



# Single frequency CMB B-mode inference with realistic foregrounds from a single training image

Niall Jeffrey<sup>1,2</sup>★ François Boulanger,<sup>1</sup> Benjamin D. Wandelt,<sup>3,4</sup> Bruno Regaldo-Saint Blancard,<sup>1,5</sup> Erwan Ally<sup>1</sup> and François Levrier<sup>1</sup>

<sup>1</sup>*Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France*

<sup>2</sup>*Department of Physics & Astronomy, University College London, Gower Street, London, UK*

<sup>3</sup>*Institut d'Astrophysique de Paris (IAP), Sorbonne Université, UMR F-7095, CNRS, France*

<sup>4</sup>*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY, USA*

<sup>5</sup>*Observatoire de Paris, PSL University, Sorbonne Université, LERMA, F-75014 Paris, France*

Accepted 2021 October 25. Received 2021 October 25; in original form 2021 September 22

## ABSTRACT

With a single training image and using wavelet phase harmonic augmentation, we present polarized Cosmic Microwave Background (CMB) foreground marginalization in a high-dimensional *likelihood-free* (Bayesian) framework. We demonstrate robust foreground removal using only a single frequency of simulated data for a BICEP-like sky patch. Using Moment Networks, we estimate the pixel-level posterior probability for the underlying  $\{E, B\}$  signal and validate the statistical model with a quantile-type test using the estimated marginal posterior moments. The Moment Networks use a hierarchy of U-Net convolutional neural networks. This work validates such an approach in the most difficult limiting case: pixel-level, noise-free, highly non-Gaussian dust foregrounds with a single training image at a single frequency. For a real CMB experiment, a small number of representative sky patches would provide the training data required for full cosmological inference. These results enable robust *likelihood-free*, simulation-based parameter, and model inference for primordial B-mode detection using observed CMB polarization data.

**Key words:** methods: statistical – cosmology: cosmic background radiation.

## 1 INTRODUCTION

Polarized galactic foregrounds are the most significant challenge for the detection of the polarized cosmic microwave background (CMB) B-mode signal of primordial gravitational waves. Such a detection would provide a direct probe of cosmic inflation in the early Universe.

Typically, the cosmological signal is recovered using multi-frequency observations, for which there have been numerous recent developments in the context of Planck and ground-based experiments (Planck Collaboration 2020a; Darwish et al. 2021). The sensitivity expected from the next generation of experiments poses further challenges (Ade et al. 2019; CMB-S4 Collaboration 2020).

A significant hindrance is the difficulty either numerically simulating the polarized galactic dust emission or constructing analytic models. In particular, if it were possible to generate synthetic realizations of the dust polarization signal, the task could be remarkably simplified. With the ability to forward model by generating mock data, the use of simulation-based methods to perform *likelihood-free inference* of unknown parameters is relatively straightforward.

Generating a significant number of polarized foreground signal realizations has been so far impossible. Numerical simulations are computationally costly and fraught with difficulties. Furthermore, low-noise and high-precision observational data are also limited.

In this letter, we show that with only a single training image of polarized dust emission we can use *wavelet phase harmonic synthesis* to generate new representative samples. These samples form a training set with which we can perform likelihood-free inference. We test the high-dimensional posterior probability space of pixel values directly, which proves to be a powerful general validation of this combined synthesis and likelihood-free approach.

Using only a single frequency (143 GHz), we use a hierarchy of convolutional neural networks to form a Moment Network (Jeffrey & Wandelt 2020) to estimate posterior mean and marginal variance per pixel of the extra-galactic CMB signal given the observed polarization data. *Summary of our approach:*

- (i) Simulate a single polarized foreground data patch.
- (ii) Synthesize fast realizations using wavelet phase harmonics to form a large augmented set of foreground maps.
- (iii) Draw cosmological parameters from a prior probability distribution and generate mock CMB signals.
- (iv) Draw foreground amplitude parameters from a prior probability distribution to match a BICEP-like sky at a single 143 GHz frequency.
- (v) Train Moment Network to estimate the marginal posterior mean and variance for the CMB signal for each pixel.
- (vi) Validate the per-pixel posterior estimates with a quantile-type test using new *unseen* simulated (not synthesized) foreground data.

\* E-mail: [niall.jeffrey@phys.ens.fr](mailto:niall.jeffrey@phys.ens.fr)

By validating the pixel-by-pixel statistical model, we have demonstrated that this approach is robust for parameter inference and, by extension, tensor-to-scalar  $r$  detection.

For application to real data, the inference task will be in many ways easier. Here, we have chosen the set-up to be particularly difficult to demonstrate the strengths of the method: Noise-free, a single training image, and a single frequency. Measurement noise will Gaussianize the data, so would limit our demonstration of highly non-Gaussian dust foreground removal.

