Constraining the primordial power spectrum using a differentiable likelihood

Subarna Chaki,[a] **Andrina Nicola,**[a] **Alessio Spurio Mancini,**[b]
Davide Piras[c] **and Robert Reischke**[a]

[a]Argelander-Institut für Astronomie, Universität Bonn,
 Auf dem Hügel 71, D-53121 Bonn, Germany
[b]Department of Physics, Royal Holloway, University of London,
 Egham Hill, Egham TW20 0EX, UK
[c]Centre Universitaire d'Informatique, Université de Genève,
 Université de Genève, 7 route de Drize, 1227 Genève, Switzerland

 E-mail: schaki@astro.uni-bonn.de, anicola@uni-bonn.de,
 Alessio.SpurioMancini@rhul.ac.uk, davide.piras@unige.ch,
 rreischke@astro.uni-bonn.de

Abstract. The simplest inflationary models predict the primordial power spectrum (PPS) of curvature perturbations to be nearly scale-invariant. However, various other models of inflation predict deviations from this behaviour, motivating a data-driven approach to reconstruct the PPS and constrain its shape. In this work, we present a novel method that employs a fully differentiable pipeline to reconstruct the PPS using Gaussian Processes and uses neural network emulators for fast and differentiable theoretical predictions. By leveraging gradient-based sampling techniques, such as Hamiltonian Monte Carlo, our approach efficiently samples the high-dimensional parameter space of cosmological parameters and the free-form PPS, enabling joint constraints on both. Applying this framework to *Planck* 2018 Cosmic Microwave Background (CMB) temperature anisotropy data we find our reconstructed PPS to be consistent with near scale-invariance on small scales, while exhibiting large uncertainties at large scales, driven mostly by cosmic variance. Our results show an overestimation of the PPS amplitude compared to ΛCDM predictions from the *Planck* 2018 analysis, which we attribute to our choice of a conservative prior on the optical depth τ based on *Planck* 2015 measurements. Adopting a prior consistent with *Planck* 2018 measurements brings our results into full agreement with previous work. To ensure robustness of our results, we validate our differentiable pipeline against a non-differentiable framework, and also demonstrate that our results are insensitive to the choice of Gaussian process hyperparameters. These promising results and the flexibility of our pipeline make it ideally suited for application to additional data sets such as CMB polarisation as well as Large-Scale Structure probes, thus moving towards multi-probe primordial power spectrum reconstruction.

Contents

1   Introduction

Constraining the physics of the early Universe is one of the key goals in cosmology, and various models have been proposed to describe the processes governing these early times and high energies. Among these, inflation has emerged as the most compelling framework [1–3], and it now constitutes one of the key pillars of the standard cosmological model, ΛCDM. According to the theory of inflation, quantum fluctuations during an early period of accelerated expansion give rise to initial perturbations, which evolve into the structures we see in the Universe today. The statistical properties of these fluctuations are encoded in the primordial power spectrum (PPS), making it a vital probe for understanding inflationary physics. These early perturbations are also imprinted on cosmological observables, with the Cosmic Microwave Background (CMB) serving as a particularly powerful tool for studying the PPS. This is largely due to the CMB's high-precision measurements and its ability to probe a wide range of spatial scales, providing strong constraints on the physics of the early Universe [4, 5]

    The simplest models of inflation, particularly canonical single-field slow-roll models [2], predict initial scalar[1] perturbations to be adiabatic and Gaussian. These fluctuations are fully described by a nearly scale-invariant PPS with amplitude A_s and scale-dependence

---

[1]While inflation also predicts the generation of tensor perturbations, we only focus on scalar perturbations for the remainder of this work.

characterised by the scalar spectral index n_s. However, more complex models of inflation predict deviations from near scale-invariance and can also lead to distinctive features in the PPS. As a specific example, models with step-like features or periodic oscillations in the inflaton potential can generate corresponding localised features in the PPS [6–9]. While recent observations have shown remarkable consistency with near-scale-invariance of the PPS (see e.g. Refs. [10, 11]), the physics of inflation remains unknown. Constraining the shape of the PPS thus offers a unique possibility to constrain the dynamics of the early Universe and test different models of inflation.

Deviations from near-scale invariance can also be constrained by going beyond the standard A_s and n_s parametrization of the PPS. The simplest model extensions allow for a scale-dependence of the scalar spectral index n_s, characterised by the so-called running of the spectral index, $\mathrm{d}n_s/\mathrm{d}\log k$, or the running of the running, $\mathrm{d}^2 n_s/\mathrm{d}\log k^2$. Several analyses, such as Refs. [11–16], have found these running terms to be small, consistent with single-field slow-roll inflation. In addition, various studies have explored oscillatory features that might be superimposed on the standard nearly scale-invariant spectrum, which could arise from more complex inflationary models or early-Universe physics [17–19].

Going beyond standard parametric models, the PPS can also be directly reconstructed from observational data. This approach allows for arbitrary shapes of the PPS, free from any assumptions about inflation, and thus facilitates the comparison with a variety of inflationary models. A number of different methods for PPS reconstruction have been considered in the literature. One common approach is wavenumber binning, where the PPS is divided into discrete intervals of k, with the amplitude estimated independently in each bin [12, 20–22]. Alternatively, wavelet expansions decompose the PPS into localized basis functions at different scales [23, 24]. Smoothing splines provide a flexible, regularized method to fit the PPS with smooth functions [25–27]. Other techniques include using inversion methods to directly solve the integral equations linking the PPS to cosmological observables (e.g., CMB anisotropies) using transfer functions [28–30], Bayesian approaches leveraging probabilistic frameworks for PPS inference [31–34], and deconvolution algorithms such as the Richardson-Lucy method iteratively correcting for the blurring effect of transfer functions, similar to image deblurring techniques [35–38].

Building upon these analyses, in this work we use Gaussian processes (GPs) to reconstruct the PPS. While not entirely non-parametric, GPs allow for modelling functions with minimal assumptions about their functional form [39] as well as robust characterization of statistical uncertainties. GPs have thus been employed for model-independent reconstruction of cosmological quantities in several works. For instance, they have been used to reconstruct the inflaton's speed of sound from *Planck* 2018 data [40], to perform model-independent cosmography by mapping the cosmic expansion history [41], and to reconstruct the growth factor [42].

One key challenge in PPS reconstruction is the large number of parameters required to model cosmological observables, which makes traditional Markov chain Monte Carlo (MCMC) algorithms such as Metropolis-Hastings, extremely inefficient. In this work, we address this challenge by developing a fully differentiable pipeline built on neural network (NN) emulators for efficient theoretical predictions. This allows us to employ gradient-based sampling techniques such as Hamiltonian Monte Carlo (HMC) [43, 44] for efficiently constraining the large parameter space spanned by cosmological parameters and the free-form PPS. In this proof-of-concept study, we apply our pipeline to reconstruct the PPS from the CMB temperature power spectrum measured by the *Planck* Collaboration in 2018 [5, 10], and use it to derive
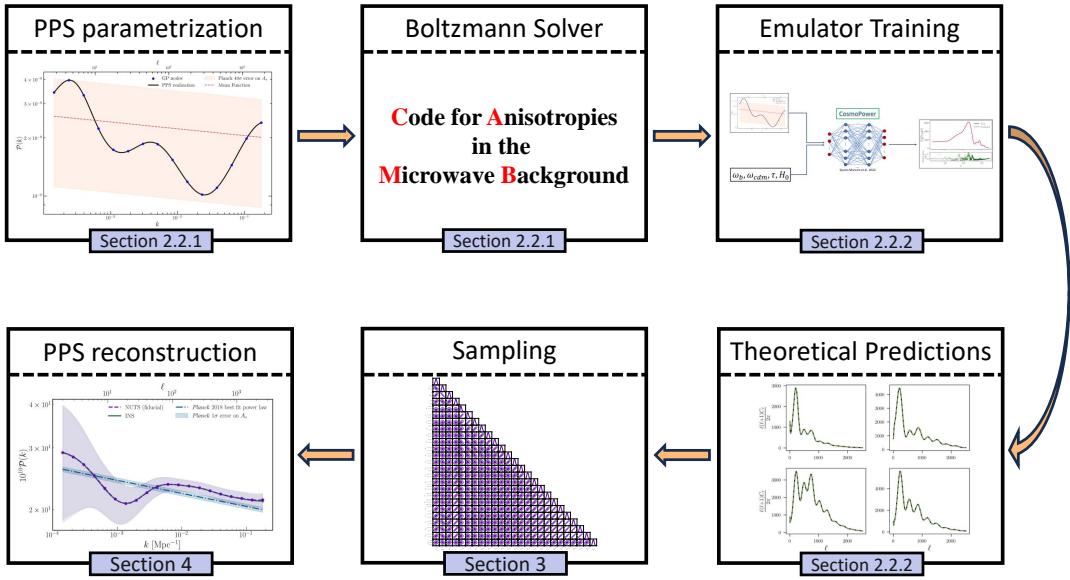
Figure 1: Overview of the differentiable pipeline for reconstructing the PPS. The top panel illustrates the data generation process and emulator training using CosmoPower (CP). The bottom panel shows the inference pipeline, leading to joint constraints on cosmological parameters and the free-form PPS, thus enabling its reconstruction.

joint constraints on cosmological parameters and the Gaussian process describing the PPS.

