

Hierarchical Bayesian CMB component separation with the No-U-Turn Sampler

R. D. P. Grumitt ¹★, Luke R. P. Jew¹ and C. Dickinson ^{2,3}

¹Sub-department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

²Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Alan Turing Building, Oxford Road, Manchester M13 9PL, UK

³Cahill Centre for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2020 June 22. Received 2020 May 29; in original form 2019 October 30

ABSTRACT

In this paper, we present a novel implementation of Bayesian cosmic microwave background (CMB) component separation. We sample from the full posterior distribution using the No-U-Turn Sampler (NUTS), a gradient-based sampling algorithm. Alongside this, we introduce new foreground modelling approaches. We use the mean shift algorithm to define regions on the sky, clustering according to naively estimated foreground spectral parameters. Over these regions we adopt a complete pooling model, where we assume constant spectral parameters, and a hierarchical model, where we model individual pixel spectral parameters as being drawn from underlying hyperdistributions. We validate the algorithm against simulations of the *LiteBIRD* and C-Band All-Sky Survey (C-BASS) experiments, with an input tensor-to-scalar ratio of $r = 5 \times 10^{-3}$. Considering multipoles $30 \leq \ell < 180$, we are able to recover estimates for r . With *LiteBIRD*-only observations, and using the complete pooling model, we recover $r = (12.9 \pm 1.4) \times 10^{-3}$. For C-BASS and *LiteBIRD* observations we find $r = (9.0 \pm 1.1) \times 10^{-3}$ using the complete pooling model, and $r = (5.2 \pm 1.0) \times 10^{-3}$ using the hierarchical model. Unlike the complete pooling model, the hierarchical model captures pixel-scale spatial variations in the foreground spectral parameters, and therefore produces cosmological parameter estimates with reduced bias, without inflating their uncertainties. Measured by the rate of effective sample generation, NUTS offers performance improvements of $\sim 10^3$ over using Metropolis–Hastings to fit the complete pooling model. The efficiency of NUTS allows us to fit the more sophisticated hierarchical foreground model that would likely be intractable with non-gradient-based sampling algorithms.

Key words: methods: data analysis – methods: statistical – cosmic background radiation – cosmology: observations – radio continuum: general.

1 INTRODUCTION

One of the major outstanding goals of cosmic microwave background (CMB) cosmology is the detection of primordial B -modes in the CMB polarization (BICEP2/Keck Collaboration et al. 2015; Kamionkowski & Kovetz 2016). The challenge of detecting these B -modes has become a problem of accurate component separation, i.e. the extraction of the CMB B -mode signal from foreground contaminated observations of the radio and microwave sky (Betoule et al. 2009; Dunkley et al. 2009; Errard et al. 2016). We can parametrize the strength of the CMB B -mode signal through the tensor-to-scalar ratio, r , which gives the ratio of the amplitude of tensor-to-scalar perturbations in the early Universe (Dodelson 2003). Given current constraints on r , along with the targeted sensitivities of next-generation CMB experiments of $\sigma(r) \sim 10^{-3}$, future CMB experiments must be able to detect a CMB signal that is potentially subdominant to foreground emission across all of the sky, at all frequencies (Kogut et al. 2011; Abazajian et al. 2016; BICEP2 Collaboration et al. 2018; Remazeilles et al. 2018; Sekimoto et al. 2018; Ade et al. 2019; Hanany et al. 2019; Shandera et al. 2019).

This presents two primary challenges. First, we must be sure to have data of a sufficient sensitivity with enough frequency coverage to be able to model foreground spectral energy distributions (SEDs) with sufficient accuracy and precision. Secondly, our component separation algorithms must be able to extract the CMB signal from our noisy observations with high fidelity, and properly quantify the uncertainty in the extracted signal.

In this paper, we focus on the second challenge, in particular studying Bayesian parametric component separation. Numerous CMB component separation algorithms have been developed (see e.g. Delabrouille, Cardoso & Patanchon 2003; Martínez-González et al. 2003; Eriksen et al. 2008; Leach et al. 2008; Dunkley et al. 2009; Stivoli et al. 2010; Remazeilles, Delabrouille & Cardoso 2011; Planck Collaboration XII 2014; Planck Collaboration IX 2016; Planck Collaboration X 2016; Remazeilles et al. 2018; Seljebotn et al. 2019). Each of these algorithms offers different advantages and disadvantages in terms of accuracy, computational efficiency, and quantification of uncertainty. One of the main motivations for using Bayesian parametric component separation is the ability to obtain properly motivated probability distributions for our model parameters, and hence a proper quantification of the uncertainty. This does come at the cost of Bayesian inference being compu-

* E-mail: richard.grumitt@physics.ox.ac.uk

tationally expensive, especially when using Markov chain Monte Carlo (MCMC) techniques. Further, uncertainties in our modelling of foreground emission can lead to biases in our inference. This is potentially highly problematic in the case of B -mode science, where the potential biases can be of the same order as the value of r we are attempting to measure. The impact of such modelling errors has received significant previous attention (see e.g. Remazeilles et al. 2016, 2018; Chluba, Hill & Abitbol 2017). These modelling problems are very closely linked with the available data. If it becomes apparent that more complex modelling is required, more comprehensive data covering a wide range of frequencies will be required to constrain the additional model parameters (Jew et al. 2019).

Bayesian parametric component separation has notably been applied through the COMMANDER algorithm, which utilizes Gibbs sampling to jointly sample the CMB sky signal, power spectrum, and foreground and instrumental parameters (Gelfand & Smith 1990; Eriksen et al. 2004, 2008; Planck Collaboration XII 2014; Planck Collaboration IX 2016; Planck Collaboration X 2016). Gibbs sampling consists of drawing successive parameter samples from the conditional distributions of your parameters, as opposed to directly sampling from the full joint distribution. In the case of complex, high-dimensional distributions this can offer significant performance improvements over sampling directly from the joint distribution using algorithms such as Metropolis–Hastings, which quickly prove intractable for CMB component separation (Hastings 1970). However, even Gibbs sampling becomes computationally challenging as we move to higher resolution analyses. This has largely limited the application of Bayesian parametric component separation to studies of the large-scale CMB signal, up to multipoles of $\ell \sim 50$.

In this paper, we present a new implementation of Bayesian parametric CMB component separation, using the No-U-Turn Sampler (NUTS) to explore the parameter space. NUTS is a self-tuning variant of Hamiltonian Monte Carlo (HMC), originally presented in Hoffman & Gelman (2014). Since this original exposition, the algorithm has undergone a number of developments, with state-of-the-art implementations in STAN (Stan Development Team 2012; Carpenter et al. 2017) and PYMC3 (Salvatier, Wiecki & Fonnesbeck 2016). HMC algorithms make use of first-order gradient information to generate efficient proposal steps. This allows HMC to avoid the random walk behaviour of standard Metropolis–Hastings and Gibbs sampling methods, which becomes particularly problematic as the dimension of the parameter space increases.

To validate the performance of our component separation algorithm, we apply it to simulated observations of *LiteBIRD*, a planned next-generation CMB satellite (Sekimoto et al. 2018), and the C-Band All-Sky Survey (C-BASS), a 5 GHz ground-based experiment observing the sky in total intensity and polarization (Jones et al. 2018). Previous CMB component separation analyses have considered the extent to which foreground spectral parameters should be allowed to vary, such that a balance can be struck between model realism and simplicity. These studies have considered various approaches to defining regions on the sky, over which foreground spectral parameters are typically assumed to be constant (Stomp et al. 2009; Errard, Stivoli & Stomp 2011; Stomp, Errard & Poletti 2016; Alonso et al. 2017; Irfan et al. 2019; Khatri 2019; Thorne et al. 2019). In this paper, we use the mean shift algorithm to define regions on the sky, clustering according to naively estimated synchrotron and dust spectral parameters. We initially fit a complete pooling model where we assume foreground spectral parameters to be constant in each region. We then fit a hierarchical foreground model. Hierarchical modelling has recently been employed in the context of blind CMB component separation in Wagner-Carena

et al. (2020). For the hierarchical analysis in this paper, we assume individual pixel spectral parameters are drawn from underlying Gaussian distributions, jointly fitting for the mean and variance of the Gaussian hyperdistributions, and the individual pixel-by-pixel spectral parameters in each region. In doing so, we are able to provide a faithful generative description of the underlying foreground emission, whilst reducing the propensity for fitting noisy outliers when assuming total independence between pixel spectral parameters (Gelman 2006b; Gelman & Hill 2007).

Benchmarking against the rate at which the algorithm generates effective CMB amplitude samples, we find that NUTS offers performance improvements of $\sim 10^3$ over sampling with the Metropolis–Hastings algorithm for the complete pooling model. Sampling from the posterior distribution of the hierarchical model is particularly challenging. Hierarchical models are known to exhibit geometrical pathologies that make it extremely difficult to achieve convergence using non-gradient-based sampling algorithms. In these situations variants of HMC are often the only tractable approaches to sampling from the posterior (Betancourt & Girolami 2015).

The outline of this paper is as follows. In Section 2, we describe the diffuse component SED models used in our analysis. In Section 3, we describe the *LiteBIRD* and C-BASS simulations used in our algorithm validation. In Section 4, we describe the mean shift clustering algorithm used to define regions on the sky. In Section 5, we describe the component separation algorithm, give a general description of the NUTS algorithm, and discuss the complete pooling and hierarchical foreground models used during validation. In Section 6, we present the results from our algorithm validation. We conclude in Section 7.

2 DIFFUSE COMPONENT MODELS

In this section, we describe the diffuse component models employed in the validation of our component separation algorithm. Our focus here is on component separation for CMB polarization studies, and as such we consider only spectral models for synchrotron, thermal dust, and CMB emission. Additional contributions can potentially arise from polarized anomalous microwave emission (AME) and free–free emission. However, in both of these cases the level of polarized emission is expected to be very low. For AME, theoretical considerations from spinning dust models suggest that AME should be very weakly polarized, with expected polarization fractions of $\sim 10^{-6}$ (Draine & Hensley 2016). Existing measurements place upper limits on the AME polarization fraction of ~ 1 per cent (Dickinson et al. 2018). Free–free emission is caused by the scattering of electrons off ions in the interstellar medium. Given the random nature of this scattering, free–free emission is intrinsically unpolarized, with upper limits on the polarization fraction of ~ 1 per cent (Macellari et al. 2011). At the edges of bright H II regions, higher polarization fractions of ~ 10 per cent are possible due to additional Thompson scattering (Rybicki & Lightman 1985; Keating et al. 1998). However, for the purposes of CMB polarization studies, these effects are expected to be largely negligible.

It is important to note here that the true underlying foreground SEDs are more complicated than those used in this analysis. Mismodelling the underlying sky emission can result in biases in recovered cosmological parameter estimates. This has received significant attention in B -mode forecasting analyses (see e.g. Errard et al. 2016; Remazeilles et al. 2016, 2018). In this paper, we present a hierarchical foreground model (described in Section 5.4) that enables us to model spatial variations in foreground spectral parameters, without inflating parameter uncertainties. Given the simulated data sets considered here, fitting foreground spectral

parameters with any parametric method will ultimately be limited by the number of frequency channels available with sufficient sensitivity. For example, in Jew et al. (2019), it was found that to constrain models including the synchrotron curvature would require additional low-frequency channels between 5 and 30 GHz. Given a maximum observed frequency of ~ 400 GHz, dust spectral parameters in the modified blackbody (MBB) model are also very poorly constrained (Remazeilles 2018; Jew et al. 2019). This problem could be alleviated to some extent by considering model reparametrizations that yield more tractable posterior geometries. However, at its core this is a problem in the data.

2.1 Synchrotron

Synchrotron emission is caused by electrons spiralling in the Galactic magnetic field (GMF), and is the dominant diffuse component at low frequencies ($\nu \lesssim 100$ GHz in polarization), contributing to both total intensity and polarized emission. In ordered magnetic fields, the polarization fraction of synchrotron emission can be as high as 70 per cent, with more typical values of around 40 per cent at high Galactic latitudes (Rybicki & Lightman 1985; Vidal et al. 2015; Planck Collaboration XXV 2016). Over a wide range of frequencies, from $\mathcal{O}(10)$ MHz up to $\mathcal{O}(100)$ GHz, the synchrotron SED can be reasonably approximated as a power-law spectrum (Lawson et al. 1987; Reich & Reich 1988; Platania et al. 2003; Davies et al. 2006; Guzmán et al. 2011). We may parametrize the power-law spectrum as

$$S_s = A_s \left(\frac{\nu}{\nu_0} \right)^{-\beta_s}, \quad (1)$$

where A_s is the reference synchrotron amplitude, ν is the observing frequency, ν_0 is some reference frequency, and β_s is the synchrotron spectral index.

In reality, the synchrotron spectrum is modified by a combination of intrinsic effects, e.g. spectral ageing, along with pixel and beam averaging effects (Mitton & Ryle 1969; Harwood et al. 2016; Chluba et al. 2017; Remazeilles et al. 2018). These complications can be modelled through additional spectral curvature terms in the power law, or through the moment expansion method presented in Chluba et al. (2017). However, for the purposes of our validation analysis in this paper, it is sufficient to consider a simple power-law model.

Analysis in Jew et al. (2019) showed that experiments such as *LiteBIRD* struggle to constrain synchrotron spectral parameters, with additional low-frequency data being necessary to begin to constrain the synchrotron spectral index. To constrain more complex synchrotron curvature models would require additional low-frequency channels. The European Low-Frequency Survey (ELFS) is a proposed experiment, covering frequencies between 6 and 30 GHz, that would allow us to constrain more complex synchrotron SEDs (Ganga et al. 2019; Hill-Valler 2019).

2.2 Thermal dust

Thermal dust emission is caused by thermal emission from interstellar dust grains. In general, dust grains are not spherically symmetric and emit preferentially along their longer axis (Planck Collaboration XIX 2015; Planck Collaboration XI 2018). These interstellar dust grains align with local magnetic fields, resulting in emission in both total intensity and polarization. Thermal dust polarization fractions can be as high as 20 per cent, with a median value across the sky of approximately 8 per cent (Planck Collaboration XIX 2015). Thermal dust is the dominant component at high frequencies ($\nu \gtrsim 100$ GHz).

The thermal dust spectrum can be approximated by an MBB model, given by

$$S_d = A_d \left(\frac{\nu}{\nu_0} \right)^{\beta_d+1} \frac{\exp(\gamma \nu_0) - 1}{\exp(\gamma \nu) - 1}, \quad (2)$$

where A_d is the reference dust amplitude, ν_0 is some reference frequency, $\gamma = h/k_B T_d$, T_d is the dust temperature, and β_d is the dust spectral index. This model is a simplification, in reality multiple dust populations will exist along the line of sight. More complex models have previously been considered in Hensley & Bull (2018). However, these more complex models encounter the same issues as for synchrotron emission. The additional parameters introduced require additional data at frequencies $\nu \gtrsim 400$ GHz to constrain them. Given the simulated frequency coverage we consider in this paper, we already struggle to place constraints on the spectral parameters of the single MBB model.