Aylor et al. (2020) previously demonstrated an adversarial networks, using  $>10^3$  cleaned (Planck) training images, for total intensity foreground cleaning. Using the results of this letter, we can now use very few representative foreground data patches for polarization (B-mode) analysis as part of our new Bayesian validated likelihood-free approach.

## 2 STATISTICAL MODEL

### 2.1 Likelihood-free inference

#### 2.1.1 Overview:

In Bayesian parameter inference of unknown parameters  $\theta$ , we aim to evaluate the posterior probability distribution

$$p(\theta|\mathbf{d}_{\text{obs}}, \mathcal{M}) = \frac{p(\mathbf{d}_{\text{obs}}|\theta, \mathcal{M}) p(\theta|\mathcal{M})}{p(\mathbf{d}_{\text{obs}}|\mathcal{M})}, \quad (1)$$

for some statistical model  $\mathcal{M}$  given the observed data (or summary statistics of the observed data)  $\mathbf{d}_{\text{obs}}$  (see Jaynes 2003 for details).

In a typical analysis, the likelihood  $\mathcal{L}(\theta) = p(\mathbf{d}_{\text{obs}}|\theta, \mathcal{M})$  is assumed or modelled analytically. In *likelihood-free inference* (also known as simulation-based inference) the likelihood is not assumed. Instead, the sampling distribution of the data  $p(\mathbf{d}|\theta, \mathcal{M})$  as a function of the unknown parameters is estimated from forward modelled data.

With density estimation likelihood-free inference (Papamakarios & Murray 2016; Alsing, Wandelt & Feeney 2018; Alsing et al. 2019), the inference task is posed as a density estimation problem. The simulated mock data  $\mathbf{d}$  realizations and their respective parameters  $\theta$  form a cloud of points in  $\{\mathbf{d}, \theta\}$  space. In this space, we could estimate the following distributions: The joint  $p(\mathbf{d}, \theta, \mathcal{M})$ , the conditional  $p(\theta|\mathbf{d}, \mathcal{M})$ , or the conditional  $p(\mathbf{d}|\theta, \mathcal{M})$  known as the sampling distribution (which becomes likelihood if evaluated for observed data  $\mathbf{d}_{\text{obs}}$ ).

Provided mock data  $\mathbf{d}$  can be generated, with each realization having an associated  $\theta$  label, the problem of inference given the actual observed data is relatively straightforward. Using this density estimation approach, a number of cosmological analysis have now been carried out (e.g. Brehmer et al. 2019; Kodi Ramanah et al. 2020; Jeffrey, Alsing & Lanusse 2021; Lemos et al. 2021) that infer model parameters  $\theta$  without needing to assume or approximate a closed-form likelihood.

In this work, thanks to wavelet phase harmonic synthesis (Section 2.3.1), we now have the means to generate fast realizations of dust foregrounds for polarization CMB analysis (given our single training image). However, rather than using likelihood-free inference to infer only a set of cosmological parameters, here we choose to demonstrate a pixel-by-pixel inference at the level of the map as a powerful test of this approach.

#### 2.1.2 Moment Networks:

These are a simple hierarchy of fast neural regression models that compute increasing moments of any lower-dimensional marginal posterior density (Jeffrey & Wandelt 2020).

For a pixel-level inference, each unknown pixel value of the underlying CMB signal  $s$  is a parameter to be inferred. The probability space of all pixels in a given  $B$  signal map  $s_B$  has a dimension of  $\mathcal{D} = 256 \times 256 > 6.5 \times 10^4$ . Evaluation of the full joint posterior probability at such a dimensionality  $\mathcal{D}$  is intractable.

Moment Networks side-step the problem of estimating the posterior density. Particularly useful for high-dimensional parameter spaces, they directly estimate moments of the marginal posterior distribution of single parameters or subsets of parameters. For example, we could estimate the mean and variance of the marginal posterior probability for each pixel of some unknown signal  $\theta = s$ ,

$$p(s_\alpha|\mathbf{d}_{\text{obs}}, \mathcal{M}) = \int p(s_\alpha, s'|\mathbf{d}_{\text{obs}}, \mathcal{M}) \, ds', \quad (2)$$

where  $s_\alpha$  is any given pixel (element of the signal vector) and  $s'$  are all other pixels. We could also estimate moments (e.g. covariance) of two-dimensional marginal posteriors for pairs of parameters:

$$p(s_\alpha, s_\beta|\mathbf{d}_{\text{obs}}, \mathcal{M}) = \int p(s_\alpha, s_\beta, s'|\mathbf{d}_{\text{obs}}, \mathcal{M}) \, ds'. \quad (3)$$