This paper is structured as follows: in Section 2, we outline our methodology for reconstructing the PPS from CMB temperature anisotropy data using GPs and NN emulators. Section 3 is dedicated to describing the data utilised in our study, along with the employed likelihoods. In Section 4, we present and discuss our results, and conclude in Section 5, where we also outline potential avenues for future research. Additional details regarding the analysis and methods are deferred to the appendices.

2 Methods

In the following, we describe our analysis pipeline to reconstruct the PPS. The main steps and their dedicated sections are illustrated in Fig. 1.

2.1 Primordial power spectrum reconstruction

The primordial scalar power spectrum is most commonly defined in terms of the variance of super-horizon comoving curvature perturbations, χ, in early radiation-domination through

$$\langle |\chi|^2 \rangle = \int d \log k \, \mathcal{P}_{\chi\chi}(k) \, , \tag{2.1}$$

where $\mathcal{P}_{\chi\chi}(k)$ denotes the dimensionless primordial curvature power spectrum, and log refers to natural logarithm throughout this work. As discussed in Sec. 1, the simplest models of

inflation predict a nearly scale-invariant PPS, which can be parametrized as

$$\mathcal{P}_{\chi\chi}(k) = A_s \left(\frac{k}{k_0}\right)^{n_s - 1} . \tag{2.2}$$

Here, A_s denotes the power spectrum amplitude and n_s is the scalar spectral index, with $n_s = 1$ for a scale-invariant spectrum. The quantity k_0 denotes a pivot scale, which, following Ref. [10], we set to $0.05\ \mathrm{Mpc}^{-1}$ throughout our analysis. Finally, to ease notation, we will omit the subscript χ for the remainder of this work, i.e. we set $\mathcal{P}(k) \equiv \mathcal{P}_{\chi\chi}(k)$.

The statistical properties of the primordial fluctuations are reflected in the temperature and polarisation anisotropies of the CMB. While CMB temperature anisotropies are characterised by the scalar field T, the spin-2 CMB polarisation field is fully described by the curl-free E- and divergence-free B-modes. The relation between the PPS, $\mathcal{P}(k)$, and the CMB anisotropy power spectrum, C_ℓ^{XY}, is established through the transfer function $T_\ell^{\mathsf{X}}(k)$, which describes how primordial fluctuations evolve into observed anisotropies. This can be expressed through (see e.g. Ref. [45])

$$C_\ell^{\mathsf{XY}} \propto \int \mathrm{d}\log k\ \mathcal{P}(k) T_\ell^{\mathsf{X}}(k) T_\ell^{\mathsf{Y}}(k) , \tag{2.3}$$

where X and Y denote T, E, or B. The transfer function for a given wavenumber k approximately peaks at an angular multipole ℓ given by

$$\ell \simeq kd , \tag{2.4}$$

where d denotes the comoving distance to the last scattering surface, which is roughly given by $d = 14,000\ \mathrm{Mpc}$ for the best-fit cosmological parameters from *Planck* 2018 [10].[2]

In this work, we employ Gaussian processes to reconstruct the PPS from the CMB temperature power spectrum in order to study deviations from near scale-invariance. The main reasons for this choice are twofold. First, GPs allow for reconstructing functions with minimal assumptions on their form except for smoothness, which is a general expectation for the PPS (see e.g. Ref. [22, 26, 46]). And second, GPs allow for straightforward quantification of uncertainties, which facilitates the interpretation of the reconstructed PPS (see Section 4 for more details).

Generally, a GP can be thought of as a framework for describing a distribution of functions, where the values of the function at different points (nodes) are treated as an ensemble of random variables. These are drawn from a multivariate Gaussian distribution governed by a mean function and a covariance function or kernel. The mean function represents the expected value of the process at each node, while the covariance function describes how the values at different nodes are correlated, effectively encoding the smoothness and structure of the process across scales. Therefore, a generic GP at nodes $\mathbf{x}$, $g(\mathbf{x})$, is fully specified by a mean function $m(\mathbf{x}) = \langle g(\mathbf{x})\rangle$ as well as a covariance function $K(\mathbf{x}, \mathbf{x}') = \langle (g(\mathbf{x}) - m(\mathbf{x}))(g(\mathbf{x}') - m(\mathbf{x}'))^T\rangle$ where $\langle \cdots \rangle$ denotes expectation values. In the following, we will briefly discuss the aspects of GPs most relevant to this work but refer the reader to Ref. [39] for a comprehensive review.

In this work, we aim to reconstruct the PPS, $\mathcal{P}(k)$, by modelling the logarithm of the PPS as a GP over a set of nodes in $\log k$-space, i.e., $\log \mathcal{P}(\log k)$. This choice ensures that correlations between nodes are applied in logarithmic space, leading to a constant correlation

---

[2]We note that, based on the *Planck* 2018 best-fit parameters from Table 2, column 2 (`TT+lowE`) of Ref. [10], we obtain $d \simeq 13,876\ \mathrm{Mpc}$. However, for simplicity and consistency in our analysis, we use $d = 14,000\ \mathrm{Mpc}$.
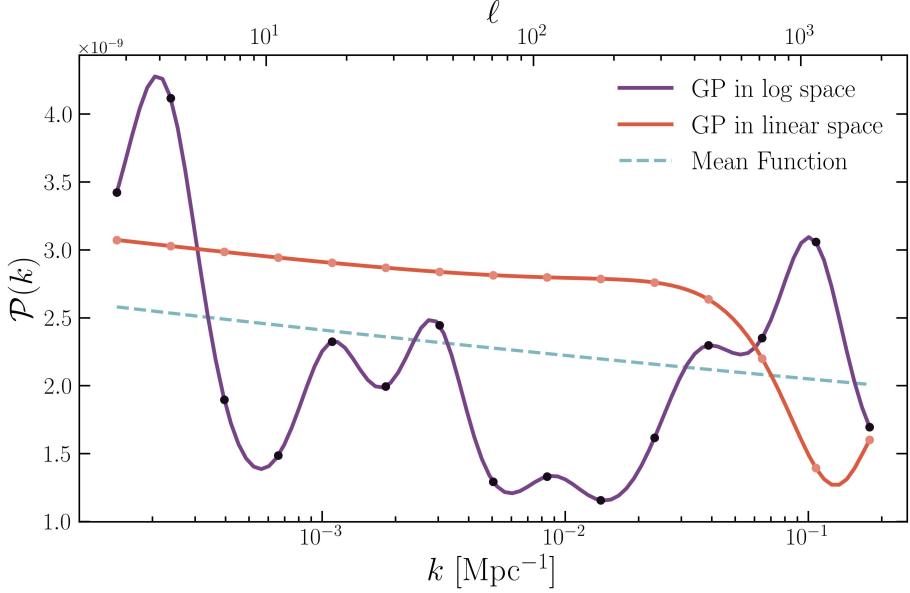
Figure 2: Comparison of two GP realizations of the PPS (see Eq. 2.7), generated by imposing correlations between the nodes in linear- and logarithmic-space. Note that the GP hyperparameters used to create the realizations are chosen solely to highlight the differences between the two spaces for clearer visualization; these are not the hyperparameters used in the analysis.

length (as defined by the covariance kernel in Eq. 2.6) across the entire wavenumber range. In contrast, given the range of scales relevant for PPS reconstruction, modelling correlations in linear space causes strong correlations at large scales (small k) and weak or negligible correlations at small scales (large k). This effect is illustrated in Fig. 2, where two GP realizations of the PPS show that correlations between nodes remain constant when imposed in logarithmic space, but decay as a function of k when imposed in linear space. Additionally, modelling $\log \mathcal{P}(\log k)$ naturally prevents unphysical negative values of the PPS, ensuring physical consistency in the reconstruction.

