

COSMOGLOBE: Towards end-to-end CMB cosmological parameter estimation without likelihood approximations

J. R. Eskilt^{1,2}, K. Lee¹, D. J. Watts¹, V. Anshul³, R. Aurlien¹, A. Basyrov¹, M. Bersanelli⁴, L. P. L. Colombo⁴, H. K. Eriksen¹, K. S. F. Fornazier⁵, U. Fuskeland¹, M. Galloway¹, E. Gjerløw¹, L. T. Hergt⁶, H. T. Ihle¹, J. G. S. Lunde¹, A. Marins⁵, S. K. Nerval^{7,8}, S. Paradiso^{9,10}, F. Rahman¹¹, M. San¹, N.-O. Stutzer¹, and I. K. Wehus¹

¹ Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway
e-mail: j.r.eskilt@astro.uio.no

² Imperial Centre for Inference and Cosmology, Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK

³ Department of Physics, Indian Institute of Technology (BHU), Varanasi 221005, India

⁴ Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria, 16, Milano, Italy

⁵ Instituto de Física, Universidade de São Paulo, 66318, CEP 05315-970, São Paulo, Brazil

⁶ Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T1Z1, Canada

⁷ David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

⁸ Dunlap Institute for Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

⁹ Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

¹⁰ Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

¹¹ Indian Institute of Astrophysics, Koramangala II Block, Bangalore 560034, India

Received 4 July 2023 / Accepted 24 August 2023

ABSTRACT

We implement support for a cosmological parameter estimation algorithm in `Commander` and quantify its computational efficiency and cost. For a semi-realistic simulation similar to *Planck* LFI 70 GHz, we find that the computational cost of producing one single sample is about 20 CPU-hours and that the typical Markov chain correlation length is ~ 100 samples. The net effective cost per independent sample is ~ 2000 CPU-hours, in comparison with all low-level processing costs of 812 CPU-hours for *Planck* LFI and WMAP in COSMOGLOBE Data Release 1. Thus, although technically possible to run already in its current state, future work should aim to reduce the effective cost per independent sample by one order of magnitude to avoid excessive runtimes, for instance through multi-grid preconditioners and/or derivative-based Markov chain sampling schemes. This work demonstrates the computational feasibility of true Bayesian cosmological parameter estimation with end-to-end error propagation for high-precision CMB experiments without likelihood approximations, but it also highlights the need for additional optimizations before it is ready for full production-level analysis.

Key words. polarization – cosmic background radiation – cosmology: observations

1. Introduction

High-precision cosmic microwave background (CMB) measurements currently provide the strongest constraints on a wide range of key cosmological parameters (e.g., [Planck Collaboration VI 2020](#)). Traditionally, these constraints are derived by first compressing the information contained in the high-dimensional raw time-ordered data into pixelized sky maps and angular power spectra (e.g., [Bennett et al. 2013](#); [Planck Collaboration II 2020](#); [Planck Collaboration III 2020](#)). This compressed dataset is typically summarized in terms of a low-dimensional power spectrum-based likelihood from which cosmological parameters are derived through Markov chain Monte Carlo (MCMC) sampling ([Lewis & Bridle 2002](#); [Planck Collaboration V 2020](#)).

A key element in this pipeline procedure is the so-called likelihood function, $\mathcal{L}(C_\ell)$, where C_ℓ denotes the angular CMB power spectrum (e.g., [Planck Collaboration V 2020](#)). This function summarizes both the best-fit values of all power spectrum coefficients and their covariances and correlations. The accuracy of the resulting cosmological parameters, both in terms

of best-fit values and their uncertainties, corresponds directly to the accuracy of this likelihood function. As such, this function must account for both low-level instrumental effects (such as correlated noise and calibration uncertainties) and high-level data selection and component separation effects. In general, this function is neither factorizable nor Gaussian, and, except for a few well known special cases, it has no analytical exact and closed form. Unsurprisingly, significant efforts have therefore been devoted to establishing computationally efficient and accurate approximations. Perhaps the most well known example of this is a standard multivariate Gaussian distribution in units of the angular power spectrum of the sky signal, C_ℓ , with a covariance matrix tuned by end-to-end simulations (e.g., [Planck Collaboration V 2020](#)). A second widely used case is that of a low-resolution multivariate Gaussian defined in pixel space with a dense pixel-pixel noise covariance matrix ([Hinshaw et al. 2013](#); [Planck Collaboration V 2020](#)). Yet other examples include log-normal or Gaussian-plus-lognormal distributions (e.g., [Verde et al. 2003](#)), as well as various hybrid combinations (e.g., [Gjerløw et al. 2013](#)).

However, as the signal-to-noise ratios of modern CMB data sets continue to increase, the relative importance of systematic uncertainties grows, and they are often highly nontrivial to describe through low-dimensional and simplified likelihood approximations, both in terms of computational expense and accuracy. These difficulties were anticipated more than two decades ago, and an alternative end-to-end Bayesian approach was proposed independently by Jewell et al. (2004) and Wandelt et al. (2004). The key aspect of this framework is that the analysis is global, meaning all aspects of the data are modeled and fitted simultaneously, as opposed to compartmentalized, which is the more traditional CMB pipeline approach. Technically speaking, this is implemented in the form of an MCMC sampler in a process that is statistically analogous to the very last cosmological parameter estimation step in a traditional pipeline, such as CosmoMC (Lewis & Bridle 2002) or Cobaya (Torrado & Lewis 2021), with the one fundamental difference that all parameters in the entire analysis pipeline are sampled over jointly within the MCMC sampler. Thus, while a standard cosmological parameter code typically handles a few dozen, or perhaps a hundred, free parameters, the global approach handles billions of free parameters. One practical method of dealing with such a large number of parameters of different types is Gibbs sampling, which draws samples from a joint distribution by iterating over all relevant conditional distributions.

In principle, a global approach is preferable, since better results are in general obtained by fitting correlated parameters jointly rather than separately. However, and equally clearly, this approach is also organizationally and technically more complicated to implement in practice, simply because all aspects of the analysis have to be accounted for simultaneously within one framework. One example of this type of approach is Commander3 (Galloway et al. 2023), which is an end-to-end CMB Gibbs sampler developed by the BeyondPlanck Collaboration (2023) to reanalyze the *Planck* Low Frequency Instrument (LFI) observations. To date, using Gibbs sampling within this framework is the only way to sample over the entire joint distribution of low-level instrumental parameters, component maps, and cosmological parameters. This work was subsequently superseded by the COSMOGLOBE project¹, which is a community-wide open science collaboration that ultimately aims to perform the same type of analysis for all available state-of-the-art large-scale radio, microwave, and submillimeter experiments to then use the resulting analysis to derive one global model of the astrophysical sky. The first COSMOGLOBE Data Release (DR1) was made public in March 2023 (Watts et al. 2023), and included the first joint end-to-end analysis of both *Planck* LFI and the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al. 2013), resulting in lower systematic residuals in both experiments.

There are many other methods of Bayesian inference that can be used to estimate cosmological parameters from CMB maps. For example, Anderes et al. (2015) show how to efficiently sample the joint-posterior of the CMB lensing potential and sky map using Hamiltonian Monte Carlo techniques, and Carron (2019) created a framework to measure the tensor-to-scalar ratio using an approximate, analytic marginalization of lensed CMB maps. The marginal unbiased score expansion method has also been shown to be a powerful algorithm for high-dimensional CMB analyses (Millea & Seljak 2022).

The numerical challenges involved in implementing cosmological parameter estimation without likelihood approxima-

tions in Gibbs sampling were originally explored and partially resolved by Jewell et al. (2009) and Racine et al. (2016). A similar approach was taken by Millea et al. (2020), who derived a re-parameterization of the unlensed sky map and the lensing potential to improve the runtime of Gibbs sampling, similar to the method of Racine et al. (2016). This framework was later enhanced in Millea et al. (2021) to include several systematic parameters and was applied to SPTpol data.

In principle, each of the aforementioned methods could be used as the CMB map-to-parameter sampling step within a larger time-ordered-data-to-parameter end-to-end Gibbs sampler. The goal of this work is to focus on the algorithm of Racine et al. (2016) and implement it in Commander3. We will apply it to simulations with realistic noise and sky coverages, validate the implementation and evaluate its performance. Until now, the highest-level output from Commander3 has been the angular CMB power spectrum. With the code development that led to this paper, it is now technically able to also directly output cosmological parameters.