We require that the unknown parameters are drawn from prior distribution  $s_i \sim p(s|\mathcal{M})$  and, through the forward model, the associated data are drawn from the sampling distribution  $\mathbf{d}_i \sim p(\mathbf{d}|s_i, \mathcal{M})$ . The first layer of the hierarchy in the Moment Network finds some function  $\mathcal{F}(\mathbf{d})$  of our data that minimizes a squared loss over the distribution of possible training examples  $\{\mathbf{d}_i, s_i\}$ ,

$$J_0 = \int \|s - \mathcal{F}(\mathbf{d})\|^2 p(\mathbf{d}, s) \, d\mathbf{d} \, ds, \quad (4)$$

then  $\mathcal{F}$ , which is a neural network, evaluated for the observed data is the mean of the posterior distribution  $\mathcal{F}(\mathbf{d}_{\text{obs}}) = \langle s \rangle_{p(s|\mathbf{d}_{\text{obs}})}$ .

Moment Networks allow the use of far simpler neural network architectures, which reduces training failure risks and improves inference speed, and have recently been applied to Cosmic Void inference (Kreisch et al. 2021).

In this work, at the next level of the hierarchy, the function  $\mathcal{G}$  minimizes

$$J_1 = \int \|(s - \mathcal{F}_{\text{fixed}}(\mathbf{d}))^2 - \mathcal{G}(\mathbf{d})\|^2 p(\mathbf{d}, s) \, d\mathbf{d} \, ds, \quad (5)$$

for fixed, already trained  $\mathcal{F}$ . If  $J_1$  is minimized over the training data (drawn from the correct prior and forward model) then  $\mathcal{G}(\mathbf{d}_{\text{obs}})$  gives the posterior variance for the marginalized posterior of the signal for each pixel. This result is exact and independent of the true underlying posterior or prior distributions (Jaynes 2003; Adler & Öktem 2018).

In this work, we train the first network  $\mathcal{F}$  of the two-layer hierarchy to return  $\mu_B = \langle s_B \rangle_{p(s|\mathbf{d}_{\text{obs}})}$ , which is the mean of the marginal posterior of the signal  $s_B$  for each pixel given the data  $\{\mathbf{d}_E, \mathbf{d}_B\}$ :

$$\mathcal{F}(\{\mathbf{d}_E, \mathbf{d}_B\}) = \int s_B p(s_B|\mathbf{d}_E, \mathbf{d}_B, \mathcal{M}) \, ds_B. \quad (6)$$

The second network  $\mathcal{G}$  gives the variance  $\sigma_B^2 = \langle s_B^2 \rangle_{p(s|\mathbf{d}_{\text{obs}})} - \mu_B^2$  of the marginal posterior distribution for each pixel of  $s_B$  given the data  $\{\mathbf{d}_E, \mathbf{d}_B\}$ :

$$\mathcal{G}(\{\mathbf{d}_E, \mathbf{d}_B\}) = \int (s_B^2 - \mu_B^2) p(s_B|\mathbf{d}_E, \mathbf{d}_B, \mathcal{M}) \, ds_B. \quad (7)$$

In this work, we show the result for  $s_B$  here as it is the most significantly contaminated by polarized dust foregrounds. Both stages of the Moment Network use a UNet convolutional neural network from the DeepMass<sup>1</sup> package (Jeffrey et al. 2020) – see Section 3.

<sup>1</sup><https://github.com/NiallJeffrey/DeepMass>

## 2.2 CMB data model

We select a flat  $\Lambda$ -Cold Dark Matter cosmological model for this analysis. This choice of model is not a requirement for this approach; one could choose any cosmological model  $\mathcal{M}$  provided it is possible to generate mock observables where the unknown cosmological parameters  $\Theta_C$  are drawn from some prior probability distribution  $p(\Theta_C | \mathcal{M})$ .

The parameters used for the CMB signal generation must be drawn from the assumed prior  $\Theta_{C,i} \sim p(\Theta_C | \mathcal{M})$ , so that the Moment Network output matches the expected properties of the posterior probability distribution. In this work, we choose to use the Planck Collaboration (2020b) `base_plikHM_TTTEEE_lowL_lowE2` parameter posterior distribution as our prior.

Our sampled cosmological parameters are: The Hubble parameter  $H_0$ , the baryon density  $\Omega_b$ , the cold dark-matter density  $\Omega_c$ , the Reionization optical depth  $\tau$ , primordial amplitude  $A_s$ , scalar spectral index  $n_s$ . We fix the tensor-to-scalar ratio parameter  $r = 0$ , but this could be easily allowed to vary over a wide prior range for such an analysis. We thin the Markov chain Monte Carlo Planck chains uniformly to sample approximately 3000 sets of parameters  $\Theta_{C,i}$ .