As our mean function, we choose the best-fit power law (see Eq. 2.2) obtained from *Planck* 2018 [10], i.e.

$$\log \mathcal{P}_{\mathrm{P18}}(\log k) = \log(A_s) + (n_s - 1)(\log k - \log k_0) \ . \tag{2.5}$$

We set $A_s = 2.092 \times 10^{-9}$ and $n_s = 0.9626$ based on the results reported by the *Planck* Collaboration in Ref. [10][3]. Following the approach of Ref. [40], we use a squared exponential kernel as our covariance function, i.e. we set

$$\mathbf{K}(\log k_i, \log k_{i+1}) = \sigma_f^2 \exp\left(-\frac{(\log k_i - \log k_{i+1})^2}{2l^2}\right) \ . \tag{2.6}$$

The squared exponential kernel depends on two so-called hyperparameters, l and σ_f^2, which control the behaviour of the GP. The correlation length l determines the rate at which the

---

[3]Specifically, we take these values from the `TT+lowE` constraints listed in Table 2, column 2 of Ref. [10].

correlation between the function values decays as the distance between the nodes increases, controlling the smoothness of the generated values. The variance σ_f^2 controls the overall scale of the function values around the mean function and thus effectively sets the spread of the generated realizations[4]. The choice of using this kernel is motivated by the fact that we expect the PPS to be a continuous function with smooth deviations from its power law form (Eq. 2.5), two properties that are naturally imposed on the GP by our use of the infinitely differentiable exponential kernel.

In our fiducial analysis, we evaluate the above GP at 20 nodes in k, logarithmically spaced between 1.429×10^{-4} Mpc^{-1} and 1.791×10^{-1} Mpc^{-1}. This range is calculated using Eq. 2.4 for the multipole range $\ell = 2$ to $\ell = 2508$ covered by the *Planck* 2018 data [5, 10]. In other words, we choose to reconstruct the PPS only within the scales covered by *Planck* and do not perform any extrapolations. The choice of the number of nodes is arbitrary a priori, and we have investigated the impact of varying this quantity in Appendix A.

Putting everything together, our full GP PPS is given by

$$\log \mathcal{P}(\log k) = \log \mathcal{P}_{\text{P18}}(\log k) + \mathbf{L}_C(\mathbf{K}) \cdot \mathbf{z}, \tag{2.7}$$

where $\log \mathcal{P}_{\text{P18}}(\log k)$ is the mean function of our GP given in Eq. 2.5, and $\mathbf{L}_C(\mathbf{K})$ denotes the lower triangular matrix obtained from the Cholesky decomposition of the covariance matrix $\mathbf{K}$ defined in Eq. 2.6. Finally, $\mathbf{z}$ is a vector of random variables, sampled from a standard normal, which we use to generate realizations of the PPS around the mean function. Eq. 2.7 can be interpreted as follows: the matrix $\mathbf{L}_C(\mathbf{K})$ transforms the initially uncorrelated elements of $\mathbf{z}$ into a set of correlated variations that follow the structure specified by the covariance function $\mathbf{K}$. By combining these correlated variations with the mean function $\log \mathcal{P}_{\text{P18}}(\log k)$, we obtain GP realizations of the PPS that smoothly vary across scales. Thus, in this model, reconstructing the PPS is equivalent to constraining the values of the random vector $\mathbf{z}$.

As has been noted in several works (see e.g. Refs. [47–49]), the choice of the GP mean can bias the final reconstruction. This is of particular concern in regions where there is little data, in which the GP is naturally drawn towards its mean function. We do not expect this to be a major issue in our analysis as, given previous measurements, we anticipate only small deviations of the reconstructed PPS from the assumed power law (Eq. 2.5). In addition, we manifestly restrict our reconstruction to regions covered by the observational data. Nevertheless, in Sec. 4, we test the robustness of our results to changes in the mean function.

2.2 Theoretical predictions

As seen in Eq. 2.3 the CMB power spectrum depends both on the PPS and the cosmological parameters that determine the shape of the transfer function, leading to potential degeneracies and thus necessitating a joint estimation. Jointly estimating these cosmological parameters and the nodes of the PPS GP (in our fiducial setup we use 20 nodes), presents a significant computational challenge due to the high dimensionality of the parameter space.

To address this challenge, we use emulators to compute theoretical predictions. Essentially, emulators are NNs that approximate the output of computationally-intensive functions, such as Boltzmann solvers in our case. These types of emulators have become essential in cosmology in order to sample the large parameter space needed to describe observables in a

---

[4]Throughout this work, we denote the second hyperparameter as σ_f, representing the standard deviation of the signal, which is the square root of the variance σ_f^2.

Parameter	Prior Distribution
ω_b	$\mathcal{U}[0.015, 0.03]$
ω_{cdm}	$\mathcal{U}[0.05, 0.3]$
H_0	$\mathcal{U}[50, 90]$
τ	$\mathcal{U}[0.01, 0.3]$
GP nodes	$\mathcal{N}(0, 1)$

Table 1: Prior distributions for the cosmological parameters and GP nodes used in the NN emulators described in Section 2.2.1. Here, $\mathcal{U}[a,b]$ denotes a uniform distribution and $\mathcal{N}(0, 1)$ a standard normal distribution.

reasonable time (see e.g. Refs. [50–57]). In this work, we use the popular CP framework [56] for building our emulators. An additional advantage of emulators is that, being NNs, they are differentiable by definition, and thus allow for the use of gradient-based sampling methods, enabling us to explore the large parameter space more effectively.

2.2.1 Emulator

In this work, we model the CMB temperature power spectrum as a function of four cosmological parameters: the physical baryon density, ω_b, the physical cold dark matter density, $\omega_{\rm cdm}$, the Hubble parameter today, H_0, and the optical depth to reionization, τ. In our fiducial setup, we additionally allow the PPS to vary at the 20 discrete k-nodes as described in Section 2.1, and denote these by $\{\mathcal{P}_i\}_{i=1}^{20}$, where $\mathcal{P}_i$ represents the power spectrum at the i-th k-mode. The remaining parameters specifying our cosmological model are kept constant throughout our analysis: in particular we assume a flat universe (i.e. we set the curvature density parameter Ω_K to 0), and assume dark energy to be described by a cosmological constant. Furthermore, we assume one massive neutrino with mass of 0.06 eV.

We train a NN emulator to learn the mapping between these 24 input parameters and the CMB temperature power spectrum at angular multipoles ℓ ranging from 2 to 2508 as computed using the Boltzmann code CAMB [58]. We use standard CAMB accuracy settings, which have been verified for *Planck* precision, and as discussed in Appendix B find these to be sufficient for the purpose of our analysis. CAMB uses the cosmological parameters and $\mathcal{P}_i$, as inputs to compute the corresponding CMB temperature power spectrum. To input our PPS into CAMB, we first subsample the 20 parametrized PPS values onto a finer grid of 100 nodes using our GP model, and then create a third-order spline interpolator, which we finally pass to the Boltzmann solver.

To ensure the robustness of this approach, we have tested various interpolation methods (including performing the interpolation within CAMB) and found that the choice of method does not significantly affect the results in our analysis.

To generate the training data for the NN emulator, we follow the workflow described in Ref. [56][5]. We sample the cosmological parameters and PPS nodes using Latin hypercube sampling (LHS) within the prior ranges specified in Tab. 1. We note that as the priors on the PPS nodes are Gaussian, we use inverse transform sampling to map the LH samples to standard normal random variables. Our training set consists of approximately 6×10^4

---

[5]For a comprehensive tutorial, please refer to the GitHub repository: CosmoPower.

samples[6], and for each point in parameter space we compute the CMB temperature power spectrum as described above.

Here, we do not attempt to also constrain the GP hyperparameters and thus set them to fixed values when training our emulator. In order to determine a suitable value for σ_f, in our fiducial analysis, we choose it such that the generated primordial power spectra (PPSa) cover a broad range of variations around the *Planck* best-fit power law $\mathcal{P}_{\mathrm{P18}}$. Specifically, we set it to encompass deviations of around $30 - 40\sigma$, where σ denotes the *Planck* 2018 uncertainty on A_s [10]. As we are imposing the correlations between the nodes in logarithmic space, this amounts to setting $\sigma_f = 0.39$[7]. In our fiducial analysis, we furthermore set the length scale hyperparameter to $l = 0.95$. We find that the rapidly varying PPSa obtained for low correlations lengths can cause the NN training to fail due to more features or "spikes" in the generated data. Therefore, we set l to the lowest possible value that still allows our NN to learn efficiently.