The rest of the paper is organized as follows. In Sect. 2 we provide a brief review of both the COSMOGLOBE Gibbs sampler method in general and the specific cosmological parameter sampler described by Racine et al. (2016). In Sect. 3 we validate our implementation by comparing it to two special cases, namely (i) Cobaya coupled to a uniform and (ii) full-sky case and a special-purpose Python sampler for a case with uniform noise but a constant latitude Galactic mask. We also characterize the computational costs of the new sampling step for various data configurations. Finally, we summarize and conclude in Sect. 4.

2. Algorithms

We start by briefly reviewing the Bayesian framework for global end-to-end analysis as described by the BeyondPlanck Collaboration (2023) and Watts et al. (2023) and the cosmological parameter estimation algorithm proposed by Racine et al. (2016), and discuss how these can be combined in Commander3 (Galloway et al. 2023).

2.1. Bayesian end-to-end CMB analysis

The main algorithmic goal of the COSMOGLOBE framework is to derive a numerical representation of the full posterior distribution $P(\omega | \mathbf{d})$, where ω is the set of all free parameters and \mathbf{d} represents all available data. In this notation, ω simultaneously accounts for instrumental, astrophysical, and cosmological parameters, which are typically explicitly defined by writing down a signal model, for instance taking the form

$$\mathbf{d} = \mathbf{s}(\omega) + \mathbf{n}, \quad (1)$$

where \mathbf{d} is the data vector; $\mathbf{s}(\omega)$ represents some general data model with free parameters ω ; and \mathbf{n} is instrumental noise. In practice, one typically assumes that the noise is Gaussian distributed with a covariance matrix \mathbf{N} , and the likelihood can therefore be written as

$$\mathcal{L}(\omega) = P(\mathbf{d} | \omega) \propto e^{-\frac{1}{2}(\mathbf{d}-\mathbf{s}(\omega))^T \mathbf{N}^{-1}(\mathbf{d}-\mathbf{s}(\omega))}. \quad (2)$$

The full posterior distribution reads $P(\omega | \mathbf{d}) \propto \mathcal{L}(\omega)P(\omega)$, where $P(\omega)$ represents some set of user-defined priors.

As a concrete example, COSMOGLOBE Data Release 1 (Watts et al. 2023) adopted the following data model for the time-ordered data to describe *Planck* LFI and WMAP,

$$\mathbf{d} = \text{GPB} \mathbf{M} \mathbf{a} + \mathbf{s}^{\text{orb}} + \mathbf{s}^{\text{fsl}} + \mathbf{s}^{\text{inst}} + \mathbf{n}^{\text{corr}} + \mathbf{n}^{\text{w}}, \quad (3)$$

¹ <https://cosmoglobe.uio.no>

where \mathbf{G} is a time-dependent gain factor; \mathbf{P} is a pointing matrix; \mathbf{B} denotes instrumental beam convolution; \mathbf{M} is a mixing matrix that describes the amplitude of a given sky component at a given frequency; \mathbf{a} describes the amplitude of each component at each point in the sky; \mathbf{s}^{orb} is the orbital CMB dipole; \mathbf{s}^{fsl} is a far sidelobe contribution; \mathbf{s}^{inst} is an instrument-specific correction term; \mathbf{n}^{corr} denotes correlated noise; and \mathbf{n}^{w} denotes white noise. This expression directly connects important instrumental effects (e.g., gains, beams, and correlated noise) with the astrophysical sky (e.g., \mathbf{M} and \mathbf{a}), and provides a well-defined model for the data at the low-level; this approach is what enables global modeling.

For our purposes, the most important sky signal component is the CMB anisotropy field, \mathbf{a}_{CMB} . The covariance of this component reads $\mathbf{S} = \langle \mathbf{a}_{\text{CMB}}^T \mathbf{a}_{\text{CMB}} \rangle$. Under the assumption of a statistically isotropic universe, this matrix is diagonal in harmonic space, $\langle a_{\ell m} a_{\ell' m'}^* \rangle = C_\ell(\theta) \delta_{\ell\ell'} \delta_{mm'}$, where $a_{\text{CMB}}(\hat{n}) = \sum_{\ell, m} a_{\ell m} Y_{\ell m}(\hat{n})$ is the usual spherical harmonics expansion and C_ℓ is called the angular power spectrum. This power spectrum depends directly on a small set of cosmological parameters, θ , and for most universe models can be computed efficiently by Boltzmann solvers such as CAMB (Lewis et al. 2000) or CLASS (Lesgourgues 2011).

With this notation, the goal is now to compute $P(\omega | \mathbf{d})$, where $\omega = \{\mathbf{G}, \mathbf{n}_{\text{corr}}, \mathbf{M}, \mathbf{a}, \theta, \dots\}$. Unfortunately, directly evaluating or sampling from this distribution is unfeasible. In practice, we therefore resort to MCMC sampling in general, and Gibbs sampling in particular, which is a well-established method for sampling from a complicated multivariate distribution by iterating over all conditional distributions. For the data model described in Eq. (3), this process can be described schematically in the form of a Gibbs chain,

$$\mathbf{G} \leftarrow P(\mathbf{G} | \mathbf{d}, \mathbf{n}_{\text{corr}}, \mathbf{M}, \mathbf{a}, \theta) \quad (4)$$

$$\mathbf{n}_{\text{corr}} \leftarrow P(\mathbf{n}_{\text{corr}} | \mathbf{d}, \mathbf{G}, \mathbf{M}, \mathbf{a}, \theta) \quad (5)$$

$$\mathbf{M} \leftarrow P(\mathbf{M} | \mathbf{d}, \mathbf{G}, \mathbf{n}_{\text{corr}}, \mathbf{a}, \theta) \quad (6)$$

$$\mathbf{a} \leftarrow P(\mathbf{a} | \mathbf{d}, \mathbf{G}, \mathbf{n}_{\text{corr}}, \mathbf{M}, \theta) \quad (7)$$

$$\theta \leftarrow P(\theta | \mathbf{d}, \mathbf{G}, \mathbf{n}_{\text{corr}}, \mathbf{M}, \mathbf{a},) \quad (8)$$

where \leftarrow indicates replacing the parameter on the left-hand side with a random sample from the distribution on the right-hand side.

As described by the [BeyondPlanck Collaboration \(2023\)](#) and [Watts et al. \(2023\)](#), this process has been implemented in `Commander3` by `BEYONDPLANCK` and `COSMOGLOBE`. However, in the code described there, the last step only supports a power spectrum estimation (i.e., $\theta = C_\ell$). In this work, we did not sample any instrumental or foreground parameters, and we assumed a foreground-cleaned map of the CMB as a starting point. Rather, the goal of the paper is to replace the last step, Eq. (8), with actual cosmological parameter estimation of a CMB map, in which θ takes on the usual form of the dark matter density Ω_c , the Hubble constant H_0 , the spectral index of scalar perturbations n_s , etc.

2.2. Joint sampling of CMB sky signal and cosmological parameters

Cosmological parameter estimation through end-to-end CMB Gibbs sampling has been discussed in the literature for almost two decades, starting with [Jewell et al. \(2004\)](#) and [Wandelt et al. \(2004\)](#). However, as pointed out by [Eriksen et al. \(2004\)](#), a major difficulty regarding this method is a very long correlation length in the low signal-to-noise regime, i.e., at high multipoles. Intuitively, the origin of this problem lies in the fundamental Gibbs

sampling algorithm itself, namely that it only allows for parameter variations parallel or orthogonal to the coordinate axes of each parameter in question, and not diagonal moves. For highly degenerate distributions, this makes it very expensive to move from one tail of the distribution to the other. A solution to this problem was proposed by [Jewell et al. \(2009\)](#), who introduced a joint CMB sky signal and C_ℓ move. This idea was further refined by [Racine et al. \(2016\)](#), who noted that faster convergence could be obtained by only rescaling the low signal-to-noise multipoles. In the following, we give a brief summary of these ideas, both standard Gibbs sampling in Sect. 2.2.1 and joint sampling in Sect. 2.2.2.