For each set of parameters drawn from the prior, we calculate the theoretical CMB polarization  $EE$  and  $BB$  power spectrum  $C_\ell(\Theta_{C,i})$  using the CAMB software (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012). For each  $C_\ell(\Theta_{C,i})$  sample, we generate a full-sky Gaussian realization of the  $E$  and  $B$  signals with HEALPix (Górski et al. 2005). These Gaussian signal realizations do not include higher order N-point corrections caused by any primordial non-Gaussianity or gravitational lensing, and therefore, would be insufficient for a full delensing analysis, though such signals would not be difficult to simulate if required (Millea, Anderes & Wandelt 2020). From each full-sky HEALPix map (resolution `nside=1024`), we extract 18 non-overlapping patches using gnomonic projection for both  $s_E$  and  $s_B$ . Each patch is  $20^\circ \times 20^\circ$  to match a BICEP-like survey area (BICEP2 Collaboration 2018) with  $256 \times 256$  pixels. Each  $s_E$  and  $s_B$  map is convolved with a Gaussian beam smoothing with the size of a single pixel width ( $\sigma \approx 4.69$  arcmin). In total, we generate 52 224 sets of  $\{s_E, s_B\}$  maps.

In reality, the CMB polarization signal is measured in terms of the  $\{Q, U\}$  Stokes fields, rather than the pseudo-scalar  $\{E, B\}$  fields (Zaldarriaga 2001), which generally introduces artefacts when transforming between these representations for masked or non-periodic data. However, this is already a solved problem in a single frequency, unlike polarization dust foreground marginalization. Again, though this  $E$ - $B$ -leakage effect should be included for analyses with actual observational data, adding this effect here would diminish our clear demonstration of non-Gaussian foreground marginalization.

The left hand of Fig. 1 shows an example  $B$  signal simulation. The centre-left hand shows the same map with simulated foregrounds significantly obscuring the signal.

## 2.3 Foreground modelling

### 2.3.1 Wavelet phase harmonic synthesis:

Using a single simulation of the polarized dust foregrounds  $f_Q^*$ ,  $f_U^*$ , we synthesize many further realizations using wavelet phase harmonic synthesis  $\{f_Q, f_U\}$ . The original simulated foreground image  $f_Q^*, f_U^*$  is of a polarization signal as produced by the thermal

emission of dust in the diffuse ISM. It is built from the same magnetohydrodynamic (MHD) simulation as described in Regaldo-Saint Blancard et al. (2020).

Wavelet phase harmonic statistics are descriptions of signals that have been designed taking advantage of comparisons between convolutional neural networks and non-linear harmonic analysis operations. These statistics form models that can be used to generate new realizations of a given signal.

We can define the wavelet phase harmonic coefficients  $\alpha$  of a foreground map in relation to a wavelet phase harmonic operator, where  $\alpha = \phi(f^*)$ . These statistics are designed to characterize the coherent structures that appear in non-Gaussian random fields, by quantifying the phase alignment between different scales (Mallat, Zhang & Rochette 2020; Zhang & Mallat 2021). For cosmic-web density fields, Ally et al. (2020) demonstrated that these statistics include most of the information captured by various high-order statistics.

For a detailed mathematical description, we refer the reader to Appendix A of Regaldo-Saint Blancard et al. (2021), in which these statistics were measured for noisy dust polarization Planck data. However, for an overall understanding of wavelet phase harmonic statistics, the reader may consider that these statistics emulate similar properties to information capture in a convolutional neural network. Unlike neural networks, these statistics require no training, so it is possible to estimate the phase harmonic coefficients from our single training image  $\alpha^* = \phi(f^*)$ . New synthesized  $f$  maps are generated by minimizing a loss,

$$W_\alpha(f) = \|\phi(f) - \alpha^*\|^2, \quad (8)$$

with  $f$  initialized as Gaussian white noise. Using differentiable  $\phi$  operators, the `pyWPH3` code uses gradient-based optimization.

As we are working in  $Q, U$  space to synthesize, we can apply the operator directly on the complex field (as in Regaldo-Saint Blancard et al. 2021). Here, we extend this to include the individual components, giving a new loss:  $W'_\alpha = W_\alpha(f_Q) + W_\alpha(f_U) + W_\alpha(f_Q + i f_U)$ .

A further development of phase harmonic synthesis in this work is the use of differentiable histograms to match the Kullback–Leibler divergence of the 1-point distributions of the synthesized  $f_Q$  and  $f_U$  fields with the target (see Appendix B – Supporting Materials).

These techniques combine to give a realistic new set of synthesized polarized foregrounds that match the single initial training simulation. The robustness of the synthesized maps for this problem is tested in the following section.