Fixing the hyperparameters of the GP can potentially lead to biases in the reconstructed PPS. As per our discussion of the GP mean function, this is of particular concern in those regions of k-space where data is sparse (see e.g. Ref. [47] for a related discussion). As we intentionally restrict our reconstruction to those scales fully covered by the CMB data, this is less of a concern, but we nevertheless test the robustness of our analysis to these hyperparameter choices. Specifically, we run a suite of different emulators with different hyperparameter settings and compare their performance in Sec. 4 and Appendix A. It will be interesting to extend the framework presented in this analysis to also constrain GP hyperparameters alongside the cosmological parameters and PPS nodes, but we leave an investigation thereof to future work.

2.2.2 Neural network training and evaluation

We evaluate the performance of our emulator training using a loss curve, and select the model with the minimum validation loss. For further details on the architecture of our emulators and the training procedure, the reader is referred to Appendix B.

As we are essentially training an emulator to predict CMB temperature power spectra for free-form PPSa, it is important to make sure that our emulators can capture the complex features arising in such an analysis. To this end, we generate a designated test set, which we produce in the same manner as the training set and which consists of approximately 6×10^3 samples. We then evaluate the performance of our emulators by comparing the prediction accuracy to the statistical uncertainties for two experiments: following Ref. [56], we first compare to the expected instrumental noise and cosmic variance from the upcoming Simons Observatory (SO) [59]. In addition, we also compare against the statistical errors obtained in the *Planck* 2018 analysis [5].

Following [56], the emulation error can be computed as

$$\frac{\left| C_{\ell,\mathrm{emulated}}^{\mathrm{TT}} - C_{\ell,\mathrm{true}}^{\mathrm{TT}} \right|}{\sigma_\ell^{\mathrm{TT}}} \, , \tag{2.8}$$

---

[6]Note that it took ~ 2 hours to generate the samples on 48 CPU cores. Also, the exact number of samples may vary, as certain parameter combinations from the hypercube can lead to integration timeouts during the computation of the CMB temperature power spectrum using CAMB. In such cases, we exclude these sets from the analysis.

[7]By comparing the spread of uncertainties around the mean function in logarithmic space with those in linear space, we ensure that this choice appropriately captures the expected deviations from $\mathcal{P}_{\mathrm{P18}}$.
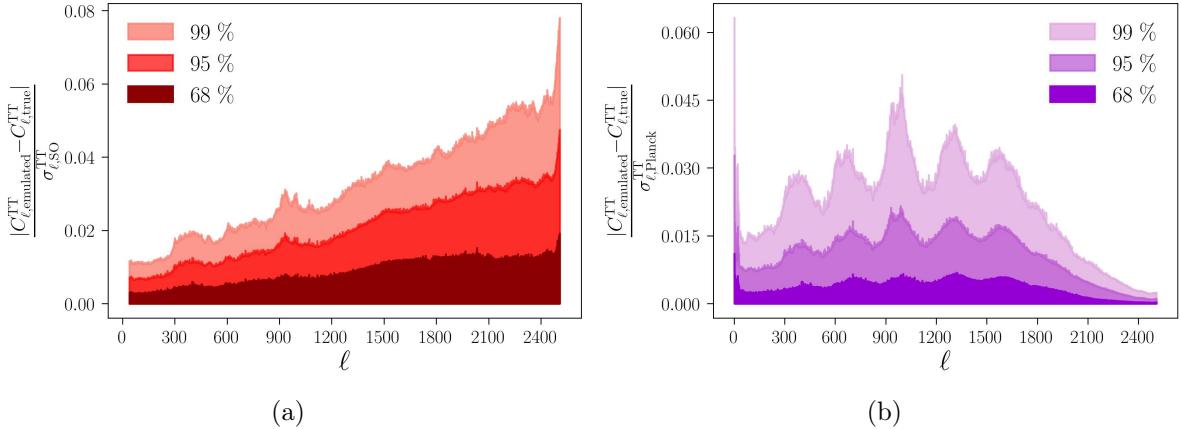
Figure 3: Emulation accuracy of the predicted CMB temperature power spectra compared with (a) expected Simons Observatory uncertainties across the multipole range used in the analysis and (b) *Planck* 2018 error bars. The different colours indicate regions encompassing the 68th, 95th, and 99th percentiles of the test set.

where $C^{\mathrm{TT}}_{\ell,\mathrm{emulated}}$ is the emulated CMB temperature power spectrum and $C^{\mathrm{TT}}_{\ell,\mathrm{true}}$ is the true power spectrum generated using CAMB. The quantity $\sigma^{\mathrm{TT}}_{\ell}$ denotes the experimental uncertainty. For SO, we follow Ref. [56] and set

$$\sigma^{\mathrm{TT}}_{\ell,\mathrm{SO}} = \sqrt{\frac{2}{f_{\mathrm{sky}}(2\ell + 1)}}\left(C^{\mathrm{TT}}_{\ell,\mathrm{true}} + N^{\mathrm{TT}}_{\ell}\right) \ . \tag{2.9}$$

Here, $f_{\mathrm{sky}} = 0.4$ represents the observed sky fraction, while N^{TT}_{ℓ} denotes the goal noise curves that account for instrumental and atmospheric effects, as specified for the Simons Observatory [59][8]. For *Planck* we use $\sigma^{\mathrm{TT}}_{\ell,\mathrm{Planck}}$ which represents the error bars from the *Planck* 2018 CMB temperature anisotropy data [5, 10]. The results of this analysis can be seen in Fig. 3 where we show the emulation error as a function of angular multipole ℓ for both cases. As can be seen, nearly 99% of the obtained deviations fall within less than $0.08\sigma^{\mathrm{TT}}_{\ell,\mathrm{SO}}$ and $0.06\sigma^{\mathrm{TT}}_{\ell,\mathrm{Planck}}$ respectively, thus demonstrating the reliability of the emulator and its ability to reproduce the complex power spectra required for our analysis.

To further illustrate the emulator performance, in Fig. 4 we show the comparison between an arbitrary ground truth CMB temperature power spectrum from our test set and its corresponding emulator prediction. As can be seen, these two agree remarkably well, despite the non-trivial features exhibited by the CMB power spectrum[9]. Given the large number of parameters involved in the emulation, this robustness is especially crucial to ensure accurate predictions across the parameter space.

3   Data and Likelihood

To reconstruct the PPS, we employ CMB temperature power spectra from the third data release of *Planck* (PR3)[10][4, 5, 10]. Following Ref. [5], we employ two different likelihoods for

---

[8]For large angular scales that are not probed by SO, it incorporates Planck data over a reduced sky fraction to model the co-added noise curves [59].

[9]We note that these are entirely determined by corresponding features in the PPS.

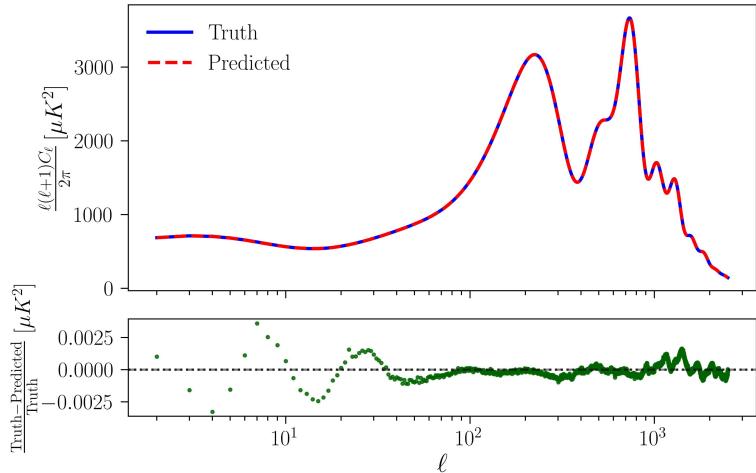[10]The data is obtained from *Planck* Legacy Archive: https://pla.esac.esa.int

Figure 4: Comparison of an emulated CMB temperature power spectrum to the ground truth from CAMB for a randomly selected sample from the test set, and associated relative residuals between the two.

the large-scale and small-scale CMB temperature power spectrum.