2.3 CMB

The CMB follows a blackbody spectrum given by

$$S_{\text{cmb}} = A_{\text{cmb}} \frac{x^2 \exp(x)}{(\exp(x) - 1)^2}, \quad (3)$$

where A_{cmb} is the CMB amplitude, $x = h\nu/k_B T_{\text{cmb}}$, and $T_{\text{cmb}} = 2.7255$ K is the mean CMB temperature (Fixsen 2009). Throughout this paper we work in units of Rayleigh–Jeans brightness temperature unless otherwise stated.

3 SIMULATIONS

To validate the performance of our component separation algorithm we generate a set of simulated Stokes Q and U maps, corresponding to the frequencies and sensitivities of the C-BASS and *LiteBIRD* experiments (Jones et al. 2018; Sekimoto et al. 2018). The frequencies and polarization sensitivities of the C-BASS and *LiteBIRD* simulations are given in Table 1.

LiteBIRD is a planned next-generation CMB satellite, aiming to measure the tensor-to-scalar ratio with a sensitivity of $\sigma(r) \sim 10^{-3}$. To accomplish this task *LiteBIRD* will target large angular scales up to $\ell \sim 200$, covering the reionization peak at $\ell \sim 10$ and the recombination peak at $\ell \sim 80$ in the primordial B -mode power spectrum. The experiment is proposed to cover frequencies from 40 to 402 GHz, with the lowest resolution 40 GHz channel having a resolution of approximately 70 arcmin. We smooth all of our simulated maps to this 70 arcmin resolution, which is sufficient for our validation analysis targeting angular scales $\ell \lesssim 200$.

C-BASS is a 5 GHz experiment, observing the whole sky in intensity and polarization at a native resolution of 45 arcmin, at a sensitivity of 0.1 mK beam^{-1} ($4320 \mu\text{K arcmin}$). The primary purpose of the survey is to provide improved constraints on polarized synchrotron emission, to aid in CMB component separation analyses (Jones et al. 2018).

We simulate maps of polarized emission, containing contributions from synchrotron, thermal dust, and the CMB, using PySM (Thorne et al. 2017). We adopt the SEDs for our sky components described in Section 2. For synchrotron emission we use the PySM s1 model. This model uses the 9-year *Wilkinson Microwave Anisotropy Probe* (WMAP) 23 GHz maps (Bennett et al. 2013), smoothed to 3° , as synchrotron Q/U templates. Small scales are added to these templates by extrapolating the map power spectra to high ℓ and obtaining Gaussian realizations of the power spectra. Details of the implementation can be found in Thorne et al. (2017). These synchrotron templates are

Table 1. Frequencies and polarization sensitivities for the C-BASS and *LiteBIRD* experiments, used in our simulations for algorithm validation. Sensitivities are given in CMB thermodynamic temperature units. Simulated maps were smoothed to a common resolution of 70 arcmin, corresponding to the lowest resolution 40 GHz *LiteBIRD* channel. This resolution is sufficient for targeting multipoles $\ell \lesssim 200$, i.e. angular scales corresponding to the reionization and recombination peaks of the primordial *B*-mode power spectrum.

Experiment name	Frequency (GHz)	Polarization sensitivity (μK_{cmb} arcmin)
C-BASS ^a	5	4320
<i>LiteBIRD</i> ^b	40	27.9
	50	19.6
	60	15.6
	68	12.3
	78	10.0
	89	9.4
	100	7.6
	119	6.4
	140	5.1
	166	7.0
	195	5.8
	235	8.0
	280	9.1
	337	11.4
402	19.6	

^aJones et al. (2018).

^bSekimoto et al. (2018).

then extrapolated to higher frequencies using a spatially varying spectral index map taken from Miville-Deschênes et al. (2008). Across the whole sky, the synchrotron spectral index map has a mean of $\langle\beta_s\rangle \approx -3.0$ and a standard deviation of $\sigma(\beta_s) \approx 0.06$. Analysis of the synchrotron angular power spectrum in Krachmalnicoff et al. (2018) found that the spectral index map used in PYSM lacks power on all angular scales. However, for the purposes of our validation analysis it is sufficient.

For thermal dust we use the PYSM d1 model, which uses the *Planck* 353 GHz maps as *Q/U* thermal dust templates. The templates are scaled using the spatially varying dust temperature and spectral index maps obtained from the *Planck* COMMANDER analysis (Planck Collaboration X 2016). Across the whole sky the mean of the dust temperature map is $\langle T_d \rangle \approx 20.9$ K and the standard deviation is $\sigma(T_d) \approx 2.2$ K. For the dust spectral index the mean value is $\langle\beta_d\rangle \approx 1.54$ and the standard deviation is $\sigma(\beta_d) \approx 0.04$. The dust templates are smoothed to $2^\circ.6$, with small scales being added using the same prescription as for synchrotron emission.

The CMB map was generated as a realization of lensed CMB power spectra, $\mathbf{C}_\ell = (C_\ell^{TT}, C_\ell^{EE}, C_\ell^{BB}, C_\ell^{TE})$, calculated using CAMB (Lewis, Challinor & Lasenby 2000). For the *B*-mode simulation we set $r = 5 \times 10^{-3}$, and the lensing amplitude $A_L = 1$. In this validation study, we do not consider the detailed impact of delensing on recovered *B*-mode estimates. In future work it would be valuable to combine the foreground analysis presented here, with a Bayesian delensing scheme such as in Millea, Anderes & Wandelt (2020), to form a complete forward model for our cosmological observations. The component amplitude templates and spectral parameter maps used in our simulations are shown in Fig. 1.

4 MEAN SHIFT CLUSTERING OF SKY REGIONS

The modelling of spectral parameters in CMB component separation presents a number of challenges. In the face of limited data it can prove difficult to properly constrain spectral parameters. Attempting to allow full pixel-by-pixel variations in the spectral parameters in this situation can result in significant increase in post-component separation noise and in the prior dominating the posterior. Given this, one may seek to reduce the number of degrees of freedom in the sky model by fitting for global spectral parameters. However, this approach will inevitably lead to modelling errors that have the potential to bias cosmological measurements made with the derived CMB map (Thorne et al. 2019). These challenges have motivated modelling approaches where spatially uniform spectral parameters are assumed over a set of defined sky regions. These have included regions defined as superpixels on low NSIDE HEALPIX maps, and regions defined according to similarities in their spectral properties (Górski et al. 2005; Stompór et al. 2009, 2016; Errard et al. 2011; Alonso et al. 2017; Irfan et al. 2019; Khatri 2019; Thorne et al. 2019). A detailed discussion of the modelling approaches that can be adopted for spectral parameters in these sky regions is given in Sections 5.3 and 5.4. In this section, we describe the mean shift clustering algorithm, implemented in SCIKIT-LEARN (Comaniciu & Meer 2002; Pedregosa et al. 2011), that we have used to define regions on the sky for our component separation analyses. The mean shift clustering algorithm has previously been used in Jew & Grumitt (2020) to identify pixels with good detections of the synchrotron spectral index.

We use the mean shift clustering algorithm to construct sky regions according to their location on the sky and their spectral properties. Specifically, we cluster according to the Cartesian coordinates of pixel centres on the unit sphere (x^1, x^2, x^3) and the naive spectral indices between the C-BASS 5 GHz and *LiteBIRD* 40 GHz channels, and the *LiteBIRD* 337 and 402 GHz channels in polarized intensity. The two sets of frequency maps are used as synchrotron and thermal dust tracers, respectively, with the naive spectral indices in a pixel, p , being given by

$$\beta_p^{i,j} = \frac{\ln(m_i(p)/m_j(p))}{\ln(\nu_i/\nu_j)}, \quad (4)$$

where $m_i(p)$ and $m_j(p)$ are the map values in the pixel p , at the frequencies ν_i and ν_j , respectively. Given these parameters we may form the feature vector,

$$\mathbf{z}_p = (x_p^1/\zeta, x_p^2/\zeta, x_p^3/\zeta, \beta_p^{5,40}, \beta_p^{337,402}), \quad (5)$$

where ζ is a spatial vector scaling factor. Setting the value of ζ to be less than 1 allows us to preferentially weight proximity on the sky as being favourable over proximity in spectral index space. The function of the mean shift algorithm is then to cluster points in this five-dimensional feature space.

The mean shift algorithm proceeds by assigning a walker to each pixel, giving us the starting vectors, $\mathbf{z}_p^0 = \mathbf{z}_p$. The walkers then step through the feature space towards regions of higher density, with the t th update being calculated as

$$\mathbf{z}_p^t = \mathbf{z}_p^{t-1} + \mathbf{s}(\mathbf{z}_p^{t-1}), \quad (6)$$

where \mathbf{s} is the mean shift vector, given by

$$\mathbf{s}(\mathbf{z}_p^t) = \frac{\sum_q K(\mathbf{z}_q^t - \mathbf{z}_p^t) \mathbf{z}_q^t}{\sum_q K(\mathbf{z}_q^t - \mathbf{z}_p^t)}. \quad (7)$$

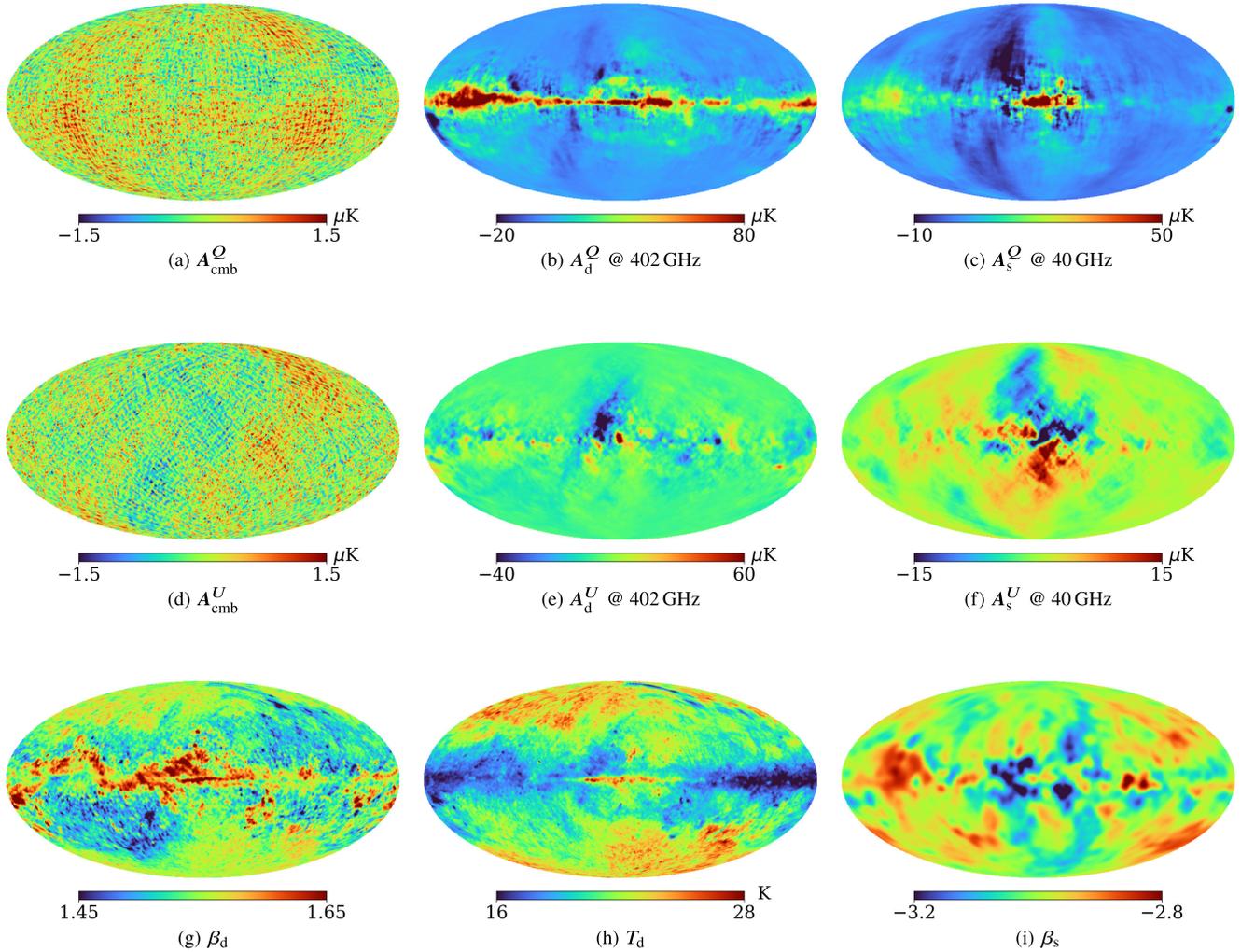


Figure 1. Input component parameter maps used in our simulations. The synchrotron and dust amplitude maps are shown here at reference frequencies of 40 and 402 GHz, respectively. When performing the component separation with C-BASS and *LiteBIRD* simulations, we set the synchrotron reference frequency to $\nu_0 = 5$ GHz.

For our purposes we choose the kernel, K , to be a top-hat defined by

$$K(\Delta) = \begin{cases} 1, & |\Delta| \leq \omega, \\ 0, & |\Delta| > \omega, \end{cases} \quad (8)$$

where ω is the bandwidth parameter. Walkers take steps until they converge, i.e. walker positions in feature space no longer change with new updates (up to some threshold). Regions are then defined as a set of pixels whose walkers have converged on the same position in the five-dimensional feature space. Any regions containing fewer pixels than some arbitrary minimum are reassigned to the nearest region in feature space containing a sufficient number of pixels.

In Fig. 2, we show the regions obtained using the frequency channels outlined above, which are used as the region definitions for the component separation analyses in this paper. We used the simulated maps at a HEALPIX NSIDE of 64, setting $\zeta = 0.5$, $\omega = 0.3$, and the minimum number of pixels in a region to 10 (Górski et al. 2005). Using the foreground tracers and parameters described here, we obtain 171 regions. The smallest region on the sky contains 138 pixels, and the largest region contains 645 pixels. The mean

number of pixels in a region is $\langle N_{\text{pix}} \rangle \approx 290$, and the standard deviation in the number of pixels is $\sigma(N_{\text{pix}}) \approx 95$. In regions of low signal-to-noise ratio (SNR) the borders of regions become less smooth. This can be mitigated to some extent by defining regions on lower NSIDE maps, at the cost of the larger pixel size meaning regions become more coarse. Using lower NSIDE maps can lead to a significant degradation in computational performance when assuming foreground spectral parameters in a given region are related in some manner.

