, PhD thesis, Ecole Normale Supérieure; Observatoire de Paris, <https://tel.archives-ouvertes.fr/tel-03600846>
- Régaldo-Saint Blancard, B. 2022, Data and Code Associated with “Generative Models of Multi-channel Data Based on a Single Example—Application to Dust Emission”, v1.0, Zenodo, doi:10.5281/zenodo.7342682
- Régaldo-Saint Blancard, B., Allys, E., Boulanger, F., et al. 2021, *A&A*, **649**, L18
- Régaldo-Saint Blancard, B., Levrier, F., Allys, E., et al. 2020, *A&A*, **642**, A217
- Reissl, S., Guillet, V., Brauer, R., et al. 2020, *A&A*, **640**, A118
- Simons Observatory Collaboration 2019, *JCAP*, **2019**, 056
- Skalidis, R., & Pelgrims, V. 2019, *A&A*, **631**, L11
- Teyssier, R. 2002, *A&A*, **385**, 337
- Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2017, *MNRAS*, **469**, 2821
- Thorne, B., Knox, L., & Prabhu, K. 2021, *MNRAS*, **504**, 2603
- Vacher, J., & Briand, T. 2021, *Image Process. Line*, **11**, 170
- Vacher, L., Chluba, J., Aumont, J., Rotti, A., & Montier, L. 2023, *A&A*, **669**, A5
- Vansyngel, F., Boulanger, F., Ghosh, T., et al. 2017, *A&A*, **603**, A62
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Zaldarriaga, M. 2001, *PhRvD*, **64**, 103001
- Zhang, S., & Mallat, S. 2021, *Appl. Comput. Harmon. Anal.*, **53**, 199
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. 1997, *ACM Trans. Math. Softw.*, **23**, 550