For the high-ℓ ($\ell \geq 30$) part of the CMB likelihood, we follow the PYTHON implementation of *Planck*'s `plik_lite` likelihood [5], `Planck-lite-py`[11], which is described in Ref. [60]. This likelihood is pre-marginalised over *Planck* high-ℓ CMB temperature foreground parameters, but has been shown to be consistent with the full likelihood in Ref. [60]. The only remaining nuisance parameters in `Planck-lite-py` is the overall multiplicative *Planck* calibration factor, A_{Planck}.

At low multipoles, $2 < \ell < 30$, the likelihood of the angular power spectrum is non-Gaussian. While the official *Planck* analysis employs a hybrid, map-based approach [5], Ref. [60] showed that the low-ℓ CMB temperature power spectrum can be efficiently compressed into two Gaussian-distributed data points. In our analysis, we employ the *Planck* 2018 version of the compressed low-ℓ likelihood presented in Ref. [60][12], which has been shown to give results consistent with the official analysis.

To perform cosmological inference from the *Planck* low- and high-ℓ likelihoods, we employ two complementary sampling algorithms and associated packages. The first approach uses Importance Nested Sampling (INS) [61] as implemented in the NAUTILUS [62] package[13].

Nested Sampling [63] is an algorithm to generate samples from posterior distributions and compute Bayesian evidences. It starts with a set of "live points" drawn from the prior distribution. At each iteration, the point with the lowest likelihood is replaced by a new point with a higher likelihood, progressively focusing on high-likelihood regions. This process continues until further contributions to the evidence are negligible. The prior volume represented by the live set decreases exponentially and the weighted posterior distribution is reconstructed using both the inactive and remaining live points.

INS enhances standard nested sampling by making better use of all likelihood evalu-

---

[11]The code is available at https://github.com/heatherprince/planck-lite-py.git.

[12]The low-ℓ-compressed likelihood, as two Gaussian bins, is implemented in the same code as an option to be included alongside the high-ℓ likelihood via a boolean flag

[13]https://github.com/johannesulf/nautilus.git

Parameter	Prior Distribution	Best Fit Value
ω_{b}	$\mathcal{U}[0.015,\ 0.03]$	0.02191 ± 0.00039
ω_{cdm}	$\mathcal{U}[0.05,\ 0.3]$	0.1218 ± 0.0029
H_0	$\mathcal{U}[50,\ 90]$	66.26 ± 1.35
τ	$\mathcal{N}(0.067,\ 0.023)$	0.0745 ± 0.0221
A_{Planck}	$\mathcal{N}(1,\ 0.0025)$	1.001 ± 0.002
GP Nodes	$\mathcal{N}(0,\ 1)$	See Fig. 5

Table 2: Cosmological and nuisance parameters varied in this work, alongside their associated priors and posterior means obtained in our fiducial analysis. The uncertainties denote the 68% C.L. To improve readability, we do not show the recovered GP nodes in the table.

ations. Instead of assigning posterior sample weights based solely on the shrinking-volume assumption, INS estimates the underlying sampling distribution and uses this to assign importance weights to all sampled points, including those outside the live set. This leads to a more accurate reconstruction of both the posterior and the Bayesian evidence [61, 62]. NAU-TILUS further advances this methodology by incorporating NN regression to refine proposal distributions, concentrating sampling in regions of high posterior probability while ensuring thorough exploration of the entire parameter space. Additionally, it allows for drawing extra samples after the initial run, increasing the effective sample size and further improving the accuracy of the estimated posterior [62].

As a consistency test, we compare the results from NAUTILUS with those from EMCEE, a standard MCMC sampler [64]. To simplify the comparison, we fix the 20 GP nodes to the *Planck* 2018 best-fit mean function (see Eq. 2.5) and sample only the 4 cosmological parameters and the nuisance parameter A_{Planck}. We find that both methods yield consistent results.

A key challenge with stochastic sampling algorithms such as traditional MCMC and INS is that they can become very inefficient in exploring complex, high-dimensional posterior distributions. This is mainly due to the fact that finding the high-posterior regions in these cases becomes increasingly difficult. In order to overcome this challenge, gradient-based samplers, such as HMC, leverage the gradient of the posterior distribution to guide the sampling process, resulting in a more efficient exploration of parameter space [43, 44, 65]. HMC simulates Hamiltonian trajectories through parameter space to efficiently propose new points, which can significantly speed up the exploration of high-dimensional parameter spaces. However, the performance of HMC relies on two key parameters, the step size and the number of steps employed when simulating the Hamiltonian trajectories, which thus need to be adapted to the problem at hand. The No-U-Turn Sampler (NUTS) [66] is an extension of HMC, which dynamically tunes these parameters. The essential idea underlying NUTS is to prevent the Hamiltonian trajectories to turn back on themselves by avoiding U-turns in parameter space.

Given the high-dimensional parameter spaces involved in our analysis, we additionally employ HMC sampling methods. Since HMC relies on gradient information, it necessitates a fully differentiable analysis pipeline. To this end, we implement the CMB temperature likelihoods using the automatically differentiable `JAX` library [67, 68]. These likelihoods are interfaced with our NN emulators through the `Cosmopower-JAX` framework[14] to ensure that

---
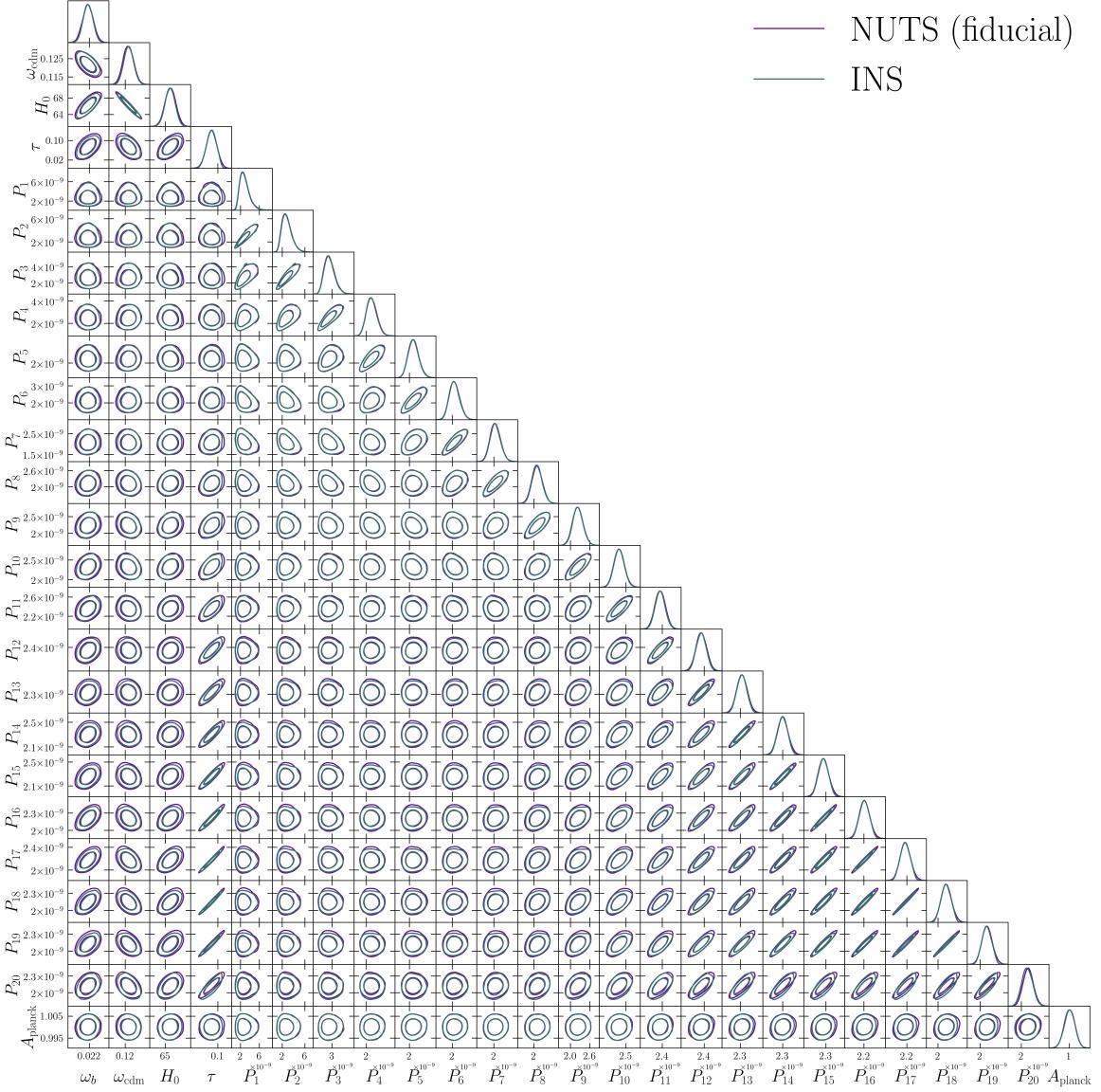
[14]https://github.com/dpiras/cosmopower-jax.git