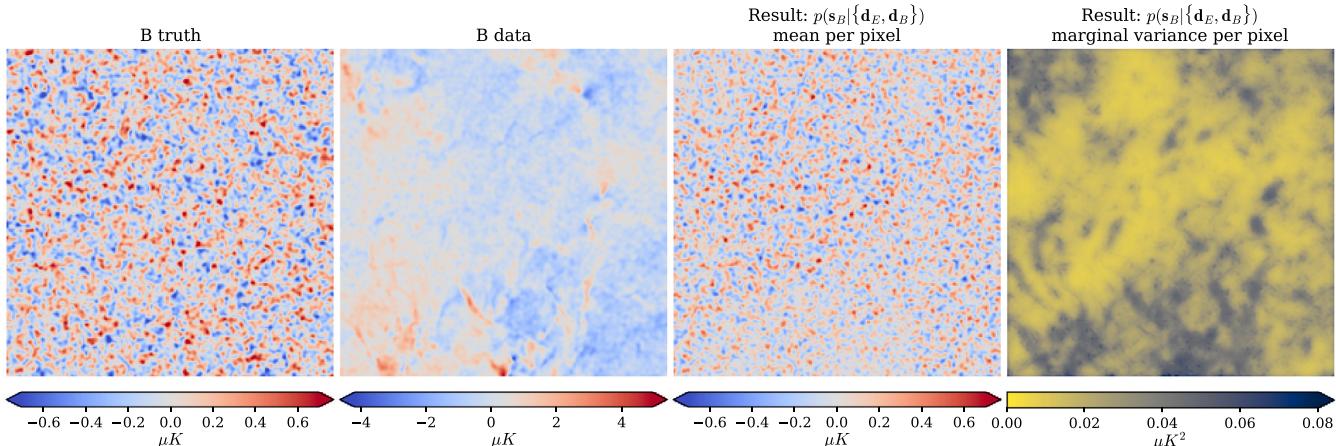
### 2.3.2 Foreground amplitude prior:

We match the synthesized foreground maps, which have  $512 \times 512$  pixels, to a  $40^\circ \times 40^\circ$  area. For each periodic synthesized map, we transform from  $Q, U$  to  $E, B$ . We then select sub-patches of  $256 \times 256$  pixels to break the periodic boundary conditions and to match the survey size of the simulated CMB signal maps. From 102 full  $512 \times 512$  synthesized foreground maps, we use data augmentation methods of image flips, translations, and rotations (applied equally to  $f_E$  and  $f_B$ ) to generate a set of 52 224 foreground maps of size  $256 \times 256$ .

These synthesized foreground maps are matched to realistic foreground properties for a BICEP-like experiment. We model the polarized dust foreground amplitude of the BICEP field using the Planck analysis values and uncertainties. We first assume that the

<sup>2</sup><https://pla.esac.esa.int/>

<sup>3</sup><https://github.com/bregaldo/pywph>



**Figure 1.** The two left hands show the simulated clean signal  $s_B$  and the foreground-contaminated data  $\mathbf{d}_B$  (validation data A – Section 3.3). The centre right hand shows the mean of the marginal posterior probability per pixel  $\mathcal{F}(\mathbf{d}_E, \mathbf{d}_B)$  and the far right shows the variance of the marginal posterior per pixel  $\mathcal{G}(\mathbf{d}_E, \mathbf{d}_B)$ . This CMB signal has been inferred (the posterior probability estimated) using only a single frequency and a single training image. Patches of reduced power in the pixel posterior mean are not artefacts; the mean is expected to move closer to 0  $\mu\text{K}$  when the posterior variance is higher.

large scale foregrounds are known well enough that we can subtract the mean foreground signal so that  $\langle \mathbf{f}_E \rangle = \langle \mathbf{f}_B \rangle = 0$  for each map. The foreground amplitude parameters  $A_E$  and  $A_B$  are defined with respect to a power-law spectrum:

$$C_\ell^{XX} = A_X (\ell/80)^{-0.42} (2\pi) / (\ell (\ell + 1)). \quad (9)$$

The Planck analysis values and uncertainties for each  $A_X$  contribute to our prior  $p(A_E, A_B)$ . As a conservative least-informative approach, we assume independent Gaussians:  $p(A_E) = \mathcal{N}(0.358, 0.0917^2)$  &  $\mathcal{N}(0.172, 0.0440^2)$ , where the model and values are derived from Planck Collaboration (2020c) Section 3.2.

To match these amplitudes defined on the celestial sphere to the pixel variance of our foreground patches, we generate 1000 full-sky Gaussian realizations of foreground maps with power spectrum parameters drawn from the above prior. From these full-sky realizations, we create patches with the correct angular size and resolution (these realizations are then discarded). This Monte Carlo procedure transforms the prior distribution with respect to the power spectrum amplitude parameters  $p(A_E, A_B)$  to a prior on the pixel standard deviation of the patches  $p(\sigma_E, \sigma_B)$ . The synthesized foreground maps  $\{ \mathbf{f}_{E,i}, \mathbf{f}_{B,i} \}$  are rescaled to match the sampled amplitudes.