2.2.1. Standard CMB Gibbs sampling

Starting with the standard Gibbs sampling algorithm ([Jewell et al. 2004](#); [Wandelt et al. 2004](#)), we first note that as far as cosmological parameter estimation with CMB data is concerned, only the CMB map, \mathbf{a}_{CMB} (which we for brevity denote \mathbf{a} in the rest of this section), is directly relevant. All other parameters in Eq. (3) only serve to produce the cleanest possible estimate of \mathbf{a} and to propagate the corresponding uncertainties. From now on, we therefore consider a greatly simplified data model of the form,

$$\mathbf{r} = \mathbf{B}\mathbf{a} + \mathbf{n}, \quad (9)$$

where \mathbf{r} now represents a foreground-cleaned CMB map (or residual) obtained by subtracting all other nuisance terms from the raw data, and our goal is to estimate cosmological parameters from this map. Under the assumption that the CMB is a Gaussian random field, this could in principle be done simply by mapping out the exact marginal posterior distribution by brute force,

$$P(\theta | \mathbf{r}) \propto \mathcal{L}(\theta) P(\theta) \propto \frac{e^{-\frac{1}{2} \mathbf{r}^T (\mathbf{B}^T \mathbf{S}(\theta) \mathbf{B} + \mathbf{N})^{-1} \mathbf{r}}}{\sqrt{|\mathbf{B}^T \mathbf{S}(\theta) \mathbf{B} + \mathbf{N}|}}, \quad (10)$$

where $\mathbf{B}^T \mathbf{S}(\theta) \mathbf{B}$ is the beam-convolved CMB signal covariance matrix and \mathbf{N} is the corresponding noise covariance matrix, and we have assumed uniform priors, $P(\theta) = 1$. This, however, becomes intractable for modern CMB experiments as the computational expense scales as $\mathcal{O}(N_p^3)$, where N_p is the number of pixels, and nowadays typical CMB maps include millions of pixels.

In the special (and unrealistic) case of uniform noise and no Galactic mask, this likelihood can be simplified in harmonic space, in which it does become computationally tractable. Let us define the observed angular CMB power spectrum (including beam convolution and noise) to be $\hat{C}_\ell^\circ = \frac{1}{2\ell+1} \sum_m r_{\ell m} r_{\ell m}^*$ and the corresponding white noise spectrum to be defined by $N_{\ell m \ell' m'} = N_\ell \delta_{\ell\ell'} \delta_{mm'}$. In that case, the likelihood in Eq. (10) for temperature-only observations simplifies to

$$\ln P(\theta | \mathbf{r}) = \sum_\ell -\frac{2\ell+1}{2} \left[\frac{\hat{C}_\ell^\circ}{b_\ell^2 C_\ell(\theta) + N_\ell} - \ln \left(\frac{\hat{C}_\ell^\circ}{b_\ell^2 C_\ell(\theta) + N_\ell} \right) \right], \quad (11)$$

where b_ℓ is the beam response function and we have removed constant terms; a similar expression for polarization is available in, for example, [Larson et al. \(2007\)](#) and [Hamimeche & Lewis \(2008\)](#). This expression is used later to validate the production code.

While it is difficult to evaluate Eq. (10) directly, it is in fact possible to sample from it using Gibbs sampling ([Jewell et al. 2004](#);

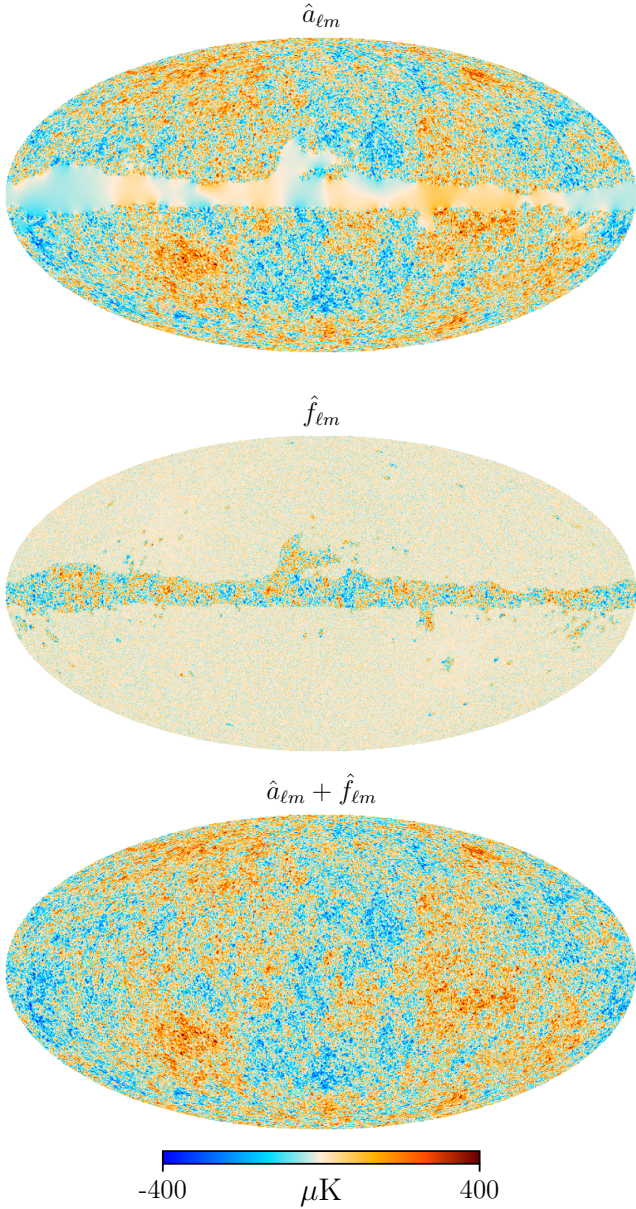


Fig. 1. Example of a constrained CMB sky map realization produced during traditional Gibbs sampling. From top to bottom, the three panels show the mean field map ($\hat{\mathbf{a}}$), the fluctuation map ($\hat{\mathbf{f}}$), and the full constrained realization ($\mathbf{a} = \hat{\mathbf{a}} + \hat{\mathbf{f}}$). This example is generated with the data configuration discussed in Sect. 3, which corresponds to *Planck* LFI 70 GHz noise properties and a realistic Galactic mask.

Wandelt et al. 2004). First, we note that the joint posterior distribution $P(\theta, \mathbf{a} | \mathbf{r})$ can be written as

$$P(\theta, \mathbf{a} | \mathbf{r}) = \frac{P(\theta, \mathbf{a}, \mathbf{r})}{P(\mathbf{r})} = P(\mathbf{r} | \mathbf{a})P(\mathbf{a} | \theta) \frac{P(\theta)}{P(\mathbf{r})} \\ \propto \frac{e^{-\frac{1}{2}(\mathbf{r}-\mathbf{B}\mathbf{a})^T \mathbf{N}^{-1}(\mathbf{r}-\mathbf{B}\mathbf{a})}}{\sqrt{|\mathbf{N}|}} \frac{e^{-\frac{1}{2}\mathbf{a}^T \mathbf{S}^{-1}\mathbf{a}}}{\sqrt{|\mathbf{S}|}} \frac{P(\theta)}{P(\mathbf{r})}, \quad (12)$$

and drawing samples from this distribution can be achieved through Gibbs sampling,

$$\mathbf{a}^{i+1} \leftarrow P(\mathbf{a} | \theta^i, \mathbf{r}), \quad (13)$$

$$\theta^{i+1} \leftarrow P(\theta | \mathbf{a}^{i+1}, \mathbf{r}). \quad (14)$$

With these joint samples in hand, a numerical representation of $P(\theta | \mathbf{r})$ can be obtained simply by marginalizing over \mathbf{a} , which is equivalent to simply making histograms of the θ samples.

For this algorithm to work, we need to be able to sample from each of the two conditional distributions in Eqs. (13) and (14). Starting with the cosmological parameter distribution in Eq. (14), we note that $P(\theta | \mathbf{a}^{i+1}, \mathbf{r}) = P(\theta | \mathbf{a}^{i+1})$, which is equivalent to the distribution in Eq. (11) with no beam and noise. This can be coupled to a standard Boltzmann solver and Metropolis-Hastings sampler, similar to CAMB and CosmoMC. We return to this step in the next section.

Next, for the amplitude distribution in Eq. (13), one can show by “completing the square” of \mathbf{a} in Eq. (12) that

$$P(\mathbf{a} | \theta, \mathbf{r}) \propto e^{-\frac{1}{2}(\mathbf{a}-\hat{\mathbf{a}})^T (\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B})(\mathbf{a}-\hat{\mathbf{a}})}, \quad (15)$$

where we have defined the so-called mean field map,

$$\hat{\mathbf{a}} \equiv [\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B}]^{-1} \mathbf{B}^T \mathbf{N}^{-1} \mathbf{r}. \quad (16)$$

This is a multivariate Gaussian distribution with mean $\hat{\mathbf{a}}$ and a covariance matrix $[\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B}]^{-1}$ that can be sampled from solving the following equation using conjugate gradients (Shewchuk 1994; Seljebotn et al. 2019),

$$[\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B}] \mathbf{a} = \mathbf{B}^T \mathbf{N}^{-1} \mathbf{r} + \mathbf{S}^{-\frac{1}{2}} \mathbf{w}_0 + \mathbf{B}^T \mathbf{N}^{-\frac{1}{2}} \mathbf{w}_1, \quad (17)$$

where \mathbf{w}_0 and \mathbf{w}_1 are randomly drawn Gaussian maps with unit variance and zero mean. The resulting sample is typically referred to as a “constrained realization”.