Figure 5: Constraints on the cosmological parameters, the PPS values calculated at each mode using the constrained GP nodes and the nuisance parameter obtained in our fiducial analysis with NUTS and INS. The contours represent the 68% and 95% confidence intervals, with 1-D marginals for each parameter.

our theoretical predictions remain compatible with `JAX` [57]. We then use the NUTS variant of HMC as implemented in the `NumPyro` library[15] [69, 70] to obtain joint constraints on cosmological and PPS parameters.

The priors used in our analysis are summarized in Tab. 2; these priors are mostly flat and match the ranges used for our emulators (see Tab. 1), with two exceptions. First, we follow Ref. [5] and impose a Gaussian prior on the *Planck* calibration parameter A_{Planck}. Second, as we do not include polarisation information in our analysis, we need to incorporate

---

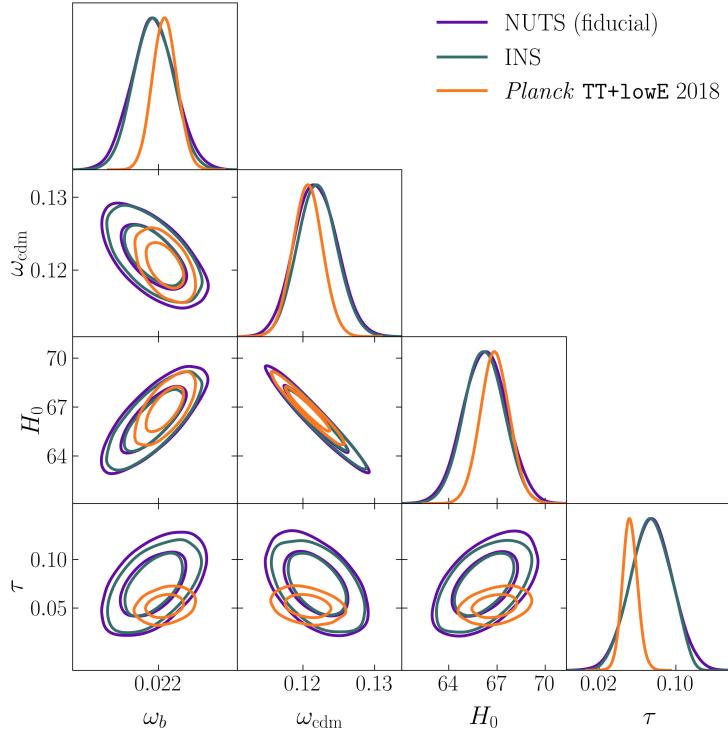[15]https://github.com/pyro-ppl/numpyro.git

Figure 6: Constraints on the cosmological parameters from our fiducial analysis compared to the *Planck* 2018 `TT+lowE` results. The PPS and nuisance parameter have been marginalized over. The contours represent the 68% and 95% confidence intervals, with 1-D marginals for each parameter.

it in the form of a prior on the optical depth to reionization, τ, as otherwise we would not be able to constrain the amplitude of the PPS due to its degeneracy with τ. In our fiducial analysis, we follow a conservative approach and impose a Gaussian prior on τ that compresses the *Planck* 2015 low-ℓ polarisation data [71], following Ref. [60]. This is an approximation, as low-ℓ polarisation data also constrains other cosmological parameters and is thus not fully independent of the temperature information. However, we leave a full joint analysis of CMB temperature and polarisation to future work, but we do test the sensitivity of our results to the choice of τ prior in Sec. 4; in particular, we investigate the effect of incorporating the higher-precision τ measurements from *Planck* 2018 [5].

4 Results

Using our differentiable pipeline, we can now move on to reconstructing the PPS using *Planck* 2018 CMB temperature anisotropy data. In Fig. 5, we show the joint constraints on the PPS values obtained at each k mode, alongside the cosmological and nuisance parameters, as obtained in our fiducial analysis both for INS and NUTS. Figure 6 illustrates the same constraints but focuses on the cosmological parameters, and compares our results to those obtained from *Planck* 2018. Focusing first on the results obtained using our fiducial sampling algorithm NUTS, we can see from Figs. 5, 6 and Tab. 2 that the inferred constraints on the cosmological parameters ω_b, $\omega_{\rm cdm}$, and H_0, agree with the official *Planck* 2018 results to within 1σ [10], as expected. Specifically, we compare to the `TT+lowE` constraints listed in
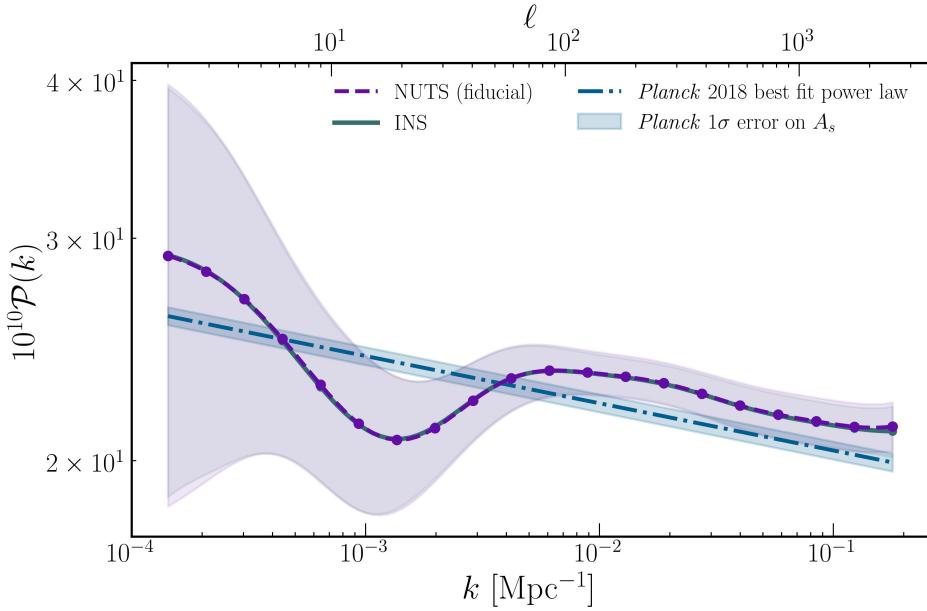
Figure 7: The mean reconstructed PPS for our fiducial setup obtained using INS and NUTS. The shaded regions denote the 68% confidence intervals, and the points highlight the GP nodes used in our analysis.

Table 2, column 2 of Ref. [10], which is the analysis variant closest to the data combination considered in this work. The difference in our inferred value of $\tau = 0.0745 \pm 0.0221$ and the value of $\tau = 0.0522 \pm 0.0080$ reported by the *Planck* Collaboration in their 2018 results is due to our use of a conservative prior based on the *Planck* 2015 results and Ref. [60]. The freedom allowed by our conservative prior choice pushes our fiducial constraints to higher values for the optical depth as compared to the results from *Planck* 2018 as we describe below.