### 3 RESULTS

#### 3.1 Marginal posterior moments

The previous section described the statistical model that generated mock data samples. For each  $i$  sample, we have: Foreground contaminated data maps  $(\mathbf{d}_{E,i}, \mathbf{d}_{B,i})$ , clean signal maps  $(s_{E,i}, s_{B,i})$ , associated cosmological and foreground parameters  $(A_{E,i}, A_{B,i}, \Theta_{C,i})$ .

In this work, we target the posterior probability distribution of the pixel values of the  $B$ -mode signal given the data  $p(s_B | \mathbf{d}_E, \mathbf{d}_B)$  with all other parameters marginalized away. We choose the  $B$ -mode signal for this demonstration as the signal power is much weaker than  $E$ -mode (by a factor of 300) so is most obscured by polarized dust foregrounds. By validating the posterior probability at pixel-level, this is a powerful demonstration of the robustness of the synthesized forward model and the likelihood-free inference.

For both stages of the Moment Network, the posterior mean and variance networks,  $\mathcal{F}(\mathbf{d}_E, \mathbf{d}_B)$  and  $\mathcal{G}(\mathbf{d}_E, \mathbf{d}_B)$ , we train the UNet

convolutional neural network from DeepMass for 100 epochs with varied batch size and learning rates (Convergence is achieved at epoch 50). From the synthesized training samples, we retain 5120 as a test set to confirm that over-fitting does not occur. Moment Networks are fast and inexpensive, taking 94 s per epoch to train with an Nvidia V100 32GB and evaluation of moments taking  $\approx 1$  ms per new data map.

To validate the inference, we do not use further synthesized images, but we take a new unseen simulation. We use four non-overlapping foreground patches from an approximately-independent snapshot of the same steady-state MHD simulation (see section 3.1 of Regaldo-Saint Blanchard et al. 2020 for details). The same steps of the forward model are applied to these new patches, including drawing the foreground amplitude from a prior for each map. Four new simulated CMB signals are added to generate four new data sets, which are labelled A, B, C, and D.

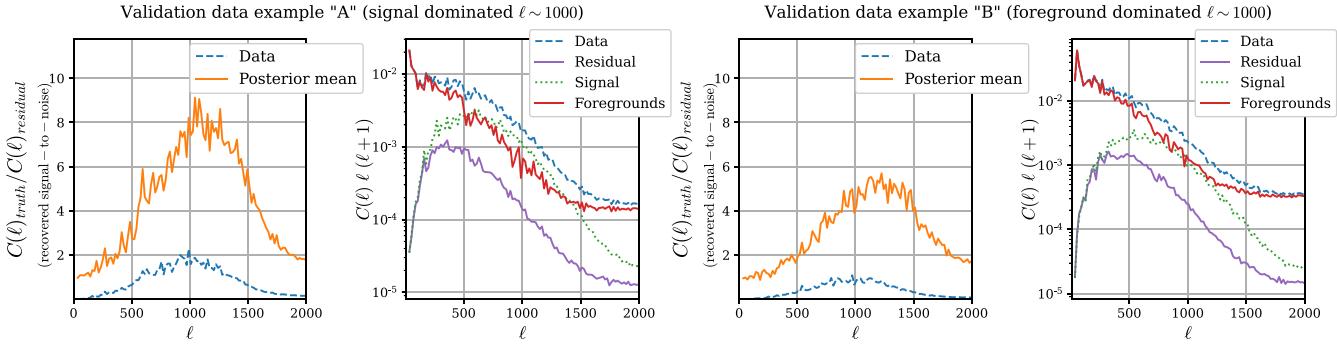
Fig. 1 shows the results of the trained Moment Network applied to validation data A. The left hand shows the clean target  $s_B$  and the foreground-contaminated data  $\mathbf{d}_B$ , though the posterior is conditioned on the  $E$ - and  $B$ -mode data. This figure is a demonstration of 5 of the 6 targets of the ‘Summary of our approach’ described in the Introduction (Section 1).

Though naively the behaviour of the mean map appears to match what would be expected from a linear filter (e.g. Wiener filtering of noise), here, at a single frequency, the ability to distinguish between target signal and the statistically-homogeneous foregrounds is entirely due to non-Gaussianity. Appendix B (Supporting Materials) shows the same example with a Gaussian foreground model.