For the purposes of the next section, it is useful to decompose the full constrained realization into two separate components, namely the mean field map defined by Eq. (16) and the fluctuation map,

$$\hat{\mathbf{f}} \equiv [\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B}]^{-1} (\mathbf{S}^{-\frac{1}{2}} \mathbf{w}_0 + \mathbf{B}^T \mathbf{N}^{-\frac{1}{2}} \mathbf{w}_1) \quad (18)$$

such that $\mathbf{a} = \hat{\mathbf{a}} + \hat{\mathbf{f}}$. For a data configuration with uniform noise and full-sky temperature data, these equations can be solved exactly in spherical harmonics space, fully analogously to Eq. (11),

$$\hat{a}_{\ell m} = r_{\ell m} \frac{b_{\ell} C_{\ell}}{N_{\ell} + b_{\ell}^2 C_{\ell}}, \quad (19)$$

$$\hat{f}_{\ell m} = w_{0\ell m} \frac{N_{\ell} \sqrt{C_{\ell}}}{N_{\ell} + b_{\ell}^2 C_{\ell}} + w_{1\ell m} \frac{\sqrt{N_{\ell}} b_{\ell} C_{\ell}}{N_{\ell} + b_{\ell}^2 C_{\ell}}. \quad (20)$$

As summarized in Appendix A, we can find a closed-form expression for $N_{\ell m' m''}^{-1}$ for a constant latitude mask with isotropic noise which greatly simplifies solving Eq. (17), and we have made a Python script that performs these calculations in order to validate the main Commander3 code discussed below.

To build intuition, these maps are illustrated for one random sample with semi-realistic instrument characteristics in Fig. 1. The top panel shows the mean field map (or “Wiener filter map”); this map summarizes all significant information in \mathbf{r} , and each mode is weighted according to its own specific signal-to-noise ratio taking into account both instrumental noise and masking. As a result, all small scales in the Galactic plane are suppressed, and only the largest scales are retained, which are constrained by the high-latitude observations in conjunction with the assumptions of both Gaussianity and statistical isotropy. The second panel shows the fluctuation map, which essentially serves

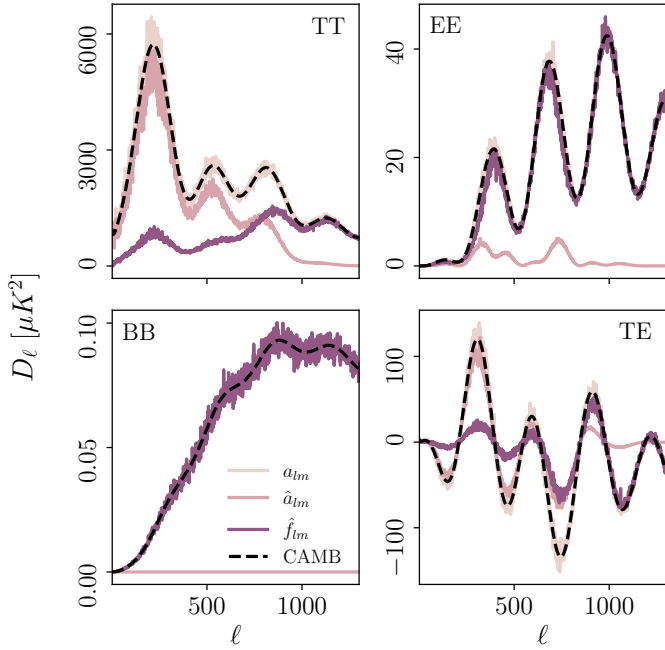


Fig. 2. Angular power spectra for each of the constrained realization maps shown in Fig. 1. The Λ CDM cosmological model corresponding to θ is shown as a dashed black line, while the colored curves show (from dark to light) spectra for the fluctuation map, the mean field map, and the full constrained realization. Note that in the fully noise-dominated cases, such as the entire B-mode power spectrum and all high multipoles, the fluctuation term $\hat{f}_{\ell m}$ and the CMB estimate $a_{\ell m}$ have nearly identical power spectra by design.

to replace the lost power in $\hat{\mathbf{a}}$ with a completely random fluctuation. The sum of the two terms, shown in the bottom panel, represents one possible full-sky realization that is consistent with both the data and the assumed cosmological parameters, θ . If we were to produce a second such constrained realization, it would look almost identical at high Galactic latitudes, where the data are highly constraining, while the Galactic plane would exhibit different structures on small and intermediate scales.

A corresponding comparison of angular power spectra is shown in Fig. 2. The relative signal-to-noise ratio of each multipole moment can be seen as the ratio between the intermediate and dark-colored lines. In temperature, this value is higher than unity up to $\ell \lesssim 700$, while for EE polarization it is less than unity everywhere except at the very lowest multipoles. A closely related way to interpret this figure is that the intermediate curve shows the total information content of the data. Thus, for the BB spectrum there is no visually relevant information present at all, and the full constrained realization is entirely defined by the fluctuation map, which is given by the assumed cosmological model, θ . This explains why the standard Gibbs sampler works well in the high signal-to-noise regime but poorly in the low signal-to-noise limit: When we have high instrumental noise, $N \rightarrow \infty$, then we have no information about $a_{\ell m}$ from the data $r_{\ell m}$, and the total sky sample only becomes a realization of the assumed variance, C_ℓ (i.e., $a_{\ell m} \rightarrow \hat{f}_{\ell m} = \sqrt{C_\ell} w_{\ell m}$). However, since we have no information of the mean field map $\hat{\mathbf{a}}_{\ell m} \rightarrow 0$, we also do not know C_ℓ . Hence, we get a strong degeneracy between the C_ℓ and $a_{\ell m}$. Gibbs sampling is well known to perform poorly for strongly correlated parameters, as one cannot move diagonally through the parameter space from one end of the joint posterior distribution to the other.

2.2.2. Joint sky signal and cosmological parameter sampling

As the Gibbs sampler struggles with long correlation lengths, a joint sampler was proposed by Jewell et al. (2009) and Racine et al. (2016) that was designed to establish an efficient sampling algorithm for the degeneracy between \mathbf{a} and θ in the low signal-to-noise regime. Mathematically, this algorithm is a standard Metropolis-Hastings sampler with a Gaussian transition rule for θ ,

$$w(\theta | \theta^i) = e^{-\frac{1}{2}(\theta - \theta^i)^T \mathbf{C}_\theta^{-1}(\theta - \theta^i)}, \quad (21)$$

where \mathbf{C}_θ is a user-specified (and pre-tuned) proposal covariance matrix made from a previous chain using a suboptimal, diagonal proposal matrix. This is followed by a deterministic rescaling of the fluctuation map in the signal amplitude map,

$$a_{\ell m}^{i+1} = \hat{a}_{\ell m}^{i+1} + (\mathbf{C}_\ell^{i+1})^{1/2} (\mathbf{C}_\ell^i)^{-1/2} \hat{f}_{\ell m}^i, \quad (22)$$

where we have defined the covariance matrix

$$\mathbf{C}_\ell^i = \begin{bmatrix} C_\ell^{TT}(\theta^i) & C_\ell^{TE}(\theta^i) & 0 \\ C_\ell^{TE}(\theta^i) & C_\ell^{EE}(\theta^i) & 0 \\ 0 & 0 & C_\ell^{BB}(\theta^i) \end{bmatrix} \quad (23)$$

Here, we have set the parity-odd power spectra to zero as predicted by Λ cold dark matter (CDM), $C_\ell^{EB} = C_\ell^{TB} = 0$. Therefore, we defined the rescaled fluctuation map,

$$\hat{f}_{\ell m}^{\text{scaled}, i+1} \equiv (\mathbf{C}_\ell^{i+1})^{1/2} (\mathbf{C}_\ell^i)^{-1/2} \hat{f}_{\ell m}^i. \quad (24)$$

As shown by Racine et al. (2016), the acceptance rate for this joint move is

$$A = \min \left[1, \frac{\pi(\theta^{i+1}) P(\theta^{i+1})}{\pi(\theta^i) P(\theta^i)} \right], \quad (25)$$

where $P(\theta)$ is the prior on θ and

$$\pi(\theta^i) = \exp \left[-\frac{1}{2} (\mathbf{r} - \mathbf{B}\hat{\mathbf{a}}^i)^T \mathbf{N}^{-1} (\mathbf{r} - \mathbf{B}\hat{\mathbf{a}}^i) - \frac{1}{2} \mathbf{a}^{i,T} \mathbf{S}^{i-1} \hat{\mathbf{a}}^i - \frac{1}{2} \hat{\mathbf{f}}^{i,T} \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B} \hat{\mathbf{f}}^i \right], \quad (26)$$

where we for brevity have omitted the θ^i dependence of $\hat{\mathbf{a}}^i$, $\hat{\mathbf{f}}^i$ and \mathbf{S}^i .