In Fig. 7, we show the corresponding reconstructed PPS, alongside the best-fit power-law spectrum from *Planck* 2018. We compute the power-law using Eq. 2.2, with n_s and A_s values taken from the *Planck* 2018 `TT+lowE` results from Table 2 in Ref. [10]. As a rough visual illustration of the uncertainties on the PPS obtained from the official *Planck* 2018 analysis, the shaded region around the *Planck* power-law denotes variations allowed by the 1σ uncertainty on A_s as derived in Ref. [10]. For our PPS reconstruction, the solid and dashed line shows the mean across all samples, and the shaded region encompasses the 68% confidence interval. As can be seen, the uncertainties on the reconstructed PPS at small k are quite large. However, our results are in good agreement with the *Planck* power law. Despite the low signal-to-noise, our reconstructed PPS shows a hint of a dip at a scale of around $k \sim 1.57 \times 10^{-3}$ Mpc^{-1}. This corresponds to angular multipoles in the range $\ell \sim 15-30$, where all CMB analyses so far have reported a dip in the CMB temperature power spectrum (see e.g. Ref. [5, 72]). Our results suggest that this observed lack of power causes a corresponding dip in our reconstructed PPS, consistent with previous works (see e.g. [11, 73]). The significant uncertainties observed on the largest scales can primarily be attributed to cosmic variance. Additionally, we suspect that our use of the compressed `Planck-lite-py` low-ℓ likelihood, as discussed in Section 3, may contribute to these uncertainties. While Ref. [60] shows that the likelihood compression is lossless for a ΛCDM cosmological model, this might not be the case when allowing a free-form
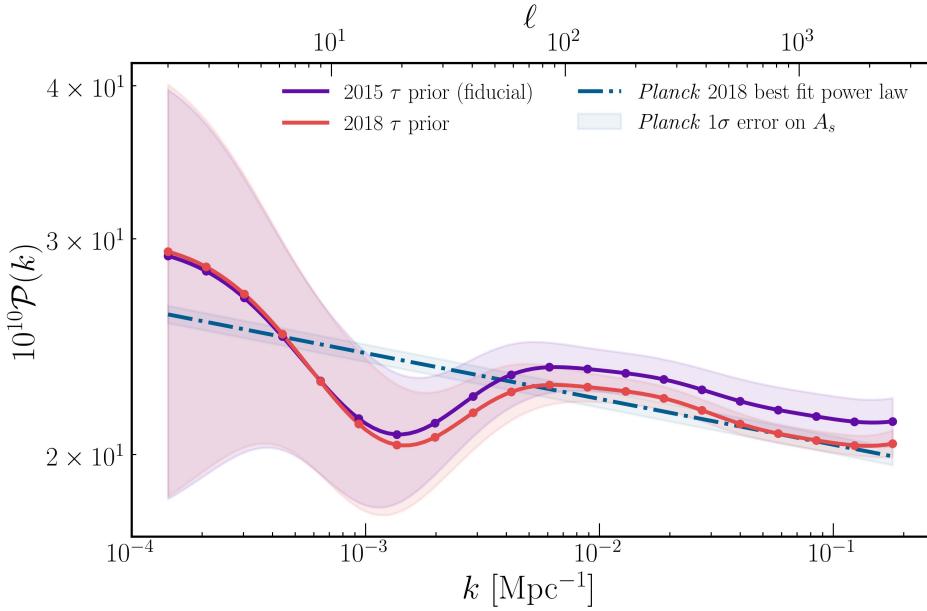
Figure 8: Comparison of the mean reconstructed PPS for our fiducial setup using two different τ priors from *Planck* 2015 and 2018. The shaded regions represent the 68% confidence interval.

PPS. Focusing on small scales (i.e. high k), we recover the nearly scale-invariant behaviour predicted by slow-roll single-field inflation. Comparing to the power-law PPS obtained in *Planck* 2018, while the two results remain consistent at 1σ, we find our reconstruction to systematically overestimate the PPS amplitude.

In order to better understand this discrepancy, we repeat our analysis using a τ prior informed by the low-ℓ polarisation likelihood from the updated *Planck* 2018 measurements, specifically from the High-Frequency Instrument [5, 74]. As compared to the Low-Frequency Instrument based results reported by *Planck* 2015, the HFI measurements result in a significantly lower value of the optical depth, which following Ref. [5], we summarize into a Gaussian prior with $\tau = 0.0506 \pm 0.0086$. The resulting reconstructed PPS is shown in Fig. 8 alongside our fiducial results; the full contour plot is deferred to the Appendix C. Adopting the *Planck* 2018 prior reduces the uncertainties on the PPS at small scales, indicating that these uncertainties are primarily driven by the choice of τ prior. In contrast, the uncertainties on large scales remain largely unchanged, as they are dominated by cosmic variance or the use of the compressed likelihood. Reassuringly, our results using the *Planck* 2018 optical depth prior move closer to the *Planck* 2018 best-fit power law model. This suggests that the overestimation of the PPS amplitude observed in our fiducial analysis is driven by our conservative τ prior choice. As discussed above, this leads to an inferred value of τ significantly higher than the official results from *Planck* 2018. Due to the degeneracy between τ and A_s, this higher optical depth leads to a correspondingly higher PPS amplitude, which we observe in our reconstruction. We note that the τ-A_s degeneracy can also be observed in Fig. 5 as a degeneracy between the high-k PPS nodes $\mathcal{P}_i$ and the optical depth τ.

To test the robustness of our HMC pipeline, we compare the fiducial constraints obtained using NUTS with those derived from INS. As shown in Figs. 5 and 6, the posterior
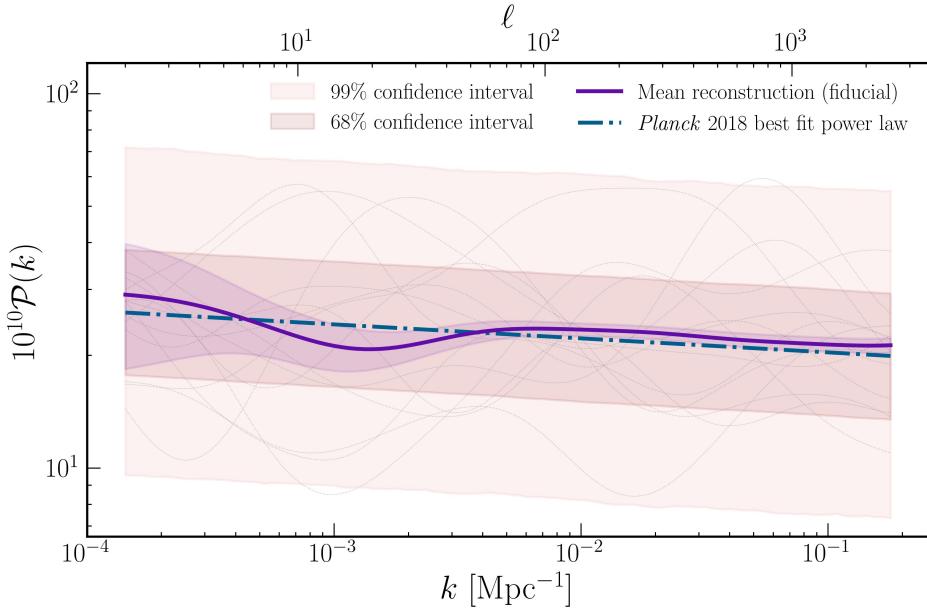
Figure 9: Fiducial PPS reconstruction compared to the prior imposed on the GP (see Sec. 2.2.1 for details on the GP prior). The shaded region around the *Planck* best-fit power law illustrates the 68% (dark) and 99% (light) percentile of the GP prior, while the gray dotted lines show random draws from this prior. The shaded region around the reconstruction represents the 68% confidence interval.

means from both NUTS and INS are consistent. This agreement is also supported by the mean reconstructed PPS in Fig. 7. In order to further assess the robustness of both the INS- and NUTS-based results, we perform a series of consistency checks by varying key hyperparameters of each algorithm. The details of these tests are provided in Appendix C. We find that our constraints remain stable under these variations.

Since the reconstructed PPS from both methods is consistent, we adopt gradient-based sampling using NUTS for the remainder of this analysis. This choice is also driven by the fact that NUTS significantly accelerates the inference process compared to INS, with NUTS taking 2 hours to obtain the constraints shown in Fig. 5 compared to around 2 days it took with INS. Details of these runtime comparisons and hardware used are included in Appendix C.

Finally, in Fig. 9, we show the comparison of our fiducial PPS reconstruction to the GP prior employed in our analysis. We show both the 68% and the 99% interval of random realizations drawn from the GP prior. As can be seen, our constraints on the PPS are prior-dominated at the largest scales (up to $k_{\min} \sim 3 \times 10^{-4}$), which we attribute to cosmic variance and our use of a compressed low-ℓ likelihood. On smaller scales on the other hand, the likelihood dominates over the prior, showing that both the lack of power observed at around $k \sim 1.57 \times 10^{-3}$ and the near scale-invariant behaviour at small scales are robust to our prior choice (we will come back to this when discussing GP hyperparameter choices below).