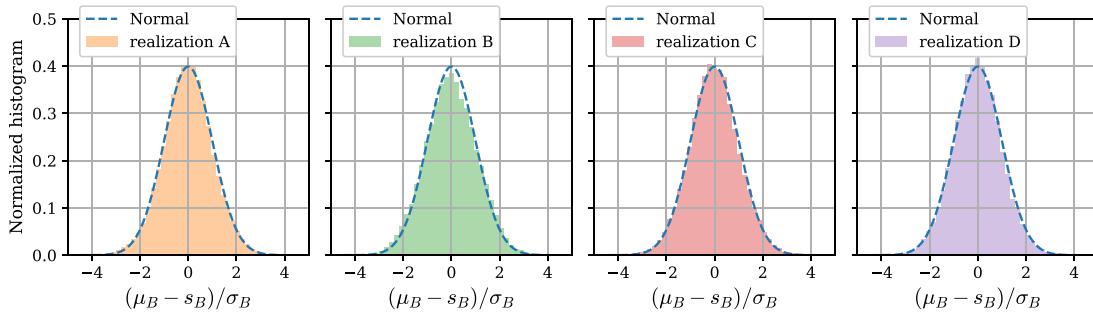
#### 3.2 Posterior mean as foreground cleaning

The aim of this high-dimensional likelihood-free inference approach is to estimate properties of the pixel posterior distribution. By validating this procedure, we can be confident in the forward model and the inference scheme for cosmological parameters or models.

However, the mean of the pixel posterior distribution is expected to be close to the true pixel values. As such, we can interpret the centre right hand of Fig. 1 as a foreground-cleaned map. It is important to note that the mean of the posterior is not a sample from the posterior, so it does not have the properties of *in-painting*; pixels with high



**Figure 2.** Each pair of panels on the left and right correspond to the validation data examples ‘A’ and ‘B’, respectively. The right hand of each pair shows the power spectra of data components and residuals. The left hand shows the ratio of the true power to the power of the residuals, between the posterior mean and the truth (orange solid line) or between the data and the truth (blue dashed line);  $C(\ell)_{\text{truth}}/C(\ell)_{\text{residual}}$  can be interpreted as signal-to-noise (see Section 3.2 for discussion). If interpreted as a foreground-cleaned map, the posterior mean per pixel shows a significant improvement over the data, as would be expected.



**Figure 3.** For each of the validation data maps (A, B, C, D), the distribution of the rescaled pixel residuals  $(\mu_B - s_B)/\sigma_B$ . This distribution is expected to have mean zero and unit standard deviation; the apparent Gaussianity is not necessarily expected or required and, though surprising, shows that the marginal posterior distributions are Gaussian for this data model. This result validates both the synthesis procedure with a single training image and the high-dimensional likelihood-free inference with Moment Networks.

uncertainty are expected to have values close to the global mean  $0 \mu\text{K}$ . This occurs when the foreground signal is particularly obscuring, which are accompanied by an appropriate increase in the marginal variance per pixel. We can test the accuracy of the recovered pixel values, provided we remember they are the mean values of a high-dimensional posterior probability  $p(s_B | \mathbf{d}_E, \mathbf{d}_B)$ .

The two left hands of Fig. 2 show results from validation data A and the two right hand show result from validation data B. The right hand of each pair, shows the power spectra of the components and residuals. We can see that the signal power dominates the foreground at  $\ell \approx 1000$  for validation data A but the foregrounds always dominate for validation data B. Though not shown, example C is neither foreground- nor signal-dominated and D is signal dominated.

The residuals are defined as the difference between the truth and a given map, either the original data  $\mathbf{d}_B$  or mean posterior map  $\mu_B$ . The left hand of each pair show the power spectrum of the truth divided by the power spectrum of the residual; this quantity (truth divided by an effective error) can be interpreted as a signal-to-noise. We clearly see that if we take the mean per pixel as a cleaned map, we significantly reduce the relative error compared with the foreground contaminated data.

### 3.3 Posterior probability validation

For each pixel value, we have a marginal posterior mean and variance. As this is simulated data and we have a large number of pixels, we can validate that the true values are distributed around the mean with the

correct variance as predicted by our inference pipeline. In particular the rescaled pixel residuals  $(\mu_B - s_B)/\sigma_B$  are expected to have zero mean and unit variance.

Fig. 3 shows the rescaled residual distributions for each of the validation data maps A, B, C, and D. The clear Gaussianity of the residuals is not generally expected, but indicates that the marginal posterior distributions are Gaussian for this data model. The high level of agreement between a unit Normal distribution and the rescaled residuals for each of the validation maps is a strong demonstration that the forward model and the high-dimensional likelihood-free inference is robust. This is the primarily validation of our approach.

## 4 CONCLUSIONS

In this letter, we have described two significant contributions to simulation-based inference of CMB polarization data. The first is the development of WPH synthesis to take a single training image of polarized dust foregrounds to create a large set of realistic training images for use in a forward model. The second is the application of high-dimensional likelihood-free methods, in particular Moment Networks, to target the posterior probability of the CMB signal pixel values given our observed data. Finally, we validate both aspects by performing a quantile-type test using mock data derived from an unseen (not synthesized) simulation of dust foregrounds.