A range of possible, non-trivial, extensions to the clustering algorithm exist that could help to alleviate some of the issues surrounding region definition in areas of the sky with low SNR. Instead of using naive spectral indices as a tracer of the spectral properties of diffuse emission over the sky, a more sophisticated estimation of the spectral indices could be performed, accounting for noise properties across the sky. The spectral index estimation could also be improved with better tracers of synchrotron and thermal dust emission. It is worth noting here that the use of the C-BASS map as a synchrotron template is particularly important. If instead we had only used low-frequency *LiteBIRD* channels as our synchrotron

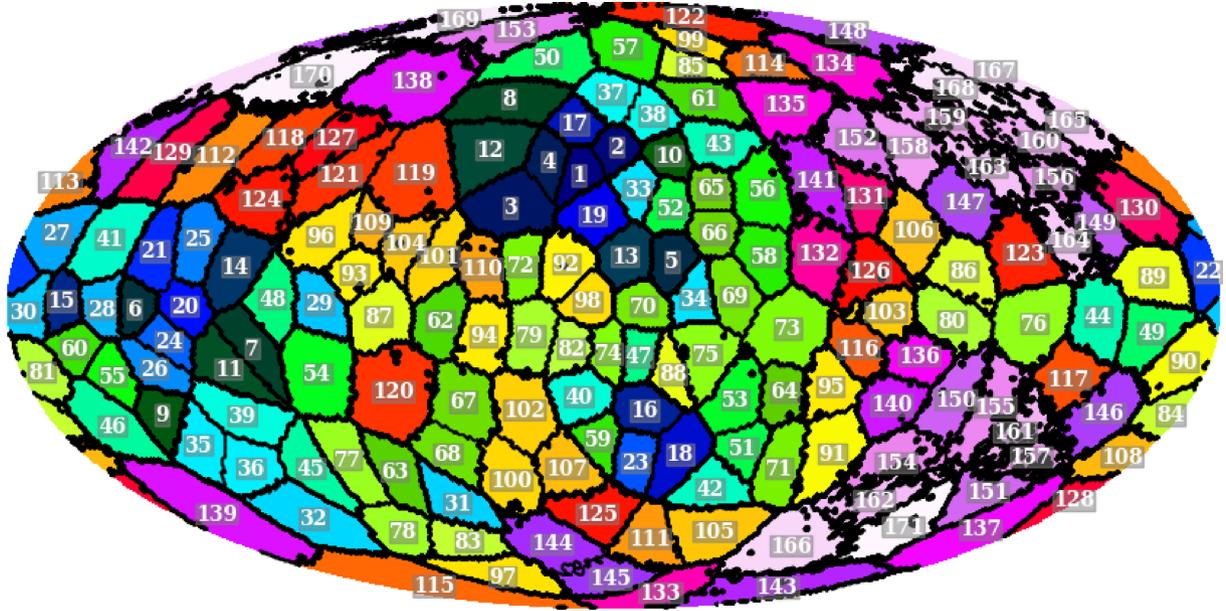


Figure 2. Regions defined using the mean shift clustering algorithm, produced using $N_{\text{SIDE}} = 64$ maps. We set $\zeta = 0.5$, $\omega = 0.3$, and the minimum number of pixels in a region to 10. Given our set-up, we obtain 171 regions. The smallest region contains 138 pixels, and the largest region contains 645 pixels. The mean number of pixels is $\langle N_{\text{pix}} \rangle \approx 290$, and the standard deviation of the number of pixels is $\sigma(N_{\text{pix}}) \approx 95$. In regions of low signal-to-noise ratio (SNR), we can see that the region borders become less smooth, due to the naive spectral indices used to cluster pixels on the sky becoming contaminated by noise.

tracers, the naive synchrotron spectral index estimates would have been heavily noise dominated. Using the methods outlined here, we found it would only be possible to define very coarse regions on $N_{\text{SIDE}} = 8$ maps.

5 PARAMETRIC BAYESIAN CMB COMPONENT SEPARATION

We have developed a new implementation of Bayesian pixel-by-pixel CMB component separation using the NUTS algorithm to explore our parameter space (Hoffman & Gelman 2014). The primary benefit in using NUTS to sample from the target distribution is in its avoidance of the random walk behaviour that slows more conventional sampling algorithms such as Metropolis–Hastings and Gibbs sampling. The component separation code is written in the PYTHON programming language, with the NUTS algorithm being implemented through the PYMC3 library (Salvatier et al. 2016). In its current form, the whole-sky component separation is parallelized over the sky regions defined using the mean shift clustering algorithm. For our validation purposes in this paper, we do not consider monopole and dipole corrections, or instrumental factors such as colour corrections in our modelling. The extension of our modelling to include such complications is left to future work.

Benchmarking the algorithm performance against the rate of effective sample generation, NUTS offers potential performance improvements of $\sim 10^3$ compared to sampling with Metropolis–Hastings.¹ Close to the Galactic plane sampling becomes more difficult, with the CMB completely subdominant to foregrounds. In this situation computational performance can be degraded such that the sampler exhibits undesirable random walk behaviour. By parallelizing over sky regions, and masking the most contaminated

sky regions close to the Galactic plane, it should be possible to achieve rapid convergence. We also note that these benchmarking tests have been performed without extensive optimization of the component separation code. By exploring reparametrizations, model prior choice, optimizing sampling parameters, etc. it is likely that we would be able to achieve further performance improvements. It would also be worthwhile considering the potential for GPU acceleration with PYMC3. For the hierarchical model, the posterior exhibits geometrical pathologies that make sampling with non-gradient-based algorithms essentially intractable (Betancourt & Girolami 2015). However, using NUTS we are able to achieve comparable computational performance to the complete pooling model.

The outline of this section is as follows. In Section 5.1, we describe the general sky model and likelihood used in our component separation analysis. In Section 5.2, we give an overview of the NUTS algorithm. In Section 5.3, we discuss the complete pooling model, where we fit for constant spectral parameters over sky regions. In Section 5.4, we discuss our hierarchical modelling approach, where we directly fit for the underlying hyperdistributions of the spectral parameters. In Section 5.5, we describe the convergence checks and diagnostics performed during sampling.

5.1 Sky model and likelihood

In attempting to observe the CMB we actually observe multiple sky components. It is the goal of our component separation to extract the CMB from these additional confusing components. For a given sky pixel, p , we may write the observed value in that pixel as

$$d_{p,\lambda}(v) = s_{p,\lambda}(v) + n_{p,\lambda}(v), \quad (9)$$

where v is the observing frequency, $\lambda = \{I, Q, U\}$ represents one of the Stokes parameters, $s_{p,\lambda}(v)$ is the true sky signal, and $n_{p,\lambda}(v)$ is the noise term. For our analysis in this paper we restrict ourselves to $\lambda = \{Q, U\}$.

¹These simple benchmarking tests were performed on a single Intel Xeon CPU, running at ~ 2.6 GHz.

The true sky signal includes contributions from diffuse emission, compact sources, line emission, etc. For our current modelling purposes we focus on diffuse emission, as discussed in Section 2. Additional contributions from compact sources, line emission, etc. are important considerations when performing component separation on real experimental data. A variety of strategies exist for mitigating their contribution, e.g. through masking point sources, and direct modelling during the fitting process. However, this sits beyond the scope of our algorithm validation.

The noise term consists largely of contributions from instrumental white noise and $1/f$ noise, which acts to introduce large-scale correlated noise in the sky maps. For ground-based experiments mitigating atmospheric noise is a significant challenge, largely limiting accessible angular scales to $\ell \gtrsim 30$. The ability of such experiments to accurately recover low multipoles will be vital for future B -mode experiments (Alonso et al. 2017).

Given a set of sky maps at frequencies $\nu = [\nu_1, \dots, \nu_N]$, the Gaussian likelihood for a pixel, p , is given by

$$-\ln \mathcal{L}_{p,\lambda}(\Theta_{p,\lambda}) = \sum_i \frac{1}{2} \ln (2\pi\sigma_{p,\lambda}^2(\nu_i)) + \frac{(d_{p,\lambda}(\nu_i) - s_{p,\lambda}(\nu_i, \Theta_{p,\lambda}))^2}{2\sigma_{p,\lambda}^2(\nu_i)}, \quad (10)$$

where $\Theta_{p,\lambda}$ are the model parameters, and $\sigma_{p,\lambda}$ is the pixel noise. We assume for simplicity that the noise is independent between pixels and frequency channels. This is not necessarily optimal when we consider the complications described above. However, the approximation is sufficient for the purposes of algorithm validation.

5.2 The No-U-Turn Sampler

The key measure of the efficiency of a sampling algorithm is in its ability to produce effective/independent samples, or equivalently reduce the correlation between samples. Indeed, this is where NUTS significantly out-performs standard Metropolis–Hastings and Gibbs sampling algorithms. Even though an individual step in these simpler sampling algorithms is less computationally expensive, their random walk behaviour results in highly correlated samples and hence very inefficient generation of effective samples. At its core, the NUTS algorithm is an extension of HMC, which was originally developed for performing calculations in lattice field theory (Duane et al. 1987). In this section, we give an overview of the HMC algorithm, and the additional tuning procedures NUTS implements to avoid the need for hand-tuned HMC implementations. For a detailed discussion of NUTS and HMC see Hoffman & Gelman (2014), Betancourt (2017), Betancourt et al. (2017), and Monahan, Thorson & Branch (2017).

HMC essentially proceeds by generating physical trajectories through parameter space, akin to simulating particle trajectories through a potential. A simple implementation of HMC may proceed as follows.

(i) Given a set of parameters, $\Theta = [\Theta_1, \dots, \Theta_d]$, with corresponding joint density, $p(\Theta)$, we introduce a set of auxiliary momentum variables, $\mathbf{r} = [r_1, \dots, r_d]$. We take the distribution over the momenta to be a Gaussian distribution centred on zero, i.e. $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{M})$, where \mathbf{M} is the mass matrix. This defines a kinetic energy term,

$$K(\Theta, \mathbf{r}) = -\ln p(\mathbf{r}|\Theta) = \frac{1}{2} \mathbf{r}^\top \mathbf{M}^{-1} \mathbf{r} + \ln |\mathbf{M}| + \text{const}. \quad (11)$$

We may then define the Hamiltonian of our system as

$$\mathcal{H}(\mathbf{r}, \Theta) = -\ln p(\mathbf{r}|\Theta) - \ln p(\Theta) = K(\Theta, \mathbf{r}) + V(\Theta), \quad (12)$$

where we define the potential energy term, $V(\Theta) = -\ln p(\Theta)$

(ii) We then evolve our position in parameter space by integrating Hamilton’s equations:

$$\frac{d\Theta}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbf{r}}, \quad (13)$$

$$\frac{d\mathbf{r}}{dt} = -\frac{\partial \mathcal{H}}{\partial \Theta}. \quad (14)$$

Practically this is done through a leapfrog algorithm. To generate a new sample in our Markov chain we draw \mathbf{r} from $\mathcal{N}(\mathbf{0}, \mathbf{M})$. The leapfrog steps then proceed as

$$\mathbf{r}^{t+\epsilon/2} = \mathbf{r}^t - \frac{\epsilon}{2} \nabla_{\Theta} V(\Theta^t), \quad (15)$$

$$\Theta^{t+\epsilon} = \Theta^t + \epsilon \mathbf{M} \mathbf{r}^{t+\epsilon/2}, \quad (16)$$

$$\mathbf{r}^{t+\epsilon} = \mathbf{r}^{t+\epsilon/2} - \frac{\epsilon}{2} \nabla_{\Theta} V(\Theta^{t+\epsilon}), \quad (17)$$

where ϵ is the leapfrog step-size. The leapfrog steps used to update our position have the convenient property of being a symplectic integrator. That is, the numerical trajectory generated by the leapfrog steps preserve the volume of phase space, as is the case for the Hamiltonian trajectories they approximate. A more detailed discussion of the numerical integration of Hamilton’s equations can be found in Leimkuhler & Reich (2004). These leapfrog steps are performed L times to generate a new proposal position, (\mathbf{r}^*, Θ^*) .

(iii) The new proposal position is then accepted with a probability of acceptance given by

$$p_{\alpha} = \min \{1, \exp(\mathcal{H}(\mathbf{r}, \Theta) - \mathcal{H}(\mathbf{r}^*, \Theta^*))\}. \quad (18)$$

We note that what we have done here is essentially generate a Metropolis–Hastings proposal step with a very high chance of being accepted.

(iv) By repeating this sampling procedure N times we may generate the parameter samples for our Markov chain.

The mass matrix used to define the distribution over the momenta acts to rotate and rescale parameter space. Choosing \mathbf{M}^{-1} to be the covariance of the target distribution will help to decorrelate the target distribution, which can lead to significant performance improvements when dealing with highly correlated parameters. For practical implementations the mass matrix can be estimated during a tuning phase. Starting with the identity matrix we can generate an initial sample set, from which we update our estimate of the mass matrix using the sample covariance. We may then iterate over this tuning process to generate an accurate estimate of the covariance of the target distribution (Betancourt 2017). Whilst estimating off-diagonal elements of the mass matrix does help in decorrelating the target distribution, using the off-diagonal elements does not necessarily scale well to high-dimensional problems given the need to invert the mass matrix at the end of tuning, and perform matrix multiplications during leapfrog steps. For our analysis in this paper we only tune diagonal elements of the mass matrix. However, this can significantly improve sampling efficiency for single pixel analyses, or analyses assuming complete independence between pixels. In this case, one can employ the tuning steps in Foreman-Mackey, Barentsen & Barclay (2019), using the default tuning schedule described in Stan Development Team (2012).

The efficiency of HMC as described above critically depends on the choice of ϵ and L used in the leapfrog steps. If ϵ is chosen to be

Table 2. Priors for parameters in the complete pooling model. We assign informative normal priors to the spectral parameters, and flat priors to the amplitude parameters. In addition to the normal priors for the spectral parameters, we multiply these by their associated Jeffreys priors as in the *Planck* COMMANDER analysis. Spectral parameters are assumed to be constant over a given region, whilst amplitude parameters are allowed to vary within each pixel, p . We restrict our analysis to polarization, so that $\lambda = \{Q, U\}$.

Θ	$p(\Theta)$
β_s	$\mathcal{N}(\mu = -3, \sigma = 0.3)$
β_d	$\mathcal{N}(\mu = 1.6, \sigma = 0.3)$
T_d	$\mathcal{N}(\mu = 21.0, \sigma = 2.0)$
$A_s^{p,\lambda}$	Unif $(-\infty, \infty)$
$A_d^{p,\lambda}$	Unif $(-\infty, \infty)$
$A_{\text{cmb}}^{p,\lambda}$	Unif $(-\infty, \infty)$

too small, the sampler will waste computation taking very small steps along the Hamiltonian trajectories. In contrast, if ϵ is chosen to be too large, the simulation of the Hamiltonian trajectory will become inaccurate and the sampler will produce proposal steps with low acceptance probabilities. If L is chosen to be too small, the sampler will generate samples close to one another, resulting in undesirable random walk behaviour. If L is chosen to be too large on the other hand, the sampler will generate paths through parameter space that loop back on themselves. This results in proposal steps close to the starting value, with the additional waste of generating the extended trajectory. In even more severe scenarios, a poor choice of L that results in the sampler jumping from one side of parameter space to another at each iteration can result in a non-ergodic chain, i.e. a chain that is not guaranteed to converge on the target distribution (Neal 2012).