We note that the acceptance rate uses the scaled fluctuation term, $\hat{f}_{\ell m}^{\text{scaled}, i+1}$, for sample $i+1$ instead of the calculated fluctuation term in Eq. (17). Hence, we only need to calculate $\hat{f}_{\ell m}^{i+1}$ from this equation if the sample is accepted. This allows us to save roughly half the computational time of discarded samples as compared to accepted samples.

The full algorithm can now be summarized in terms of the following steps:

1. We start with an initial guess of the cosmological parameters, θ^0 . From that, we calculate \mathbf{S}^0 with CAMB; the mean field map, $\hat{\mathbf{a}}^0$ from Eq. (16); and the fluctuation map, $\hat{\mathbf{f}}^0$, from Eq. (18).
2. We then draw a new cosmological parameter sample from $w(\theta | \theta^i)$, and reevaluate \mathbf{S}^{i+1} , $\hat{\mathbf{a}}^{i+1}$, and the scaled fluctuation term $\hat{f}_{\ell m}^{\text{scaled}, i+1}$.
3. We calculate the acceptance term of Eq. (25), and accept or reject according to the regular Metropolis-Hastings rule.

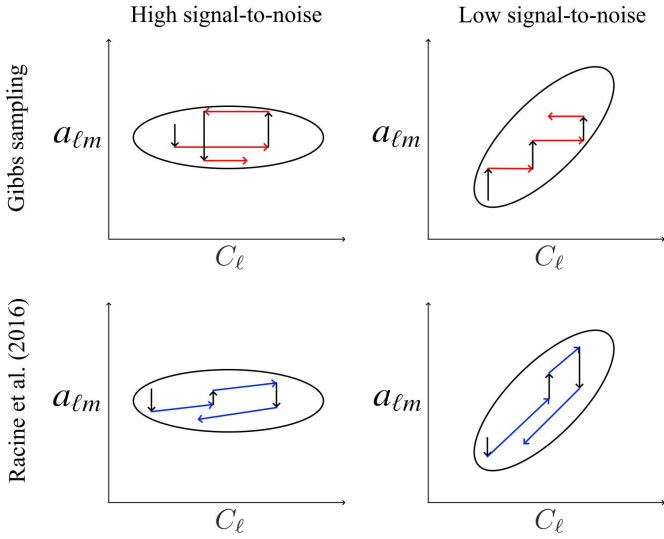


Fig. 3. Schematic comparison of the standard Gibbs sampler (top panels) and the joint sampler of Racine et al. (2016), bottom panels. The left and right columns show high and low signal-to-noise regimes, respectively. The Gibbs sampler performs poorly in the low signal-to-noise regime as it requires a large number of samples to explore the posterior distribution. The joint sampler in the lower panels performs well in both regimes as it allows the next sample to move diagonally in the $\{a, \theta\}$ parameter space.

4. If the sample is accepted, then we calculate \hat{f}^{i+1} . In the next iteration this term becomes \hat{f}^i , which will appear in the acceptance probability and the equation for $\hat{f}^{\text{scaled}, i+1}$.
5. Iterate 2–4.

The intuition behind this joint move is illustrated in Fig. 3. The top panel shows the standard Gibbs sampling algorithm, in which only parameter moves parallel to the axes are allowed. This works well in the high signal-to-noise regime (left column), where the joint distribution is close to symmetric. In the low signal-to-noise regime, however, the distribution is strongly tilted, because of the tight coupling between \hat{f} and θ discussed at the end of Sect. 2.2.1, and many orthogonal moves are required in order to move from one tail of the distribution to the other. On the other hand, with the joint algorithm, in the bottom panel, steps are non-orthogonal, with a slope defined exactly by the local signal-to-noise ratio of the mode in question. The net result is a much faster exploration of the full distribution.

3. Results

The main goal of this paper is to implement the joint sampling algorithm into Commander3 and characterize its performance both in terms of accuracy and computational cost on simulated data. All Commander code used in the following is available as Open Source software in a GitHub repository². For the current paper, we focused on a standard six-parameter Λ CDM model and chose $\theta = (\Omega_b h^2, \Omega_{\text{CDM}} h^2, H_0, \tau, A_s, n_s)$ as our base parameters, where Ω_b and Ω_{CDM} are the current density of baryonic and CDM; H_0 is the Hubble parameter (and h is the normalized Hubble constant, $h = \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}$); τ is the optical depth at reionization; and A_s and n_s are the amplitude and tilt of the

scalar primordial power spectrum. We used CAMB³ (Lewis et al. 2000) to evaluate all Λ CDM power spectra, $C_{\ell}(\theta)$.

In the following we consider three different simulations, corresponding to increasing levels of complexity and realism. A natural early target for this algorithm is a reanalysis of COSMOGLOBE DR1; hence, all three cases are similar to the *Planck* LFI channels. We use the *Planck* data release 4 70 GHz beam transfer functions (Planck Collaboration Int. LVII 2020), and for the two first cases, the noise level is isotropic with a value matching twice the mean of the full-sky 70 GHz rms map, while for the third case is given by one realization of the actual non-isotropic BEYONDPLANCK 70 GHz noise rms distribution.

The second difference between the three cases is their sky coverage. The first case considers full-sky data, for which the analytical posterior distribution solution is readily available through Eq. (11). The second case implements a constant-latitude Galactic mask for which an analytical expression is also available, albeit more complicated than for the full sky (see Appendix A for details). The third case uses a realistic Galactic and point-source mask, for which no analytical expression is available.

To validate the main Commander3 results, we developed a special-purpose Python program⁴ that implements the same joint sampling algorithm described above for the two first cases, and this serves essentially as a basic bug-check with respect to the Fortran implementation. In addition, we also implemented a uniform, full-sky likelihood module for Cobaya (Torrado & Lewis 2021), an industry-standard cosmological parameter Metropolis sampler, and this serves as an end-to-end check on the entire code, including integration with CAMB.

3.1. Accuracy validation

The first main goal of this paper is to demonstrate that the new Commander3 implementation of the Racine et al. (2016) algorithm performs as expected with respect to accuracy. Thus, we start our discussion by considering the uniform and full-sky case discussed above. For each of the three available codes – Commander3, Python, and Cobaya – we produced 50 000 samples; we show the full posterior distributions in Fig. 4. The agreement between all three codes is excellent, and the overall variations are within $\sim 0.1\sigma$.

Second, Fig. 5 shows a similar comparison between the Commander3 and Python implementation. The sky mask is in this case defined by $|b| < 5.7^\circ$ where b is the Galactic latitude, removing 10% of the sky. We also find good agreement in this case between the two implementations, and the true input values (marked by vertical dashed lines) all fall well within the full posterior distributions. However, we do note that the introduction of a sky mask significantly increases the overall run time of the algorithm, since one now has to solve for the mean field map repeatedly at each step with conjugate gradients, rather than with brute force matrix inversions. In total, we only produce $\sim 11\,000$ and $17\,000$ samples for each of the two codes for this exercise, at a total cost of $\mathcal{O}(10^5)$ CPU-hours.

Finally, in Fig. 6 we show trace plots for each cosmological parameter for the realistic configuration as produced with Commander3, noting that none of the other codes are technically able to produce similar estimates. In this case, the computational cost is even higher, and we only produce a total of 1600 samples.