The quantile-type test validates our approach in the most difficult case: Pixel-level, noise-free, highly non-Gaussian dust

foregrounds with a single training image at a single frequency. It would be straightforward to combine this single-frequency (morphological) approach with existing multi-frequency models. Future interesting developments would include astrophysical modelling of dust frequency decorrelation (Tassis & Pavlidou 2015), improvements in the accuracy and efficiency of WPH synthesis, and extensions of the likelihood-free inference methods. But now, with a few representative patches of the galactic microwave sky, either simulated or data-driven, we can overcome a significant challenge of dust emission for CMB polarization analysis in a likelihood-free framework.

## ACKNOWLEDGEMENTS

This research was supported by the Agence Nationale de la Recherche (Project BxB: ANR-17-CE31-0022).

## DATA AVAILABILITY

Reproducible code demonstrations and associated data are available on the [DeepMass](#) repository.

## REFERENCES

Ade P. et al., 2019, *J. Cosmology Astropart. Phys.*, 2019, 056  
 Adler J., Öktem O., 2018, preprint ([arXiv:1811.05910](#))  
 Allye E., Marchand T., Cardoso J. F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, *Phys. Rev. D*, 102, 103506  
 Alsing J., Wandelt B., Feeney S., 2018, *MNRAS*, 477, 2874  
 Alsing J., Charnock T., Feeney S., Wandelt B., 2019, *MNRAS*, 488, 4440  
 Aylor K., Haq M., Knox L., Hezaveh Y., Perreault-Levasseur L., 2020, *MNRAS*, 500, 3889  
 BICEP2 Collaboration, 2018, *Phys. Rev. Lett.*, 121, 221301  
 Brehmer J., Mishra-Sharma S., Hermans J., Louppe G., Cranmer K., 2019, *ApJ*, 886, 49  
 CMB-S4 Collaboration, 2020 preprint ([arXiv:2008.12619](#))  
 Darwisch O. et al., 2021, *MNRAS*, 500, 2250  
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759  
 Howlett C., Lewis A., Hall A., Challinor A., 2012, *J. Cosmology Astropart. Phys.*, 2012, 027  
 Jaynes E. T., 2003, *Probability Theory: The Logic of Science*. Cambridge University Press, Shaftesbury Road, Cambridge CB2 8EA  
 Jeffrey N., Wandelt B. D., 2020, Accepted in the Third Workshop on Machine Learning and the Physical Sciences. NeurIPS, Vancouver, Canada  
 Jeffrey N., Lanusse F., Lahav O., Starck J.-L., 2020, *MNRAS*, 492, 5023  
 Jeffrey N., Alsing J., Lanusse F., 2021, *MNRAS*, 501, 954  
 Kodi Ramanah D., Wojtak R., Ansari Z., Gall C., Hjorth J., 2020, *MNRAS*, 499, 1985  
 Kreisch C. D., Pisani A., Villaescusa-Navarro F., Spergel D. N., Wandelt B. D., Hamaus N., Bayer A. E., 2021, preprint ([arXiv:2107.02304](#))  
 Lemos P., Jeffrey N., Whiteway L., Lahav O., Libeskind N., Hoffman Y., 2021, *Phys. Rev. D*, 103, 023009  
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473  
 Mallat S., Zhang S., Rochette G., 2020, *Information and Inference: A Journal of the IMA*, 9, 721  
 Millea M., Anderes E., Wandelt B. D., 2020, *Phys. Rev. D*, 102, 123542  
 Papamakarios G., Murray I., 2016, in Lee D., Sugiyama M., Luxburg U., Guyon I., Garnett R., eds, *Advances in Neural Information Processing Systems*. Curran Associates, Inc., p. 1028  
 Planck Collaboration, 2020a *A&A*, 641, A4  
 Planck Collaboration, 2020b, *A&A*, 641, A6  
 Planck Collaboration, 2020c, *A&A*, 641, A11  
 Regaldo-Saint Blancard B., Levrier F., Allye E., Bellomi E., Boulanger F., 2020, *A&A*, 642, A217  
 Regaldo-Saint Blancard B., Allye E., Boulanger F., Levrier F., Jeffrey N., 2021, *A&A*, 649, L18  
 Tassis K., Pavlidou V., 2015, *MNRAS*, 451, L90  
 Zaldarriaga M., 2001, *Phys. Rev. D*, 64, 103001  
 Zhang S., Mallat S., 2021, *ACHA*, 53, 199

## SUPPORTING INFORMATION

Supplementary data are available at *MNRASL* online.

Figure S1. Upper row: Simulated  $(E, B)$  foregrounds as the target  $f^*$ . Lower row: Resulting patch  $bf$  of a typical synthesis as used in this work.

Figure S2. Validation data example ‘A’: Posterior mean using an analytic Gaussian statistical model, rather than a non-Gaussian forward model using likelihood-free inference.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a *TeX/LaTeX* file prepared by the author.