The need to finely tune ϵ and L means that standard implementations of HMC typically require costly tuning runs. This can significantly reduce the utility and general applicability of HMC in realistic problems. The NUTS algorithm overcomes these problems by automatically tuning these sampling parameters. The value of ϵ is tuned during an initial tuning phase to meet some target acceptance probability. The target acceptance probability can be adjusted depending on the degree of curvature in the posterior, with a higher acceptance rate (or equivalently, smaller step-size) being needed for highly curved distributions. The value of L is modified during sampling to meet a No-U-Turn criterion. That is, the leapfrog integrator is iterated over until the simulated trajectory begins to turn back on itself, or some maximum number of leapfrog simulations are performed. In doing so, the sampler is able to maximize the distance between the proposal step and the initial position, before looping back on itself and wasting computation. Details on these tuning procedures can be found in Stan Development Team (2012), Hoffman & Gelman (2014), and Salvatier et al. (2016).

5.3 Complete pooling of spectral parameters

As discussed in Section 4, in the face of limited data and low SNR, allowing spectral parameters to vary completely from pixel-to-pixel is suboptimal, resulting in increased levels of post-component separation noise and the posterior potentially becoming prior dominated. As a first alternative to allowing full pixel-by-pixel variations we may instead assume spectral parameters to be constant over the regions defined as in Section 4, i.e. we assume a complete pooling of the spectral parameters.

The priors used in the complete pooling model are given in Table 2. We assign informative normal priors to the spectral parameters. The

standard deviations on the priors for β_s and β_d are chosen to be 0.3, corresponding to the bandwidth used in clustering sky regions and encompassing most of the range over which these parameters have been measured (Planck Collaboration X 2016; Krachmalnicoff et al. 2018). For the dust temperature we set a prior based on constraints on the dust temperature found in the *Planck* COMMANDER analysis (Planck Collaboration IX 2016; Planck Collaboration X 2016). These help to downweight the more extreme regions of parameter space, offering significant computational performance improvements, and helping to regularize the posterior by reducing the biasing effect of the probability mass associated with extreme parameter values. Detailed discussion of prior choice, in particular around the use of weakly informative priors, can be found in Gelman (2006a), Gelman et al. (2008), Evans & Jang (2011), Polson & Scott (2012), Gelman & Hennig (2017), Simpson et al. (2017), and Gelman, Simpson & Betancourt (2017). In addition to the normal priors on the spectral parameters, we multiply these by the associated Jeffreys priors as in the *Planck* COMMANDER analysis (Jeffreys 1946, 1961; Eriksen et al. 2008; Planck Collaboration XII 2014; Planck Collaboration IX 2016; Planck Collaboration X 2016).

Our emission model for a pixel, p , in some sky region is given by

$$s_{p,\lambda}(v) = A_s^{p,\lambda} f_s(v, \beta_s) + A_d^{p,\lambda} f_d(v, \beta_d, T_d) + A_{\text{cmb}}^{p,\lambda} f_{\text{cmb}}(v). \quad (19)$$

The functions $f_s(v, \beta_s)$, $f_d(v, \beta_d, T_d)$, and $f_{\text{cmb}}(v)$ are the spectral forms of the synchrotron, dust, and CMB components, as defined in Section 2. Note that we assume the spectral parameters to be identical for $\lambda = \{Q, U\}$. We model our data as being normally distributed, i.e. we assume the Gaussian likelihood in equation (10).

Complete pooling offers a potentially effective approach to account for the spatial variation in spectral parameters whilst avoiding the generation of excessive post-component separation noise. However, with additional data points at low and/or high frequencies it is possible to adopt a more sophisticated, hierarchical model of the spectral parameters in these regions. It is worth noting that for the complete pooling model one can analytically marginalize over the amplitude parameters as in Alonso et al. (2017). This greatly reduces the dimension of parameter space and hence improves the sampling efficiency. We have not implemented sampling of this marginal distribution for our analysis here, where we study the computational performance of NUTS in sampling the full posterior. Indeed, the intrinsic efficiency of NUTS makes this unnecessary.

5.4 Hierarchical modelling of spectral parameters

In statistical modelling, we often encounter scenarios where our model contains a set of latent variables that are related in some way. In such a scenario it is neither ideal to treat the latent variables as being entirely independent nor to simply fit for a single, global variable. Instead we can take a hierarchical approach. In a hierarchical Bayesian model, we introduce a set of population-level hyperparameters, which define the distribution from which our individual latent variables are drawn (Gelman 2006b; Gelman & Hill 2007).

In the context of CMB component separation, we may model the pixel spectral parameters within our regions as being drawn from some underlying hyperdistributions. In our particular case we model the spectral parameters as being drawn from underlying normal distributions, parametrized by the hyperparameters, $\{(\mu_{\beta_s}, \sigma_{\beta_s}), (\mu_{\beta_d}, \sigma_{\beta_d}), (\mu_{T_d}, \sigma_{T_d})\}$. Each pair corresponds to the mean and standard deviation of the underlying normal hyperdistribution for the synchrotron spectral index, the dust spectral index, and the dust temperature, respectively. During component separation we jointly

Table 3. Priors for parameters in the hierarchical model. For the means of the spectral hyperdistributions we assign normal priors, and for the standard deviations of the hyperdistributions we assign half-normal priors. The priors for the pixel-level spectral parameters are then defined through their conditional dependence on the hyperparameters. As with the complete pooling model, we assign flat priors to the amplitude parameters and allow them to vary from pixel to pixel. We restrict our analysis to polarization.

Θ	$p(\Theta)$
μ_{β_s}	$\mathcal{N}(\mu = -3, \sigma = 0.3)$
σ_{β_s}	Half-normal ($\sigma = 0.35$)
μ_{β_d}	$\mathcal{N}(\mu = 1.6, \sigma = 0.3)$
σ_{β_d}	Half-normal ($\sigma = 0.35$)
μ_{T_d}	$\mathcal{N}(\mu = 21, \sigma = 2)$
σ_{T_d}	Half-normal ($\sigma = 2.5$)
β_s^p	$\mathcal{N}(\mu = \mu_{\beta_s}, \sigma = \sigma_{\beta_s})$
β_d^p	$\mathcal{N}(\mu = \mu_{\beta_d}, \sigma = \sigma_{\beta_d})$
T_d^p	$\mathcal{N}(\mu = \mu_{T_d}, \sigma = \sigma_{T_d})$
$A_s^{p,\lambda}$	Unif ($-\infty, \infty$)
$A_d^{p,\lambda}$	Unif ($-\infty, \infty$)
$A_{\text{cmb}}^{p,\lambda}$	Unif ($-\infty, \infty$)

fit for the population-level hyperparameters and the associated pixel-level spectral parameters. The hierarchical approach allows us to model the pixel-level variations in the spectral parameters, with the hyperdistributions reducing the propensity of the model to overreact to noise, as would be the case if we assumed total independence between pixel-level spectral parameters (Katahira 2016).

Our emission model takes the same form as in equation (19), and we again assume the Gaussian likelihood in equation (10). The priors for our hierarchical model are listed in Table 3. Analogously to the complete pooling model, we set informative priors on the spectral hyperparameters. For the means of the hyperdistributions we set the same normal priors as for the spectral parameters in the complete pooling model. For the standard deviations of the hyperdistributions we set half-normal priors, with scale parameters set to correspond to the standard deviations of the mean priors. The half-normal prior constrains the standard deviations to be positive, with the scale parameters chosen to encapsulate the likely degree of variation of spectral parameters in a given region. Setting informative priors on the hyperparameters in a hierarchical model can be particularly important in ensuring the robust computational performance of the sampling algorithm. Hyperparameters are highly correlated with the associated pixel-level parameters, and small changes in the values of the hyperparameters can induce large changes in the target distribution. This can result in funnel-like geometries in the posterior when the data are limited, i.e. a region of high density but low volume below a region of low density but high volume. The funnel regions are highly curved, which can lead to major computational difficulties during sampling, in the worst case leading to a failure in geometric ergodicity. This problem can be partly mitigated through setting informative priors that downweight more extreme parameter values as we have done here (Gelman 2006a; Betancourt & Girolami 2015).

In addition to our choice of informative priors, we reparametrize our spectral parameters by introducing the auxiliary variables,

$$\Gamma_c^p \sim \mathcal{N}(0, 1), c = \{s, d, \text{cmb}\}. \quad (20)$$

In the case of the synchrotron spectral index we may re-express β_s as

$$\beta_s^p = \mu_{\beta_s} + \Gamma_s^p \sigma_{\beta_s}, \quad (21)$$

with analogous expressions for β_d^p and T_d^p . Thus, instead of directly sampling the $\{\beta_s^p, \beta_d^p, T_d^p\}$, we instead sample a set of Gaussian latent variables and obtain the pixel-by-pixel spectral parameters through a translation and scaling with the hyperparameters. This is known as the non-centred parametrization and has the convenient effect of reducing correlations between the hyperparameters and the pixel-level spectral parameters. A detailed discussion of the geometrical pathologies of hierarchical models and practical approaches to their mitigation can be found in Betancourt & Girolami (2015).

5.5 Convergence checks

Given an infinite number of samples it can be shown the NUTS algorithm will converge on the target distribution. However, it remains important to perform a number of checks to reassure ourselves of convergence after a finite number of samples. To this end, we output a number of convergence diagnostics that we describe below.

The first convergence diagnostic we output is the Gelman–Rubin statistic (Gelman & Rubin 1992; Brooks & Gelman 1998). This compares the variance between multiple, independently initialized chains with the variance within each chain, and is defined as

$$\hat{R} = \frac{\hat{V}}{\hat{W}}, \quad (22)$$

where \hat{V} is the between-chain variance and \hat{W} is the within-chain variance. If convergence has been achieved the between-chain and within-chain variance will be equal. In reality we apply the threshold, $\hat{R} \leq 1.1$, to reassure ourselves that our chains satisfy the necessary geometric ergodicity conditions.

We also output the number of effective samples in each chain, n_{eff} (Geyer 1992; Brooks et al. 2011). When sampling from a target distribution using some MCMC algorithm, we may draw a total of N samples, but these samples are not totally independent. The effective sample size provides a measure of the number of independent samples in a chain, defined as

$$n_{\text{eff}} \equiv \frac{N}{\sum_{t=-\infty}^{\infty} \rho_t} = \frac{N}{1 + 2 \sum_{t=1}^{\infty} \rho_t}, \quad (23)$$

where ρ_t is the autocorrelation within a chain at a lag t . Details on the estimation of the autocorrelation can be found in Stan Development Team (2012).² The appropriate number of effective samples to be able to properly capture the target distribution is to some extent a question of judgement. However, in Kruschke (2011) a threshold of ~ 1000 effective samples is proposed to be confident in expectations calculated with parameter chains. As such, we adopt this as a confidence threshold for our sampling output.

Finally, we also output warnings when divergences occur during sampling. A divergence takes place when the sampler encounters a region of the target distribution where the curvature is too high to be resolved given the tuned step-size. In practice, divergences are detected when the value of the Hamiltonian diverges from its initial value when simulating trajectories through parameter space. This is significant in that divergences can mean that the conditions for geometric ergodicity are not met, and therefore using the resultant chains to construct statistical estimators can lead to biased inferences (Betancourt & Girolami 2015; Betancourt 2016, 2017).

²It is worth noting that if a chain is estimated to have a negative autocorrelation between samples, one can obtain $n_{\text{eff}} > N$.

6 ALGORITHM VALIDATION

We validate the component separation algorithm’s performance against the simulated experimental observations described in Section 3. For ease of discussion, we assign the following labels to our three validation sets (i.e. simulation and modelling runs).

- (i) CP(L): the *LiteBIRD* only analysis, fitting the complete pooling model described in Section 5.3.
- (ii) CP(LC): the C-BASS and *LiteBIRD* analysis, fitting the complete pooling model.
- (iii) H(LC): the C-BASS and *LiteBIRD* analysis, fitting the hierarchical model described in Section 5.4.

In all three validation sets we use regions defined on $N_{\text{SIDE}} = 64$ maps, as described in Section 4, for our spectral modelling. We note here that we do not fit the hierarchical model to the simulation set consisting of just *LiteBIRD* observations. It was found to be very challenging to control for the occurrence of divergences when fitting the hierarchical model to *LiteBIRD*-only observations, leaving the convergence properties of the resulting MCMC chains suspect. These problems around controlling divergences can be understood when we consider the lack of low-frequency channels in *LiteBIRD*. Given limited available information to constrain synchrotron spectral parameters, the posterior geometry for the hierarchical model becomes extremely difficult to sample. Applying the hierarchical model to *LiteBIRD*-only observations likely requires a careful study of prior choice for model hyperparameters and extended tuning phases to help mitigate the occurrence of divergences.

For each validation set, we evaluate the normalized deviations of the recovered parameters. For some parameter Θ , the normalized deviations are defined as

$$\eta_{\lambda} = \frac{\Theta^{\text{in},\lambda} - \Theta^{\text{out},\lambda}}{\sigma_{\Theta}^{\lambda}}, \quad \lambda = \{Q, U\}, \quad (24)$$

where $\Theta^{\text{out},Q/U}$ is the output parameter Q/U map, $\Theta^{\text{in},Q/U}$ is the input parameter Q/U map, and $\sigma_{\Theta}^{Q/U}$ is the corresponding Q/U standard deviation map. We note that for spectral parameters the Q and U normalized deviations will be identical. If our observed $\Theta^{\text{out},Q/U}$ are drawn from a Gaussian distribution with mean given by $\Theta^{\text{in},Q/U}$ and standard deviation given by $\sigma_{\Theta}^{Q/U}$, the normalized deviations should be distributed as a standard Gaussian, $\mathcal{N}(0, 1)$.

Histograms of the normalized deviations for the model parameters recovered from each validation set are shown in Fig. 3. In Table 4, we state the median of the normalized deviations for each parameter and validation set along with the corresponding median absolute deviation (MAD) values. The distributions of the normalized deviations for each parameter are discussed in the relevant subsections outlined below.

The outline of the remainder of this section is as follows: in Section 6.1, we discuss the CMB amplitude output; in Section 6.2, we present the synchrotron and dust amplitude constraints; in Section 6.3, we discuss constraints on the synchrotron spectral index; and in Section 6.4, we show results for the dust spectral parameters. In Fig. 4, we show maps of the recovered CMB amplitudes and effective sample size for each validation set. In Fig. 7, we show the recovered synchrotron and dust amplitude maps. In Fig. 8, we show the recovered synchrotron and dust spectral parameter maps. In addition to the results presented here, we provide figures showing residual maps for our model parameters as supplementary online material.