³ <https://github.com/cmbant/CAMB>

⁴ Publicly available at <https://github.com/LilleJohs/COLOCOLA>

² <https://github.com/Cosmoglobe/Commander>

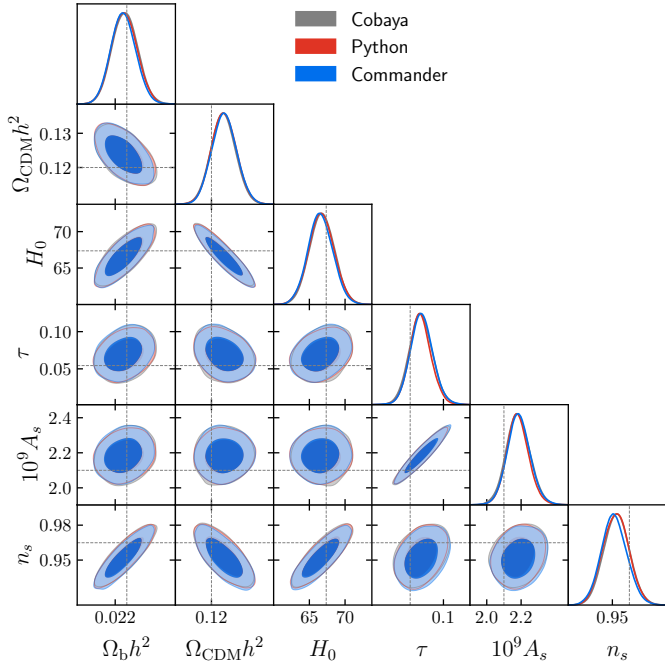


Fig. 4. Comparison of cosmological parameter posterior distributions derived with Commander3 (blue), Python (red), and Cobaya (gray) for a simulation with uniform noise and full-sky coverage. They are made from a single chain per code, and the true input cosmological parameter values are indicated by vertical and horizontal dashed lines.

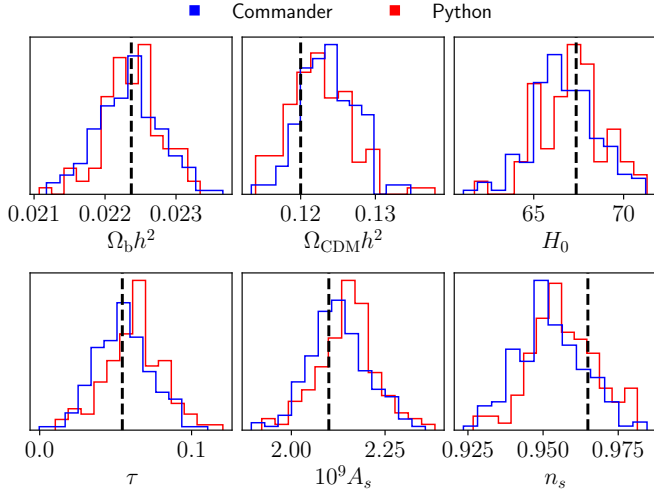


Fig. 5. Comparison of marginal cosmological parameter posterior distributions derived with Commander3 and Python for a simulation with uniform noise and a 10% constant latitude mask. The total number of independent samples produced by the two codes is $\mathcal{O}(10^2)$, which accounts for the Markov chain correlation length of the algorithm; the Monte Carlo uncertainty due to a finite number of samples is therefore significant.

Still, by comparing these traces with the true input values (marked as dashed horizontal lines), we do see that the resulting samples agree well with the expected mean values. The uncertainties are also as expected since they are smaller and on the order of the noisier 10% constant latitude and isotropic noise case in Fig. 5.

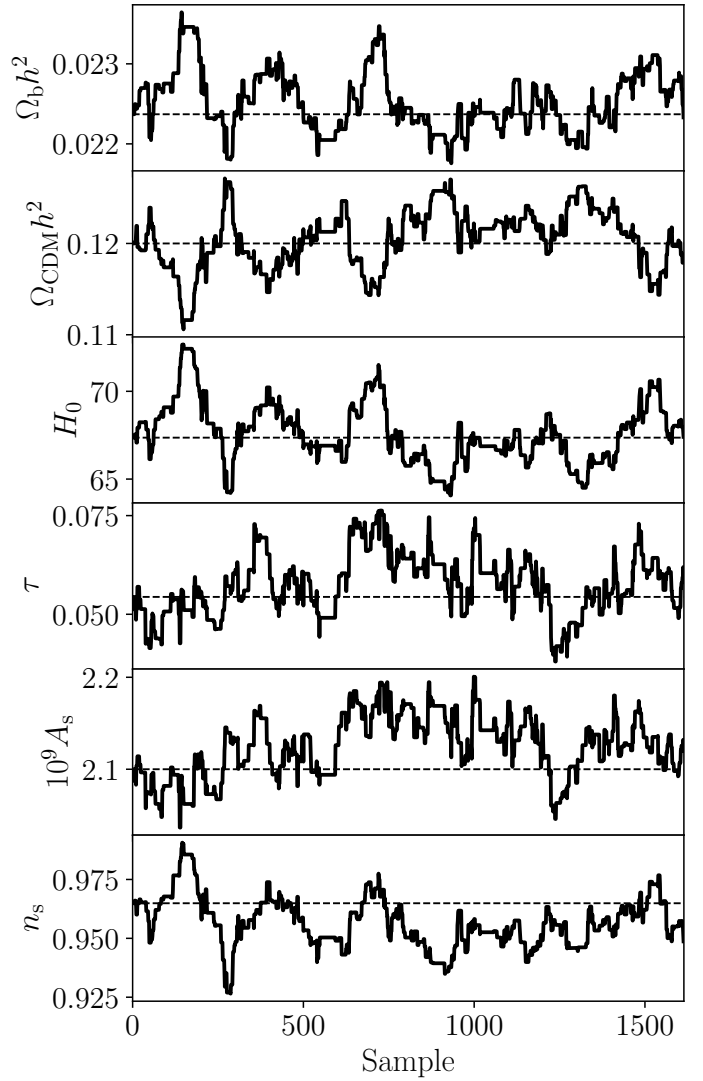


Fig. 6. Trace plot of the cosmological parameters for a semi-realistic case based on the *Planck* LFI 70 GHz noise level and a realistic Galactic mask. The dashed horizontal lines show the true input values.

3.2. Markov chain correlation length and computational costs

The second main goal of this paper is to quantify the computational costs involved in this algorithm and identify potential bottlenecks that could be optimized in future work. The effective cost per independent Markov chain sample can be written as the product of the raw sampling cost per sample and the overall Markov chain correlation length. The latter can be quantified in terms of the autocorrelation function,

$$\zeta_\theta(\Delta) \equiv \left\langle \frac{(\theta_i - \mu)(\theta_{i+\Delta} - \mu)}{\sigma^2} \right\rangle, \quad (27)$$

where μ and σ are the posterior mean and standard deviation computed from the Markov chain for θ . We define the overall correlation length to be the first value of Δ for which $\zeta < 0.1$. This function is plotted for both Cobaya and Commander3 in Fig. 7; for the latter, dashed lines show full-sky results and dot-dashed lines show results for a 10% mask. Overall, we see that the Commander3 correlation length is 50–60 for the full-sky case, increasing to 50–100 for the masked case. For comparison, Cobaya typically has a correlation length of about 10.

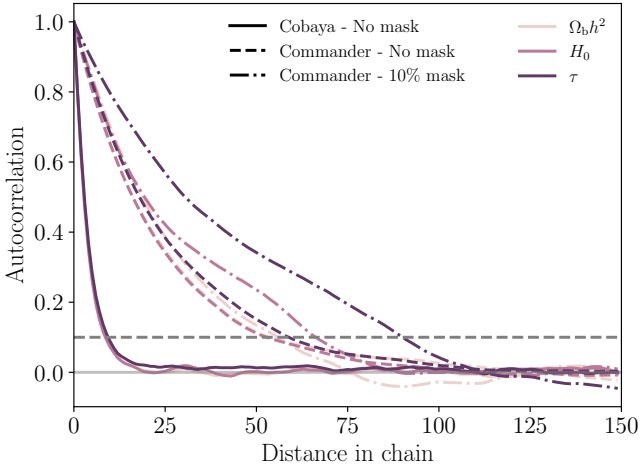


Fig. 7. Markov chain autocorrelation functions for Cobaya (solid lines) and Commander3 for a set of typical cosmological parameters. Dashed and dot-dashed lines show Commander3 results for the full-sky case and the 10% mask case, respectively. The horizontal dashed gray line indicates an autocorrelation of 0.1, which we use to define the overall correlation length.