6.1 CMB amplitude

The primary output from the NUTS component separation is CMB amplitude maps in Q and U . This consists of a set of maps corresponding to individual posterior samples, along with the summary maps of the mean and standard deviation of the amplitude maps. In Fig. 4, we show the mean CMB Q and U amplitude maps obtained for our three validation sets, along with the associated maps of the effective sample size. When using a complete pooling model obvious artefacts can be seen in the recovered CMB amplitude maps near the Galactic plane. This is to be expected, given the bright diffuse emission in these regions makes the extraction of weak CMB signals extremely challenging. When using a hierarchical model these artefacts are no longer present. By allowing the model to account for the real variation in spectral parameters in our regions, whilst constraining this variation through the fitted hyperdistributions, we are able to achieve a more accurate foreground removal and thereby remove the biases apparent from assuming constant spectral parameters. Regions of the CMB amplitude maps containing component separation artefacts are well traced by the effective sample size. In regions of low effective sample size ($n_{\text{eff}} \lesssim 1000$), the parameter chains exhibit a high degree of autocorrelation, indicative of the sampler struggling to draw independent posterior samples. Thresholding maps of the effective sample size can be used to construct confidence masks for the CMB amplitude maps, which can be used in combination with standard Galactic emission masks.

In Fig. 3(a), we show histograms of the normalized deviations of the $A_{\text{cmb}}^{Q/U}$ maps. For the validation sets studied here the distributions of the normalized deviations are slightly wider than the standard Gaussian. This means the uncertainties reported by the CMB amplitude standard deviation maps underestimate the errors on the recovered CMB amplitudes, and the CMB amplitude posterior shows slight departures from Gaussianity. The complete pooling model results in small biases away from zero in the median of the normalized deviations, which are not present with the hierarchical model.

We perform power spectrum estimation using the NAMASTER library (Alonso et al. 2019). NAMASTER is a code for performing pseudo- C_{ℓ} estimation, accounting for effects from sky masking, as well as performing full E - and B -mode purification. This is particularly important when the B -mode signal is much weaker than the E -mode signal, as is the case for CMB B -mode studies. In this situation E -to- B leakage when performing power spectrum estimation on a cut sky can lead to the variance of the B -mode power spectrum estimators being dominated by the variance of the leaked E -modes. Details on the construction of unbiased pseudo- C_{ℓ} estimators and E - and B -mode purification can be found in Hivon et al. (2002), Elsner, Leistedt & Peiris (2017), and Alonso et al. (2019).

We generate the Galactic emission mask following the procedure in Remazeilles et al. (2018), with a common mask being used to enable direct comparison between the recovered power spectra. We extrapolated 10° smoothed 5 and 402 GHz polarized intensity maps to 70 GHz, and applied a 5σ threshold against the standard deviation of the 10° smoothed CMB polarized intensity map at 70 GHz. The 5 GHz map was extrapolated using a constant spectral index of $\beta_s = -3$ and the 402 GHz map was extrapolated using an MBB SED, setting $\beta_d = 1.6$ and $T_d = 19.4$ K. We further mask all pixels with $n_{\text{eff}} \lesssim 1000$ in either the Q or U maps, although most of these pixels are already contained within the Galactic emission masks. The Galactic emission mask produced by masking all pixels below the 5σ threshold excludes approximately 60 per cent of the sky.

In Fig. 5, we show the E - and B -mode power spectra derived for the three validation sets. For the purposes of power spectrum estimation, we perform component separation on two splits of the input data with

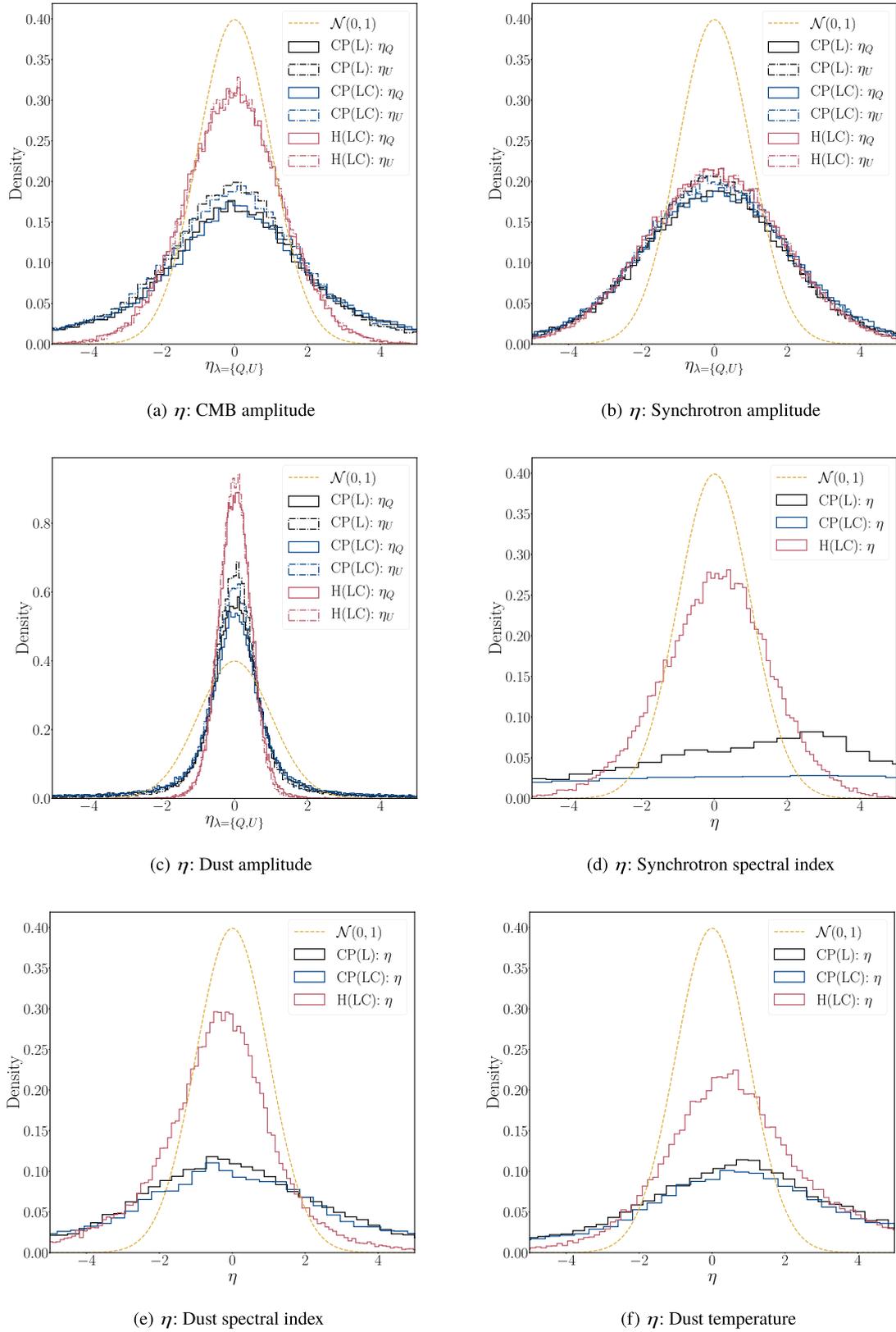


Figure 3. Histograms of the normalized deviations for our recovered parameter maps, obtained for the three validation sets. Alongside these histograms, we plot the standard Gaussian, $\mathcal{N}(0, 1)$. Mismatch between the normalized deviation histograms and $\mathcal{N}(0, 1)$ indicates departures from Gaussianity in the marginal parameter posteriors. In itself, this is not surprising, given the posterior distributions include contributions from non-Gaussian priors, and in the case of the hierarchical model contains complex correlations between hyperparameters and pixel-level parameters.

Table 4. The median of the normalized deviations for each parameter and validation set along with the corresponding median absolute deviation (MAD) values. In columns 2 and 3 we state the medians of the normalized deviations for the recovered model parameters in each validation set, and in columns 4 and 5 we state the corresponding MAD values. For spectral parameters the normalized deviations in Q and U are identical. As such, we only state results for spectral parameters under their corresponding Q columns.

Validation set	Med(η_Q)	Med(η_U)	MAD(η_Q)	MAD(η_U)
A_{cmb} : CP(L)	0.21	-0.10	2.60	2.18
A_{cmb} : CP(LC)	0.12	-0.02	2.66	2.25
A_{cmb} : H(LC)	-0.03	2×10^{-3}	1.30	1.29
A_s : CP(L)	2×10^{-3}	-0.02	2.17	1.98
A_s : CP(LC)	0.08	-2×10^{-3}	2.14	2.00
A_s : H(LC)	0.01	1×10^{-3}	1.89	1.88
A_d : CP(L)	0.01	0.01	0.74	0.62
A_d : CP(LC)	0.02	0.01	0.81	0.68
A_d : H(LC)	0.03	2×10^{-3}	0.45	0.43
β_s : CP(L)	-0.44	...	7.05	...
β_s : CP(LC)	-1.03	...	16.8	...
β_s : H(LC)	0.09	...	1.46	...
β_d : CP(L)	0.09	...	4.18	...
β_d : CP(LC)	-0.35	...	4.96	...
β_d : H(LC)	-0.35	...	1.42	...
T_d : CP(L)	0.61	...	4.33	...
T_d : CP(LC)	1.23	...	5.11	...
T_d : H(LC)	0.63	...	1.96	...

differing noise realizations. We then evaluate the cross-spectra of the recovered CMB maps from the two data splits, allowing us to avoid complications from noise bias. We show power spectra, covering multipoles $2 \leq \ell < 180$, using a bin width of $\Delta\ell = 10$ (excluding the first bin, which covers multipoles $2 \leq \ell < 10$).

For multipoles $\ell \lesssim 30$ it was found that the purified pseudo- C_ℓ estimator no longer recovered accurate point estimates for the bandpowers. At these large angular scales, the sky mask results in the pseudo- C_ℓ estimator becoming suboptimal. For these low multipoles, one can obtain better power spectrum estimates by directly sampling the C_ℓ from the joint distribution, $P(C_\ell, \mathbf{A}_{\text{cmb}} | \mathbf{d})$, as is performed in the COMMANDER component separation code (Eriksen et al. 2004, 2008; Wandelt, Larson & Lakshminarayanan 2004; Taylor, Ashdown & Hobson 2008). We leave the implementation of this additional power spectrum estimation technique to future work.

For multipoles $\ell \gtrsim 150$, the 70 arcmin beam correction resulted in higher variance pseudo- C_ℓ estimates. At these small scales the beam correction begins to inflate the noise present in the CMB amplitude maps. These effects, in combination with the pixel resolution at NSIDE = 64, result in bandpower estimates becoming unreliable for $\ell \gtrsim 180$.

To estimate the bandpower covariances, noise realizations were obtained by taking the difference between individual CMB amplitude posterior samples and the mean CMB amplitude map for the respective data splits, i.e.

$$\delta \mathbf{A}_{\text{cmb}}^{k,j,\lambda} = \mathbf{A}_{\text{cmb}}^{k,j,\lambda} - \langle \mathbf{A}_{\text{cmb}}^{j,\lambda} \rangle, \quad \lambda = \{Q, U\}, \quad (25)$$

where k denotes the k th posterior sample and $j \in \{1, 2\}$ denotes the relevant data split. The covariance matrix was then estimated by calculating the power spectra of 1000 signal plus noise simulations. For the signal simulations we generated 1000 realizations of the

theoretical input CMB power spectrum to account for cosmic variance.

The E -mode power spectrum is obtained with high accuracy for $\ell \geq 30$, confirming the overall fidelity of the component separation algorithm. The recovery of the B -mode power spectrum is more challenging, given this signal is significantly weaker than the E -mode signal and potentially subdominant to foregrounds over most of the sky, at all frequencies. It can be seen that when using a complete pooling model the recovered B -mode power spectrum contains large-scale foreground residuals, biasing the power spectrum high. In real experimental applications one could attempt to mitigate this by applying even more aggressive Galactic emission masks, although this comes at the cost of increasing the uncertainty in the recovered power spectra. For multipoles $\ell < 30$ we can see that the apparent residuals in the B -mode power spectrum are reduced when using the hierarchical model. However, as discussed above, the pseudo- C_ℓ estimates at these multipoles become suboptimal, demonstrated clearly here by the negative point estimates obtained for the E -mode bandpowers in the $10 \leq \ell < 20$ bin.

To quantify the impact of foreground residuals in the B -mode power spectrum, we can study the tensor-to-scalar ratio constraints that would be obtained from these power spectra for $\ell \geq 30$. To do this we approximate the likelihood for the CMB power spectra as a multivariate Gaussian,

$$-2 \ln \mathcal{L} = \text{const} + \sum_{\ell\ell'} (\tilde{C}_\ell - \tilde{C}_\ell^{\text{th}}) (\Sigma^{-1})_{\ell\ell'} (\tilde{C}_{\ell'} - \tilde{C}_{\ell'}^{\text{th}}). \quad (26)$$

The \tilde{C}_ℓ are the binned power spectra or bandpowers, with $\tilde{C}_\ell^{\text{th}}$ being the corresponding theoretical bandpowers, and Σ is the bandpower covariance matrix. The sum here runs over the effective ℓ values for each bin. It is worth noting that, in general, the CMB likelihood is non-Gaussian. However, for the higher multipoles we consider here the power spectrum estimates are formed by averaging over the individual $a_{\ell m}$ s corresponding to a given multipole, justifying the use of the Gaussian approximation through the central limit theorem.

We parametrize the theoretical power spectrum as

$$C_\ell^{\text{th}} = \frac{r}{0.01} C_\ell^{BB}(r = 0.01) + A_L C_\ell^{\text{lens}}, \quad (27)$$

where A_L is the lensing amplitude, $C_\ell^{BB}(r = 0.01)$ is a fiducial primordial B -mode power spectrum corresponding to $r = 0.01$, and C_ℓ^{lens} is the fiducial lensing B -mode power spectrum. In a general B -mode analysis we would fit jointly for r and A_L . However, it is challenging to constrain A_L solely through the B -mode power spectrum here. In a realistic experiment, tight constraints can be put on the lensing B -mode through analysis of the E -mode, TE , correlations, and the lensing potential power spectrum. Combined with constraints from external data sets and delensing, one can expect to be able to place a tight prior on A_L . For the sake of simplicity here, we fix A_L to the input value of 1.