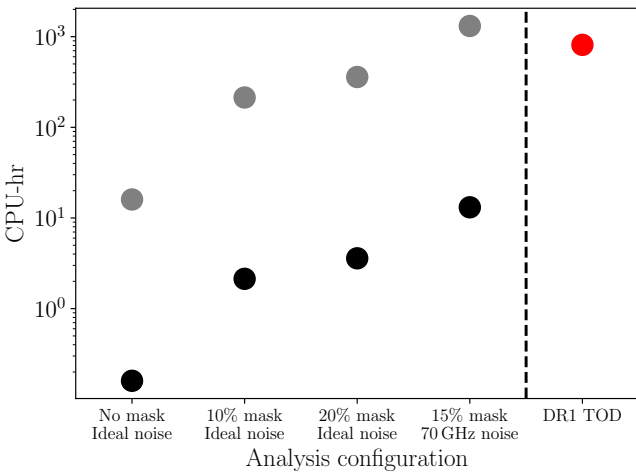


Fig. 8. CPU-hours per accepted sample. The black dots are the costs per sample, while the gray dots are the costs per effective sample, i.e., normalized by the correlation length of the chain. This is compared to the 812 CPU-hours to generate one single end-to-end sample of DR1 COSMOGLOBE *Planck* LFI and WMAP channels, in red.

The basic computational cost per sample for Commander3 is plotted as black points in Fig. 8. Here we see that the overall cost increases rapidly with decreasing sky coverage, starting at 0.2 CPU-hours for a full-sky sample to about 4 CPU-hours for a 20% constant latitude mask and uniform noise. As the isotropic noise cases have twice the noise of the mean of the 70 GHz rms map, the higher signal-to-noise causes an increase in runtime, and for the realistic LFI 70 GHz scanning pattern and 15% mask, the cost is 20 CPU-hours/sample. Taking into account that the overall correlation length is ~ 100 , the effective cost is therefore about 2000 CPU-hours per independent sample; the gray dots in Fig. 8 indicates this total cost for each configuration.

To interpret the implications of these costs, it is useful to compare with the overall cost of the full end-to-end processing pipeline within which this algorithm is designed to operate. The most relevant comparison point is therefore the cost of the COSMOGLOBE DR1 processing, which is 812 CPU-hours for the

combination of *Planck* LFI and WMAP, marked as a red dot in Fig. 8. In its current state, we find that the new cosmological parameter step alone is thus about 2.5 times more expensive than the full end-to-end processing.

4. Summary and discussion

This work had two main goals. The first goal was simply to implement the cosmological parameter sampling algorithm proposed by Racine et al. (2016) into the state-of-the-art Commander3 CMB Gibbs sampler and demonstrate that this new step works in a production environment. A head-to-head comparison with both Cobaya and a stand-alone Python implementation demonstrates the fidelity of the method. We therefore conclude that this goal has been met, and as such this work finally adds true “end-to-end” processing abilities to Commander3.

The second goal was to measure the computational performance of the method, assess whether it is already suitable for large-scale production, and identify potential bottlenecks that should be resolved in future work. In this case, we find that the overall cost per independent Gibbs sample for a data set similar to COSMOGLOBE DR1 is about 2000 CPU-hours, which is more than a factor of two higher than the total cost of all other steps in the end-to-end algorithm, including low-level calibration, map-making, and component separation. Therefore, although the code is technically operational at the current time, we also conclude that it is still too expensive to be useful in a full-scale production environment.

At the same time, this analysis has also revealed the underlying origin of these high expenses, which can be divided into two parts. First, the overall cost per sample is dominated by the expense for solving for the mean field (Eq. (16)) and fluctuation maps (Eq. (18)) using conjugate gradients. In this respect, we note that the current implementation uses a standard diagonal preconditioner to solve these equations, as described by Eriksen et al. (2004). However, after more than two decades of development, far more efficient preconditioners have been explored and described in the literature, including multi-resolution or multi-grid methods (e.g., Seljebotn et al. 2014, 2019). Implementing one of them can potentially reduce the basic cost per sample by one or two orders of magnitude. The second issue is a long correlation length. We note that the current algorithm is based on a very basic random-walk Metropolis sampler, which generally performs poorly for highly correlated parameters; in the cases considered here, the correlation between the amplitude of scalar perturbations, A_s , and the optical depth of reionization, τ , is a typical example of such a correlation. This suggests that the correlation length can be greatly reduced in several ways, for instance either by adding new data sets (such as weak gravitational lensing constraints) or by implementing a more sophisticated sampling algorithm that uses derivative information, such as a Hamiltonian sampler. Overall, we consider it very likely that future work will be able to reduce the total cost by one or two orders of magnitude, as needed for production, and we hope that the current results can provide inspiration and motivation for experts in the field to join the effort.

Finally, before concluding this paper, we note that a key application of the Bayesian end-to-end analysis framework as pioneered by BEYONDPLANCK and COSMOGLOBE is the analysis of next-generation CMB *B*-mode experiments, for instance LiteBIRD. In this case, it is worth noting that the polarization-based signal-to-noise ratio of the tensor-to-scalar ratio, r , is much lower than *Planck* and WMAP’s temperature signal-to-noise ratio to the Λ CDM parameters, and both the conjugate

gradient cost and Markov chain correlation length are likely to be much shorter than for the case considered in this paper. As such, it is conceivable that the current method already performs adequately for LiteBIRD, and this will be explored⁵ in future work.

Acknowledgements. The current work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 819478 (ERC; COSMOGLOBE) and 772253 (ERC; BITS2COSMOLOGY). In addition, the collaboration acknowledges support from RCN (Norway; grant no. 274990). Moreover, Simone Paradiso acknowledges support from the Government of Canada's New Frontiers in Research Fund, through grant NFRFE-2021-00595. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), part of the High Energy Astrophysics Science Archive Center (HEASARC). HEASARC/LAMBDA is a service of the Astrophysics Science Division at the NASA Goddard Space Flight Center. Some of the results in this paper have been derived using the healpy and HEALPix packages (Górski et al. 2005; Zonca et al. 2019).

References

- Anderes, E., Wandelt, B., & Lavaux, G. 2015, *ApJ*, 808, 152
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, *ApJS*, 208, 20
- BeyondPlanck Collaboration 2023, *A&A*, 675, A1
- Carron, J. 2019, *Phys. Rev. D*, 99, 043518
- Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, *ApJS*, 155, 227
- Galloway, M., Andersen, K. J., & Aurlen, R. 2023, *A&A*, 675, A3
- Gjerløw, E., Mikkelsen, K., Eriksen, H. K., et al. 2013, *ApJ*, 777, 150
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
- Hamimeche, S., & Lewis, A. 2008, *Phys. Rev. D*, 77, 103013
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Jewell, J., Levin, S., & Anderson, C. H. 2004, *ApJ*, 609, 1
- Jewell, J. B., Eriksen, H. K., Wandelt, B. D., et al. 2009, *ApJ*, 697, 258
- Larson, D. L., Eriksen, H. K., Wandelt, B. D., et al. 2007, *ApJ*, 656, 653
- Lesgourgues, J. 2011, ArXiv e-prints [arXiv:1104.2932]
- Lewis, A., & Bridle, S. 2002, *Phys. Rev. D*, 66, 103511
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Millea, M., & Seljak, U. 2022, *Phys. Rev. D*, 105, 103531
- Millea, M., Anderes, E., & Wandelt, B. D. 2020, *Phys. Rev. D*, 102, 123542
- Millea, M., Daley, C. M., Chou, T-L., et al. 2021, *ApJ*, 922, 259
- Planck Collaboration II. 2020, *A&A*, 641, A2
- Planck Collaboration III. 2020, *A&A*, 641, A3
- Planck Collaboration V. 2020, *A&A*, 641, A5
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Planck Collaboration Int. LVII. 2020, *A&A*, 643, A42
- Racine, B., Jewell, J. B., Eriksen, H. K., & Wehus, I. K. 2016, *ApJ*, 820, 31
- Seljebotn, D. S., Mardal, K.-A., Jewell, J. B., Eriksen, H. K., & Bull, P. 2014, *ApJS*, 210, 24
- Seljebotn, D. S., Bærland, T., Eriksen, H. K., Mardal, K. A., & Wehus, I. K. 2019, *A&A*, 627, A98
- Shewchuk, J. R. 1994, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Edition 1 $\frac{1}{4}$, <http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf>
- Torrado, J., & Lewis, A. 2021, *JCAP*, 05, 057
- Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003, *ApJS*, 148, 195
- Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, *Phys. Rev. D*, 70, 083511
- Watts, D. J., Basyrov, A., Eskilt, J. R., et al. 2023, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202346414>
- Wieczorek, M. A., & Meschede, M. 2018, *Geochem. Geophys. Geosyst.*, 19, 2574
- Zonca, A., Singer, L., Lenz, D., et al. 2019, *J. Open Source Softw.*, 4, 1298

⁵ <http://healpix.sf.net>

Appendix A: Analytic expression for a constant latitude mask

The goal of this appendix is to show how we can calculate $N_{\ell m \ell' m'}^{-1}$ analytically for a constant latitude mask and uniform noise. This expression makes the map-making equation, in Eq. (17), computationally faster to solve as we derive our expression to be $N_{\ell m \ell' m'}^{-1} \propto \delta_{mm'}$. These results are generally known, but we re-derive them here for completeness.