Sampling from this likelihood using PYMC3, we obtain the constraints on r shown in Fig. 6. Using the complete pooling model we recover biased estimates of the tensor-to-scalar ratio, obtaining $r = (12.9 \pm 1.4) \times 10^{-3}$ for the CP(L) set and $r = (9.0 \pm 1.1) \times 10^{-3}$ for the CP(LC) set. This can be understood when we consider the large residuals and artefacts present in the recovered CMB when using a complete pooling model. By comparison, the bias is effectively removed for the H(LC) set, obtaining $r = (5.2 \pm 1.0) \times 10^{-3}$. It is worth noting that, despite the increased degrees of freedom in using a hierarchical model, the uncertainties obtained for the H(LC) set are smaller than those for the CP(LC) set. By adopting a multilevel structure for the spectral parameters in each region we

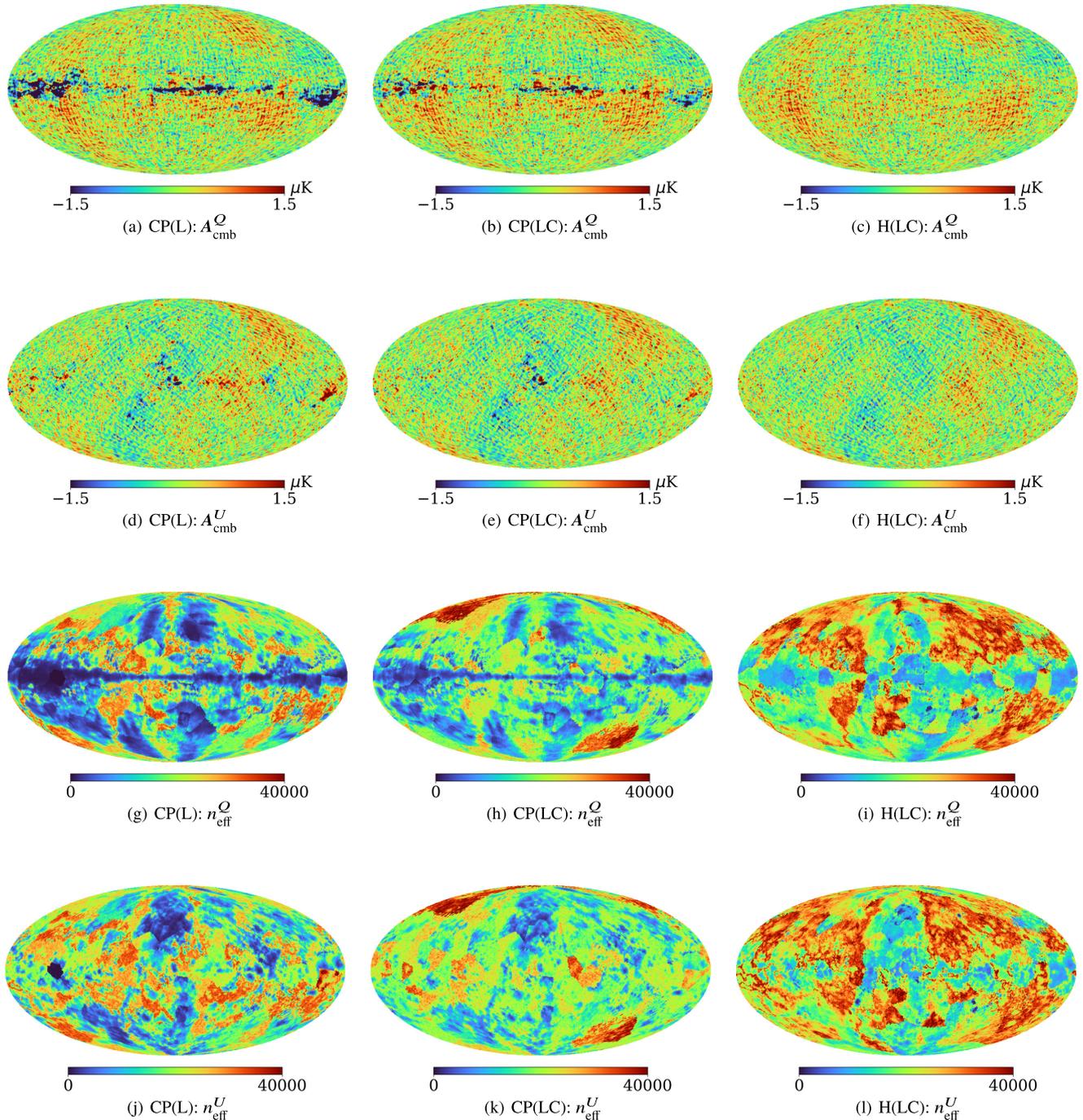


Figure 4. Output CMB amplitude maps obtained with the NUTS component separation algorithm. In panels (a)–(f), we show the output $A_{\text{cmb}}^{Q/U}$ maps for the various validation sets. In panels (g)–(l), we show the corresponding maps of the effective sample size. Obvious artefacts can be seen in the output $A_{\text{cmb}}^{Q/U}$ maps when using a complete pooling model. This is the result of extremely bright foreground emission dominating over the CMB signal, making the extraction of the CMB signal very challenging. These artefacts are well traced by the effective sample size, being most apparent in bright regions close to the Galactic plane and North Polar Spur. By using a hierarchical model we are able to remove these obvious artefacts from our recovered CMB maps.

prevent the model from overreacting to noise, whilst still capturing the spatial variation in spectral parameters. This in turn removes many of the foreground residuals present in the CMB amplitude maps obtained with the complete pooling model. These residuals result in increased uncertainties and biases on r , caused by the misspecified power spectrum model in the presence of foreground residuals.

As an aside, we do caution against interpreting these results as direct forecasts for the proposed *LiteBIRD* experiment (and/or some combination with additional experiments). *LiteBIRD* will also obtain constraints on lower multipoles around the reionization peak in the B -mode power spectrum, which we have not considered here due to the suboptimal pseudo- C_ℓ estimator used, increasing the

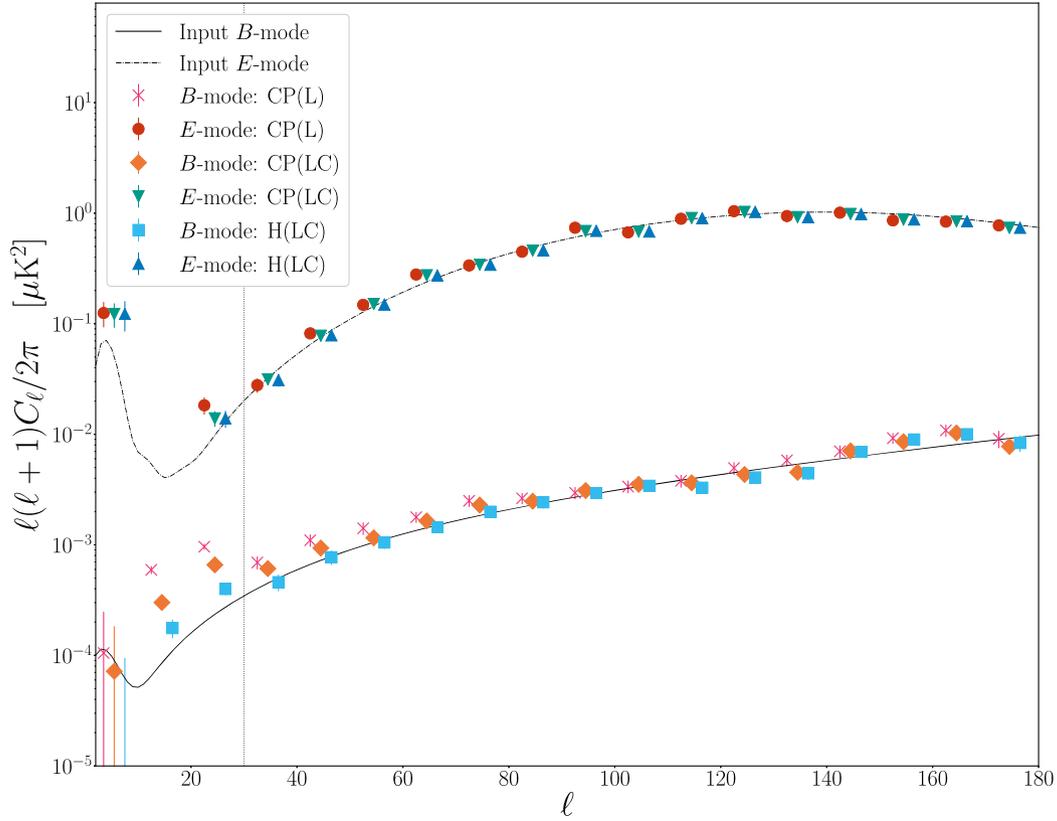


Figure 5. CMB E - and B -mode power spectra for our three validation sets, covering the multipole range, $2 \leq \ell < 180$. Power spectra are binned, using a bin width of $\Delta\ell = 10$ (excluding the first bin, which contains multipoles $2 \leq \ell < 10$). We plot the power spectra for the three validation sets together, slightly offset from one another for display purposes. The central point within each triple is at the effective multipole value for that bin. We can see that the E -mode power spectrum is recovered with high accuracy for $\ell > 30$, demonstrating the overall fidelity of the algorithm. For the B -mode power spectrum we can see the results obtained using a complete pooling model contain significant large-scale foreground residuals that are not present when using the hierarchical model. For $\ell \gtrsim 30$ the sky mask results in the pseudo- C_ℓ estimator becoming suboptimal, which is particularly apparent in the $10 \leq \ell < 20$ bin, where the point estimates for the E -mode bandpowers are negative. We plot a dotted vertical line at $\ell = 30$ to separate these angular scales. For $\ell \gtrsim 150$ the beam correction results in higher variance pseudo- C_ℓ estimates. This beam correction, along with the limiting pixel resolution at NSIDE = 64, results in bandpower estimates becoming unreliable for $\ell \gtrsim 180$.

sensitivity of any tensor-to-scalar ratio measurement. Further, for simplicity in this validation analysis we smoothed all channels to the 70 arcmin resolution of the lowest frequency *LiteBIRD* channel. This is likely a somewhat pessimistic approach. However, it is beyond the scope of this work to study the impact of the low resolution of the low-frequency *LiteBIRD* channels on the ability to recover the CMB power spectra at higher multipoles. We have also not considered the effect of mismodelling foreground SEDs or the impact of experimental systematics, both of which would significantly complicate any B -mode measurements.

As discussed previously, given the frequency coverage considered here we already struggle to constrain spatial variations in foreground spectral parameters for these simple models. Fitting more complex models will require additional low- and high-frequency data. We leave the impact of mismodelling complex foreground SEDs to more detailed forecasting analyses, focusing here on validating the ability of Bayesian hierarchical modelling to reduce biases in recovered CMB estimates without inflating parameter uncertainties. It is worth noting that we consider the performance of the hierarchical model here for an experimental frequency coverage much broader than that of next-generation ground-based experiments such as the Simons Observatory (Ade et al. 2019) and CMB-S4 (Abazajian et al. 2016). Given the reduced frequency coverage of these experiments, fitting

a full hierarchical model will be significantly more challenging, and will likely require a careful study of the prior choice for model hyperparameters, and an exploration of model reparametrizations.

6.2 Synchrotron and dust amplitudes

In Fig. 7, we show the dust and synchrotron amplitude maps. Dust amplitude maps are shown at the reference frequency of $\nu_0 = 402$ GHz for all three validation sets. Synchrotron amplitude maps are shown at a reference frequency of $\nu_0 = 40$ GHz for the CP(L) validation set, and at $\nu_0 = 5$ GHz for the CP(LC) and H(LC) validation sets. In all cases we can see the recovered component amplitude maps trace the input component amplitude maps well. However, this is to some extent a result of our choice of reference frequency, with the amplitudes being constrained by the pixel values at those frequencies. The overall level of residuals in the synchrotron amplitude maps is reduced by ~ 5 per cent for the H(LC) set compared to the CP(LC) set, and the typical residuals in the dust amplitude maps are reduced by ~ 40 per cent. These reductions were estimated using the MAD values of the residuals. Whilst the absolute value of these reductions is small at synchrotron and dust frequencies, they ultimately propagate through to significant biases in the recovered CMB maps, as seen in our tensor-to-scalar ratio estimates.

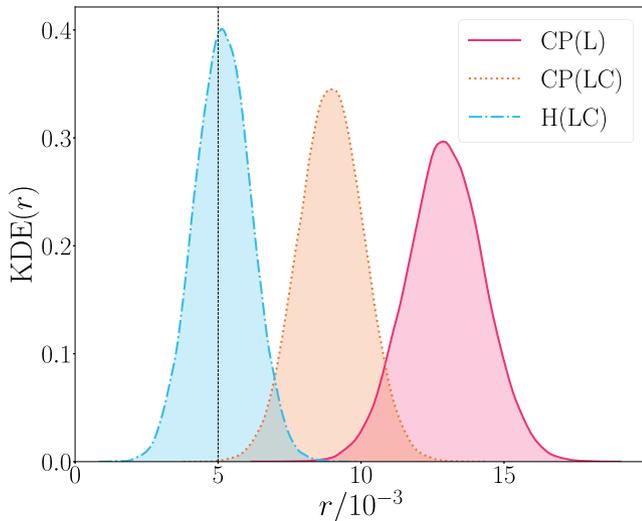


Figure 6. Tensor-to-scalar ratio constraints obtained for the three validation sets, using derived power spectra over the range $30 \leq \ell < 180$ and assuming a multivariate Gaussian likelihood. For the CP(L) set we find $r = (12.9 \pm 1.4) \times 10^{-3}$, for the CP(LC) set we find $r = (9.0 \pm 1.1) \times 10^{-3}$, and for the H(LC) set we find $r = (5.2 \pm 1.0) \times 10^{-3}$. With the addition of C-BASS we see a reduction in the bias on r with the complete pooling model, although a $\sim 3.6\sigma$ bias still remains for the CP(LC) set. This bias is removed when using a hierarchical model. The uncertainty on the measured tensor-to-scalar ratio is smaller for the hierarchical model. The hyperdistributions over foreground spectral parameters allow us to model their spatial variations without increasing parameter uncertainties. Further, foreground residuals that result from the complete pooling model mean the power spectrum model is misspecified, resulting in increased biases and uncertainties on cosmological parameters.

In Fig. 3(b), we show histograms of the normalized deviations for the synchrotron amplitude, and in Fig. 3(c), we show histograms of the normalized deviations for the dust amplitude. For the synchrotron amplitude we obtain similar distributions across all three validation sets. The normalized deviations for the synchrotron amplitudes are wider than the standard Gaussian, indicating the standard deviations of the posterior samples underestimate the uncertainties on the synchrotron amplitudes. Conversely, for the dust amplitudes the normalized deviations are narrower than the standard Gaussian.

6.3 Synchrotron spectral parameters

In Figs 8(a)–(c), we show the synchrotron spectral indices obtained for the three validation sets. In the case of the CP(L) set, we struggle to place accurate constraints on the synchrotron spectral index. This is to be expected, given the lack of low-frequency channels below 40 GHz, and is consistent with expectations from the single-pixel component separation analysis presented in Jew et al. (2019). For the CP(LC) validation set, we are able to place improved constraints on the synchrotron spectral index, with the variations in the synchrotron spectral index from region-to-region tracing the variations in the input spectral index map shown in Fig. 1.