For uniform instrumental noise with no mask, the pixel covariance matrix is $N_{pp'} = \sigma^2 \delta_{pp'}$ where σ is the white noise uncertainty per pixel. Including the constant latitude mask, the inverse of the noise covariance matrix can be written in pixel space as

$$(N^{-1})_{pp'} = \frac{1}{\sigma^2} \delta_{pp'} H(|\theta(p) - \pi/2| - b).$$

Here, H is the Heaviside function, meaning that we mask every pixel p where $|\theta(p) - \pi/2| < b$ for some latitude b in radians. For a masked pixel, we have $(N^{-1})_{pp} = 0$, meaning that the noise at that pixel is infinite, $N_{pp} = \infty$.

Defining the spherical harmonics functions $Y_{\ell m}(p) = Y_{\ell m}(\hat{n}(p))$ and $\hat{n}(p) = (\theta(p), \phi(p))$ are the spherical coordinates at pixel p , which we can transform into spherical harmonics space,

$$\begin{aligned} (N^{-1})_{\ell m \ell' m'} &= \sum_{pp'} (N^{-1})_{pp'} Y_{\ell m}^*(p) Y_{\ell' m'}(p') \\ &= \frac{1}{\sigma^2} \sum_{pp'} Y_{\ell m}^*(p) Y_{\ell' m'}(p') \delta_{pp'} H(|\theta - \pi/2| - b) \\ &= \frac{1}{\sigma^2} \sum_p Y_{\ell m}^*(p) Y_{\ell' m'}(p) H(|\theta - \pi/2| - b) \\ &= \frac{1}{\sigma^2} \sum_p \tilde{P}_{\ell m} \tilde{P}_{\ell' m'} e^{-i(m-m')\phi} H(|\theta - \pi/2| - b). \end{aligned} \quad (\text{A.1})$$

Here, we have defined $\tilde{P}_{\ell m} = \tilde{P}_{\ell m}(\theta) = \Delta_{\ell m} P_{\ell m}(\cos(\theta))$, where $\Delta_{\ell m} = (-1)^m \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}}$ and $P_{\ell m}(\cos(\theta))$ are the associated Legendre polynomials. Thus, we write the spherical harmonics functions as $Y_{\ell m}(p) = \tilde{P}_{\ell m} e^{im\phi}$.

We now switch from discrete pixels to continuous space, meaning that we change the sum to an integration where we account for the number of pixels, N_{pix} , per area element,

$$\sum_p \rightarrow \frac{N_{\text{pix}}}{4\pi} \int d\Omega = \frac{N_{\text{pix}}}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin(\theta). \quad (\text{A.2})$$

This gives us

$$\begin{aligned} (N^{-1})_{\ell m \ell' m'} &= \frac{N_{\text{pix}}}{4\pi\sigma^2} \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin(\theta) \tilde{P}_{\ell m} \tilde{P}_{\ell' m'} e^{-i(m-m')\phi} \\ &\quad \cdot H(|\theta - \pi/2| - b) \\ &= \frac{N_{\text{pix}}}{2\sigma^2} \delta_{mm'} \int_0^\pi d\theta \sin(\theta) \tilde{P}_{\ell m} \tilde{P}_{\ell' m'} H(|\theta - \pi/2| - b) \end{aligned}$$

$$\begin{aligned} &= \frac{N_{\text{pix}}}{2\sigma^2} \delta_{mm'} \\ &\quad \cdot \left(\int_0^{\pi/2-b} d\theta \sin(\theta) \tilde{P}_{\ell m} \tilde{P}_{\ell' m'} + \int_{\pi/2+b}^\pi d\theta \sin(\theta) \tilde{P}_{\ell m} \tilde{P}_{\ell' m'} \right). \end{aligned} \quad (\text{A.3})$$

Writing $x = \cos(\theta)$, we know that the associated Legendre polynomials $P_{\ell m}(x)$ are either symmetric or antisymmetric in $x \rightarrow -x$. $P_{\ell m}(x)$ is symmetric in $x \rightarrow -x$ when $\ell + m = \text{even}$ and antisymmetric when $\ell + m = \text{odd}$. Since $m = m'$, we note that the two integrals in the last line of Eq. (A.3) cancel each other if $\ell + \ell' = \text{odd}$. We, therefore, only get nonzero elements when $\ell + \ell' = \text{even}$, for which the two integrals are equal. Hence, for $\ell + \ell' = \text{even}$, we get

$$\begin{aligned} (N^{-1})_{\ell m \ell' m'} &= \frac{N_{\text{pix}}}{\sigma^2} \delta_{mm'} \int_0^{\pi/2-b} d\theta \sin(\theta) \tilde{P}_{\ell m}(\theta) \tilde{P}_{\ell' m'}(\theta) \\ &= \frac{N_{\text{pix}}}{\sigma^2} \delta_{mm'} \int_{\sin(b)}^1 dx \tilde{P}_{\ell m}(\arccos(x)) \tilde{P}_{\ell' m'}(\arccos(x)). \end{aligned} \quad (\text{A.4})$$

In the full-sky case, $b = 0$, we find $(N^{-1})_{\ell m \ell' m'} = \delta_{\ell\ell'} \delta_{mm'} \frac{N_{\text{pix}}}{4\pi\sigma^2}$ which is equivalent to a diagonal noise power spectrum $N_\ell = \frac{4\pi\sigma^2}{N_{\text{pix}}}$.

The integral in Eq. (A.4) can be solved numerically by gridding x , and $\tilde{P}_{\ell m}$ can be calculated in Python using the library *pyshools* (Wieczorek & Meschede 2018). The azimuthally symmetric Galactic mask used for this work has a sky coverage of $f_{\text{sky}} = 0.90$, giving $\sin(b) = 0.1$. Therefore, we require many fewer grid points for x if we use the following identity for $\ell + \ell' = \text{even}$,

$$\delta_{mm'} \int_{\sin(b)}^1 dx \tilde{P}_{\ell m}(\arccos(x)) \tilde{P}_{\ell' m}(\arccos(x)) = \quad (\text{A.5})$$

$$\frac{1}{4\pi} \delta_{\ell\ell'} \delta_{mm'} - \delta_{mm'} \int_0^{\sin(b)} dx \tilde{P}_{\ell m}(\arccos(x)) \tilde{P}_{\ell' m}(\arccos(x)), \quad (\text{A.6})$$

which comes from the orthonormality condition for spherical harmonics. We now only need to grid x in the interval $0 \leq x \leq \sin(b) = 0.1$, as opposed to the interval $0.1 \leq x \leq 1$.

Since $N_{\ell m \ell' m'}^{-1} \propto \delta_{mm'}$, we get simplified matrix expressions. Imagine multiplying the matrix $N^{-1} = N_{\ell m \ell' m'}^{-1}$ by the vector $\mathbf{b} = b_{\ell m}$:

$$N^{-1} \cdot \mathbf{b} = \sum_{\ell' m'} (N^{-1})_{\ell m \ell' m'} b_{\ell' m'} = \sum_{\ell'} (N^{-1})_{\ell m \ell' m} b_{\ell' m} \quad (\text{A.7})$$

$$= \sum_{\ell'} (N^{-1})_{\ell, \ell'}^{(m)} b_{\ell'}^{(m)}. \quad (\text{A.8})$$

To solve the map-making equation, Eq. (17), we get a matrix equation for each $0 \leq m \leq \ell_{\text{max}}$. This gives us $\ell_{\text{max}} + 1$ number of matrix equations where the dimensions of the matrices are maximally $(\ell_{\text{max}} + 1) \times (\ell_{\text{max}} + 1)$. This is numerically much quicker to invert rather than inverting the full matrix $[\mathbf{S}^{-1} + \mathbf{B}^T \mathbf{N}^{-1} \mathbf{B}]$ once.