When using a hierarchical model we are able to more finely model variations in the synchrotron spectral index across the sky. As can be seen in Fig. 8(c), the spectral index in regions of high SNR traces the same variations seen in the input synchrotron spectral index map. In regions of low SNR, away from the Galactic plane, we do see additional features not present in the input β_s map. This is a result of the noisier estimates of the synchrotron spectral index we

obtain in these regions. These noisy variations are constrained by the hyperdistribution, which penalizes individual estimates of β_s being too far from the population mean, μ_{β_s} . Further, even in these noisier regions of the sky we can see that many of the large-scale variations in the spectral index are still traced by the individual spectral indices. This is a well-known property of hierarchical models, known as posterior shrinkage, and is one of the main advantages of adopting the hierarchical approach, i.e. we obtain improved point estimates of our latent variables (Katahira 2016).

For the simulations we have considered here, the typical residuals in the synchrotron spectral index maps are similar between the CP(LC) and H(LC) validation sets. This is partly a result of the input synchrotron spectral index map being highly idealized, lacking in small-scale features (Krachmalnicoff et al. 2018). Improvements can also likely be obtained for the hierarchical model by refining the region definition, such that we define larger regions in areas of low SNR, increasing the smoothing effect of the hyperdistributions.

In Fig. 3(d), we show histograms of the normalized deviations for the synchrotron spectral index. When using a complete pooling model, the distributions of the normalized deviations are significantly wider than the standard Gaussian. This is particularly apparent for the CP(LC) set. With the addition of C-BASS, we have greater constraining power on the average spectral index across each sky region. However, normalized deviations are evaluated at the pixel-level. The uncertainties obtained for the average spectral index in a given region will not be representative of the pixel-level uncertainties. For the range of values shown in Fig. 3(d), the CP(L) set appears to show a positive bias. However, the median is driven to -0.44 by a set of large magnitude, negative normalized deviations. Given the lack of low-frequency channels below $\nu = 40$ GHz, *LiteBIRD* data alone are unable to properly constrain synchrotron spectral parameters.

6.4 Dust spectral parameters

Most of the constraining power for dust spectral parameters comes from high-frequency channels, i.e. $\nu \gtrsim 100$ GHz. In our three validation sets these remained identical, being the high-frequency *LiteBIRD* channels. In Jew et al. (2019), it was found that, given a *LiteBIRD*-like frequency coverage, it is difficult to constrain dust spectral parameters. Indeed, this was the case for our own analysis here, where informative priors were needed on the dust spectral parameters. The difficulty in constraining dust spectral parameters is reflected in their low effective sample size, which was typically $\lesssim 1000$ over much of the sky in all three validation sets.

In Figs 8(d)–(f), we show the recovered dust spectral index maps; and in Figs 8(g)–(i), we show the recovered dust temperature maps. In using a hierarchical model for the dust spectral parameters, we are able to fit for variations in the spectral parameters in regions of high SNR close to the Galactic plane. Away from the Galactic plane, the individual variations become much smaller in each region, with the resulting maps of β_d and T_d very obviously tracing out the crude structure of the regions used in the component separation. Given the limited frequency coverage, there is simply not enough information to constrain the low-level variations of the dust spectral parameters in each region. In this case the marginal distributions for σ_{β_d} and σ_{T_d} have a large fraction of their probability mass close to zero, constraining individual spectral parameters to be very close to their population means, μ_{β_d} and μ_{T_d} . The typical residuals for the dust spectral index maps are reduced by ~ 50 per cent for the H(LC) set compared to the CP(LC) set. The typical residuals in the recovered dust temperature maps are ~ 25 per cent lower for the H(LC) set compared to the CP(LC) set. By allowing for some of the low-level

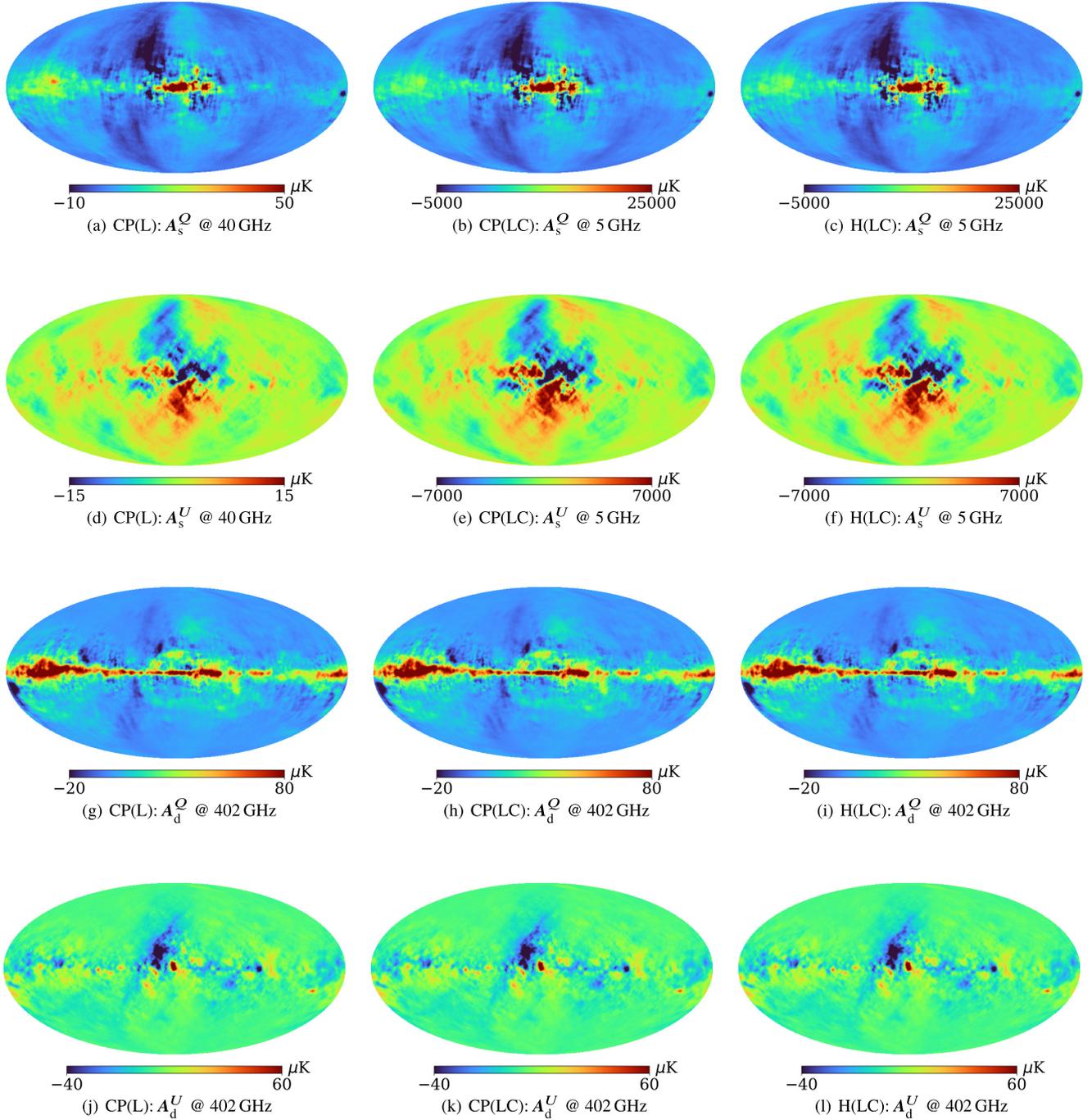


Figure 7. Synchrotron and dust amplitude maps obtained for the three validation sets. In panels (a)–(f), we show the synchrotron amplitude maps, shown at the reference frequency of $\nu_0 = 40$ GHz for the CP(L) validation set, and $\nu_0 = 5$ GHz for the CP(LC) and H(LC) validation sets. In panels (g)–(l), we show the dust amplitude maps, displayed at the reference frequency of $\nu_0 = 402$ GHz. In all three cases the component separation algorithm recovers accurate estimates of the foreground amplitude maps at their reference frequencies. This is to be expected, and is to some extent a result of our choice of synchrotron and dust reference frequencies. The typical residuals in synchrotron amplitude maps are ~ 5 per cent lower for the H(LC) set compared to the CP(LC) set. Similarly, the dust amplitude residuals are ~ 40 per cent lower for the H(LC) set compared to the CP(LC) set.

variations in the dust spectral parameters we are able to obtain smaller residuals in the recovered parameter maps, propagating through to reduced biases in the CMB amplitude estimates.

In Fig. 3(e), we show histograms of the normalized deviations for the dust spectral index; and in Fig. 3(f), we show histograms of the normalized deviations for the dust temperature. For both the dust

spectral index and the dust temperature we find biases in the medians of the normalized deviations. The MAD values obtained with a complete pooling model are significantly larger than one. This is again a result of the fact that the uncertainty obtained on the average spectral parameters in a region is not representative of the pixel-level uncertainty in those parameters. The distributions of normalized

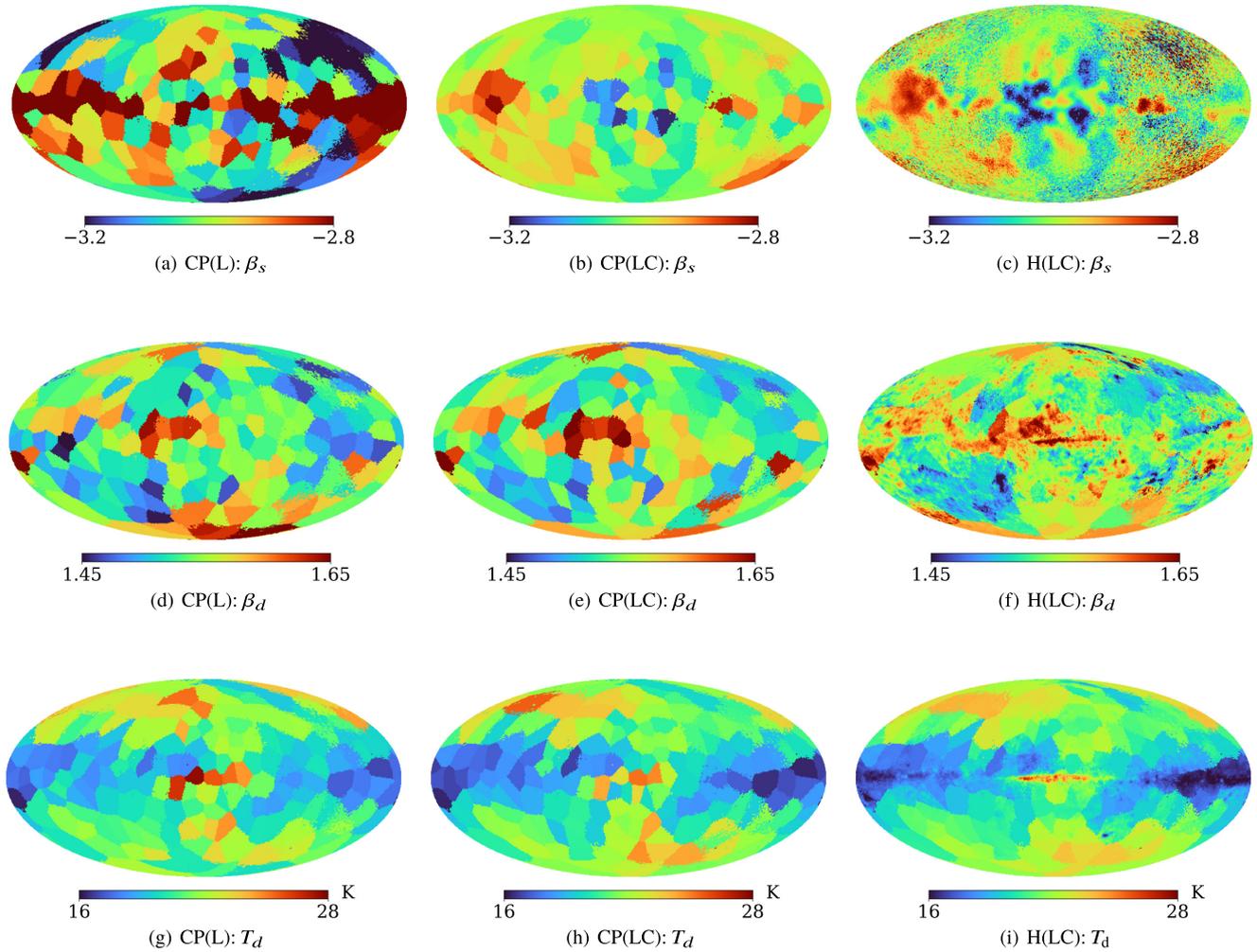


Figure 8. Spectral parameter maps obtained for the three validation sets. Panels (a)–(c): recovered synchrotron spectral index maps. Panels (d)–(f): recovered dust spectral index maps. Panels (g)–(i): recovered dust temperature maps. For the CP(L) set, we struggle to accurately constrain the synchrotron spectral index, due to the absence of low-frequency channels below 40 GHz. With the addition of a 5 GHz C-BASS channel for the CP(LC) and H(LC) sets, we are able to place improved constraints on the synchrotron spectral index. Given the frequency coverage of *LiteBIRD* we struggle to constrain dust spectral parameters for all three validation sets, reflected in low effective sample sizes for the dust spectral parameters. The hierarchical model does allow us to capture some of the pixel-level variations in our spectral parameters. Improvements in the point estimates obtained with the hierarchical model can likely be achieved by defining larger regions in areas of low SNR, increasing the smoothing effect of the hyperdistributions. The typical residuals for the synchrotron spectral index maps are comparable between the CP(LC) and H(LC) sets. However, the performance of the CP(LC) set is likely exaggerated by the lack of small-scale features in the input synchrotron spectral index map. The typical residuals for the dust spectral index maps are reduced by ~ 50 per cent for the H(LC) set compared to the CP(LC) set. The typical residuals in the recovered dust temperature maps are ~ 25 per cent lower for the H(LC) set compared to the CP(LC) set. The simulated dust spectral parameter maps contain additional small-scale features compared to the synchrotron spectral index map.

deviations for the H(LC) set are closer to the standard Gaussian than for the CP(LC) and CP(L) sets. Improved point estimates could potentially be obtained by conducting a more detailed exploration of informative priors on the hyperparameters of dust spectral parameters, or considering model reparametrizations. However, the hierarchical model is still limited by the lack of additional frequency channels at $\nu > 402$ GHz.

7 CONCLUSIONS

We have developed a new implementation of Bayesian CMB component separation, using the NUTS algorithm to draw samples from the full posterior distribution. The NUTS algorithm is a self-tuning variant of HMC that avoids the random walk behaviour that leads

to slow convergence when using standard Metropolis–Hastings and Gibbs sampling algorithms. Measured against the rate of effective sample generation, NUTS offers performance improvements of $\sim 10^3$ compared to Metropolis–Hastings when fitting the complete pooling model. Geometrical pathologies typical of hierarchical models often make variants of HMC the only reliable option for the diagnosis of divergences and biased inferences (Betancourt & Girolami 2015).

We apply this component separation algorithm to simulations of the *LiteBIRD* and C-BASS experiments to validate the algorithm performance and fidelity. These simulations use a tensor-to-scalar ratio of $r = 5 \times 10^{-3}$ and a lensing amplitude of $A_L = 1$. Component separation is performed over a set of separate sky regions, defined using the mean shift algorithm. These clusters sky regions according to the similarity in their synchrotron and dust spectral

properties, and their location on the sky. In each region we adopt two different modelling approaches, namely complete pooling and a hierarchical model. In the complete pooling model we assume the spectral parameters in each region are constant. In the hierarchical model we assume spectral parameters are drawn from underlying Gaussian distributions, fitting for the hyperparameters defining the mean and variance of the Gaussian hyperdistributions, along with the individual pixel-by-pixel spectral parameters constrained by these hyperdistributions.

When using the complete pooling model we are able to recover accurate estimates of the CMB over much of the sky. However, component separation artefacts are present close to the Galactic plane where the CMB is highly subdominant to foregrounds. Using the hierarchical model, these artefacts are removed from the recovered CMB. Estimating the CMB power spectra with these maps, we find the complete pooling model induces large-scale foreground residuals in the recovered power spectra. Using multipoles in the range $30 \leq \ell \leq 180$ and fixing the lensing amplitude, we are able to translate these power spectrum estimates into tensor-to-scalar ratio constraints. With only *LiteBIRD* frequency channels, and using the complete pooling model, we find $r = (12.9 \pm 1.4) \times 10^{-3}$. Applying the complete pooling model with an additional C-BASS channel at 5 GHz we find $r = (9.0 \pm 1.1) \times 10^{-3}$, and using the hierarchical model with C-BASS and *LiteBIRD* we find $r = (5.2 \pm 1.0) \times 10^{-3}$. The addition of C-BASS reduces the bias in the recovered tensor-to-scalar ratio for the complete pooling model. However, the crude assumptions made regarding the behaviour of spectral parameters still leaves a $\sim 3.6\sigma$ bias in the estimate of r . We find that the hierarchical model offers an effective generative approach to the modelling of spectral parameters that helps to mitigate the propensity for fitting outliers when assuming total independence between spectral parameters. The reduced foreground residuals in the recovered CMB maps also means that the hierarchical model does not suffer from the same issues around model misspecification at the power spectrum level, which results in increased uncertainties and biases on recovered cosmological parameters for the complete pooling model.

For the analysis in this paper we have developed a simple proof-of-concept implementation of our algorithm. Potential future work includes the extension of the algorithm to allow for the joint fitting of dipole and monopole corrections, along with instrumental effects such as bandpass corrections. It would also be prudent to extend the power spectrum estimation, to include the direct joint sampling of CMB amplitudes and power spectra at low multipoles, and expand the code to allow for multiresolution analyses. From the perspective of modelling, one may also consider more sophisticated approaches to clustering. For the purposes of validation in this study we used naive spectral indices as tracers of the foreground spectral properties, which will be contaminated by noise. In future work, it would be worthwhile examining improvements in region definition by using more sophisticated estimators of foregrounds spectral properties, along with studying the optimal data sets to be used as foreground templates. Finally, significant gain can potentially be made by exploring optimizations to the component separation code e.g. through reparametrizations and GPU acceleration.

ACKNOWLEDGEMENTS

The authors would like to thank David Alonso for useful comments on a draft of this paper. RDPG acknowledges support from an STFC studentship. CD acknowledges support from an ERC Starting (Consolidator) Grant (no. 307209) and an STFC Consolidated Grant (ST/P000649/1).

DATA AVAILABILITY

The simulated data used in this study were generated using the PYSM software package, which can be found at <https://pysm-public.readthedocs.io/en/latest/>

REFERENCES

- Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
- Ade P. et al., 2019, *J. Cosmol. Astropart. Phys.*, 02, 056
- Alonso D., Dunkley J., Thorne B., Naess S., 2017, *Phys. Rev. D*, 95, 043504
- Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration, 2019, *MNRAS*, 484, 4127
- Bennett C. L. et al., 2013, *ApJS*, 208, 20
- Betancourt M., 2016, preprint (arXiv:1604.00695)
- Betancourt M., 2017, preprint (arXiv:1701.02434)
- Betancourt M. J., Girolami M., 2015, in Upadhyay S. K., Singh U., Dey D. K., Loganathan A., eds, *Current Trends in Bayesian Methodology with Applications*. Chapman and Hall/CRC, New York, p. 79
- Betancourt M., Byrne S., Livingstone S., Girolami M., 2017, *Bernoulli*, 23, 2257
- Betoule M., Pierpaoli E., Delabrouille J., Le Jeune M., Cardoso J. F., 2009, *A&A*, 503, 691
- BICEP2 Collaboration et al., 2018, *Phys. Rev. Lett.*, 121, 221301
- BICEP2/Keck Collaboration et al., 2015, *Phys. Rev. Lett.*, 114, 101301
- Brooks S. P., Gelman A., 1998, *J. Comput. Graphical Stat.*, 7, 434
- Brooks S., Gelman A., Jones G., Meng X.-L., 2011, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, Boca Raton, FL
- Carpenter B. et al., 2017, *J. Stat. Software*, 76, 1
- Chluba J., Hill J. C., Abitbol M. H., 2017, *MNRAS*, 472, 1195
- Comaniciu D., Meer P., 2002, *IEEE Trans. Pattern Analysis Machine Intelligence*, 24, 603
- Davies R. D., Dickinson C., Banday A. J., Jaffe T. R., Górski K. M., Davis R. J., 2006, *MNRAS*, 370, 1125
- Delabrouille J., Cardoso J. F., Patanchon G., 2003, *MNRAS*, 346, 1089
- Dickinson C. et al., 2018, *New Astron. Rev.*, 80, 1
- Dodelson S., 2003, *Modern Cosmology*. Academic Press, New York
- Draine B. T., Hensley B. S., 2016, *ApJ*, 831, 59
- Duane S., Kennedy A., Pendleton B. J., Roweth D., 1987, *Phys. Lett. B*, 195, 216
- Dunkley J. et al., 2009, in Dodelson S. et al., eds, *AIP Conf. Proc. Vol. 1141, CMB Polarization Workshop: Theory and Foregrounds: CMBPol Mission Concept Study*. Am. Inst. Phys., New York, p. 222
- Elsner F., Leistedt B., Peiris H. V., 2017, *MNRAS*, 465, 1847
- Eriksen H. K. et al., 2004, *ApJS*, 155, 227
- Eriksen H. K., Jewell J. B., Dickinson C., Banday A. J., Górski K. M., Lawrence C. R., 2008, *ApJ*, 676, 10
- Errard J., Stivoli F., Stompor R., 2011, *Phys. Rev. D*, 84, 063005
- Errard J., Feeney S. M., Peiris H. V., Jaffe A. H., 2016, *J. Cosmol. Astropart. Phys.*, 03, 052
- Evans M., Jang G. H., 2011, *Stat. Sci.*, 26, 423
- Fixsen D. J., 2009, *ApJ*, 707, 916
- Foreman-Mackey D., Barentsen G., Barclay T., 2019, *dfm/exoplanet: exoplanet v0.1.4*, doi:10.5281/zenodo.2561395
- Ganga K. et al., 2019, *BAAS*, 51, 111
- Gelfand A. E., Smith A. F. M., 1990, *J. Am. Stat. Assoc.*, 85, 398
- Gelman A., 2006a, *Bayesian Analysis*, 1, 515
- Gelman A., 2006b, *Technometrics*, 48, 432
- Gelman A., Hennig C., 2017, *J. R. Stat. Soc. Ser. A (Stat. Soc.)*, 180, 967
- Gelman A., Hill J., 2007, *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge Univ. Press, Cambridge
- Gelman A., Rubin D. B., 1992, *Stat. Sci.*, 7, 457
- Gelman A., Jakulin A., Pittau M. G., Su Y.-S., 2008, *Ann. Appl. Stat.*, 2, 1360
- Gelman A., Simpson D., Betancourt M., 2017, *Entropy*, 19, 555
- Geyer C. J., 1992, *Stat. Sci.*, 7, 473
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759

- Guzmán A. E., May J., Alvarez H., Maeda K., 2011, *A&A*, 525, A138
- Hanany S. et al., 2019, *BAAS*, 51, 194
- Harwood J. J. et al., 2016, *MNRAS*, 458, 4443
- Hastings W. K., 1970, *Biometrika*, 57, 97
- Hensley B. S., Bull P., 2018, *ApJ*, 853, 127
- Hill-Valler J., 2019, PhD thesis, Univ. Oxford
- Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, *ApJ*, 567, 2
- Hoffman M. D., Gelman A., 2014, *J. Machine Learning Res.*, 15, 1593
- Irfan M. O., Bobin J., Miville-Deschênes M.-A., Grenier I., 2019, *A&A*, 623, A21
- Jeffreys H., 1946, *Proc. R. Soc. Lond. Ser. A*, 186, 453
- Jeffreys H., 1961, *Theory of Probability*, 3rd edn. Clarendon Press, Oxford
- Jew L., Grumitt R. D. P., 2020, *MNRAS*, 495, 578
- Jew L. et al., 2019, *MNRAS*, 490, 2958
- Jones M. E. et al., 2018, *MNRAS*, 480, 3224
- Kamionkowski M., Kovetz E. D., 2016, *ARA&A*, 54, 227
- Katahira K., 2016, *J. Math. Psychology*, 73, 37
- Keating B., Timbie P., Polnarev A., Steinberger J., 1998, *ApJ*, 495, 580
- Khatri R., 2019, *J. Cosmol. Astropart. Phys.*, 02, 039
- Kogut A. et al., 2011, *J. Cosmol. Astropart. Phys.*, 07, 025
- Krachmalnicoff N. et al., 2018, *A&A*, 618, A166
- Kruschke J. K., 2011, *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press, Burlington, MA
- Lawson K. D., Mayer C. J., Osborne J. L., Parkinson M. L., 1987, *MNRAS*, 225, 307
- Leach S. M. et al., 2008, *A&A*, 491, 597
- Leimkuhler B. J., Reich S., 2004, *Simulating Hamiltonian Dynamics*. Cambridge Univ. Press, Cambridge
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Macellari N., Pierpaoli E., Dickinson C., Vaillancourt J. E., 2011, *MNRAS*, 418, 888
- Martínez-González E., Diego J. M., Vielva P., Silk J., 2003, *MNRAS*, 345, 1101
- Millea M., Anderes E., Wandelt B. D., 2020, preprint ([arXiv:2002.00965](https://arxiv.org/abs/2002.00965))
- Mitton S., Ryle M., 1969, *MNRAS*, 146, 221
- Miville-Deschênes M. A., Ysard N., Lavabre A., Ponthieu N., Macías-Pérez J. F., Aumont J., Bernard J. P., 2008, *A&A*, 490, 1093
- Monnahan C. C., Thorson J. T., Branch T. A., 2017, *Methods Ecology Evolution*, 8, 339
- Neal R., 2012, in Brooks S., Gelman A., Jones G., Meng X.-L., eds, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, Boca Raton, FL, p. 113
- Pedregosa F. et al., 2011, *J. Machine Learning Res.*, 12, 2825
- Planck Collaboration XII, 2014, *A&A*, 571, A12
- Planck Collaboration XIX, 2015, *A&A*, 576, A104
- Planck Collaboration IX, 2016, *A&A*, 594, A9
- Planck Collaboration X, 2016, *A&A*, 594, A10
- Planck Collaboration XXV, 2016, *A&A*, 594, A25
- Planck Collaboration XI, 2018, preprint ([arXiv:1801.04945](https://arxiv.org/abs/1801.04945))
- Platania P., Burigana C., Maino D., Caserini E., Bersanelli M., Cappellini B., Mennella A., 2003, *A&A*, 410, 847
- Polson N. G., Scott J. G., 2012, *Bayesian Analysis*, 7, 887
- Reich P., Reich W., 1988, *A&AS*, 74, 7
- Remazeilles M., 2018, preprint ([arXiv:1806.01026](https://arxiv.org/abs/1806.01026))
- Remazeilles M., Delabrouille J., Cardoso J.-F., 2011, *MNRAS*, 418, 467
- Remazeilles M., Dickinson C., Eriksen H. K. K., Wehus I. K., 2016, *MNRAS*, 458, 2032
- Remazeilles M. et al., 2018, *J. Cosmol. Astropart. Phys.*, 04, 023
- Rybicki G. B., Lightman A. P., 1985, *Radiative Processes in Astrophysics*. Wiley-VCH Verlag, Weinheim, Germany
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Comput. Sci.*, 2, e55
- Sekimoto Y. et al., 2018, *Proc. SPIE*, 10698, 106981Y
- Seljebotn D. S., Bærland T., Eriksen H. K., Mardal K. A., Wehus I. K., 2019, *A&A*, 627, A98
- Shandera S. et al., 2019, *BAAS*, 51, 338
- Simpson D., Rue H., Riebler A., Martins T. G., Sørbye S. H., 2017, *Stat. Sci.*, 32, 1
- Stan Development Team, 2012, *Stan Modeling Language User's Guide and Reference Manual*, Version 1.0. Available at: <http://mc-stan.org/>
- Stivoli F., Grain J., Leach S. M., Tristram M., Baccigalupi C., Stompor R., 2010, *MNRAS*, 408, 2319
- Stompor R., Leach S., Stivoli F., Baccigalupi C., 2009, *MNRAS*, 392, 216
- Stompor R., Errard J., Poletti D., 2016, *Phys. Rev. D*, 94, 083526
- Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, *MNRAS*, 389, 1284
- Thorne B., Dunkley J., Alonso D., Naess S., 2017, *MNRAS*, 469, 2821
- Thorne B. et al., 2019, preprint ([arXiv:1905.08888](https://arxiv.org/abs/1905.08888))
- Vidal M., Dickinson C., Davies R. D., Leahy J. P., 2015, *MNRAS*, 452, 656
- Wagner-Carena S., Hopkins M., Diaz Rivero A., Dvorkin C., 2020, *MNRAS*, 494, 1507
- Wandelt B. D., Larson D. L., Lakshminarayanan A., 2004, *Phys. Rev. D*, 70, 083511

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Table S1 In columns 3 and 4 we state the medians of the residuals for the given parameter and validation set, and in columns 5 and 6 we state the corresponding MAD values.

Figure S1 CMB amplitude residual maps for the three validation sets.

Figure S2 In panels (a)–(f) we show the synchrotron amplitude residual maps, and in panels (g)–(l) we show the dust amplitude residual maps.

Figure S3 Spectral parameter residual maps obtained for each validation set.

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 $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.