As discussed in Section 2.1, the properties of GPs are governed by two hyperparameters, the length scale l and the signal standard deviation σ_f, which we have fixed in our fiducial analysis. In order to ensure that our results are not driven by our particular choice of these
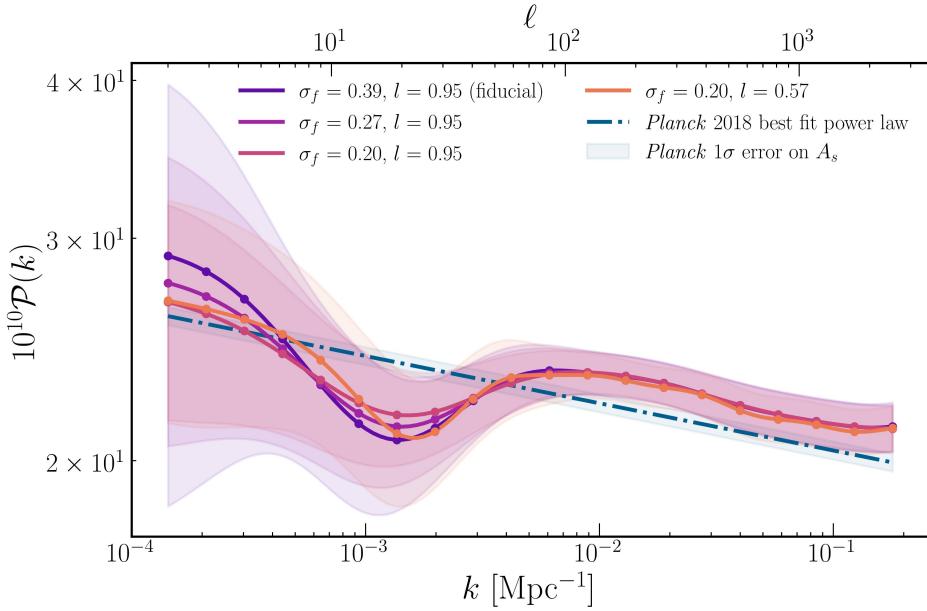
Figure 10: The mean reconstructed PPS for different GP hyperparameter settings. The shaded regions around the curves represent the 68% confidence intervals.

hyperparameters, we repeat our analysis for a number of different hyperparameter choices. Specifically, we train several emulators with different values for l and σ_f^2, while keeping all other settings consistent with those used in our fiducial analysis. The joint parameter constraints obtained for the four different cases considered are shown in Appendix A, with the corresponding reconstructed PPSa in Fig. 10. As can be seen, the mean reconstructed PPSa remain consistent for all cases considered. In particular, the nearly scale-invariant form of the reconstructed PPS at small scales is robust to any hyperparameter changes, as is the observed lack of power around $k \sim 1.57 \times 10^{-3}$ Mpc^{-1}. The only difference observed is that the uncertainties on the reconstructed PPSa on large scales decrease as we decrease σ_f. This is expected, since, as discussed above, the constraints at these scales are prior-driven and decreasing σ_f effectively tightens the prior on the GP. The constraints on the small-scale PPSa are unaffected, as in this regime our contours are driven by the data rather than the prior. To assess the robustness of our results against the choice of the GP mean function, we repeat our analysis using the reconstructed PPS from our fiducial results (which we obtain via the differentiable framework) as the new mean function. The resulting reconstructed PPS, shown in Fig. 11, remains consistent with the fiducial results, with some deviations on the prior-dominated regime which are well within the statistical uncertainties. On small scales, the reconstruction is fully data driven and consistent with the near–scale-invariant behaviour observed in the fiducial results. This demonstrates that our findings are not biased by the choice of the GP mean function, but are instead strongly driven by the data across the considered scales. An extension of this analysis would involve adopting the approach introduced in Ref. [47], where a free amplitude parameter is introduced to scale the GP mean function. This parametrization allows the mean itself to be constrained by the GP. We leave this more comprehensive investigation for future work.

Finally, the number of nodes with which we parametrize the GP represents an additional
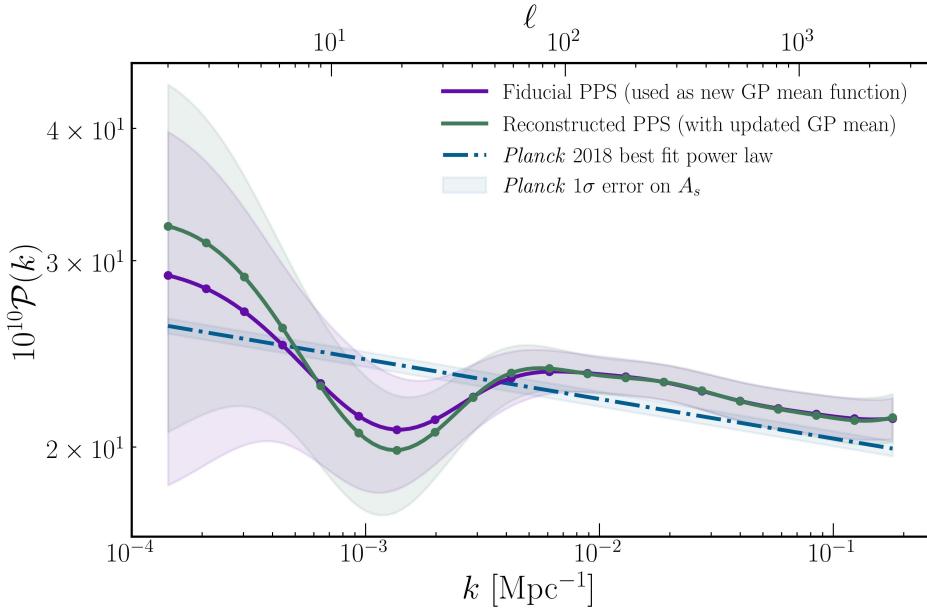
Figure 11: The fiducial reconstructed PPS is compared with the PPS reconstructed using the fiducial reconstruction as the mean function. The shaded region represents the 68% confidence interval.

hyperparameter in our analysis. As we show in Appendix A, we find our results to not depend on this choice, and thus conclude that our main results are robust to changes in the GP hyperparameters and the mean function, with only minor differences in the derived uncertainties in the prior-dominated regime.

5   Conclusions

In this work, we present a novel, fully-differentiable pipeline to reconstruct the primordial power spectrum (PPS) generated from inflation. We model the PPS using a Gaussian process (GP), which in our fiducial setup we parametrize through a set of $n = 20$ GP nodes, and obtain theoretical predictions for cosmological observables by training a neural network emulator based on the Boltzmann code CAMB and using the COSMOPOWER framework. To ensure differentiability, we implement the likelihood entirely in `JAX`, in particular making use of `Cosmopower-JAX`. This approach allows us to employ gradient-based sampling algorithms, particularly the No-U-Turn-Sampler extension of Hamiltonian Monte Carlo (HMC), to efficiently explore the parameter space spanned by the cosmological parameters and the Gaussian process nodes.

We then apply this pipeline to constrain the PPS using CMB temperature anisotropy data from *Planck* 2018, alongside a Gaussian prior on the optical depth to reionization based on *Planck* 2015 low-ℓ polarisation data. On small scales, our reconstructed PPS is consistent with near scale-invariance as predicted from single-field slow-roll inflation. On large scales, our constraints remain relatively weak, mostly due to cosmic variance. Nonetheless, we recover the well-known dip in the CMB temperature power spectrum at around $\ell = 22$ as a corresponding dip in the reconstructed PPS. Comparing to the ΛCDM results from *Planck* Collaboration

2018, we find the PPS reconstructed in this work to exhibit a slightly higher amplitude on small scales. This can be attributed to our conservative choice of prior on the optical depth to reionization τ, with our fiducial results based on the measurement from *Planck* 2015, which was later revised downwards in *Planck* 2018 due to the measurements from the High-Frequency instrument. Incorporating these latter measurements into our analysis brings our results into full agreement with the results from *Planck* 2018, which is due to the degeneracy between the primordial power spectrum amplitude and τ, such that lower optical depth leads to lower amplitude.

Finally, we test the robustness of our results by performing a suite of consistency tests. In particular, we compare our HMC-based results to results based on importance nested sampling, and we repeat our analysis using different GP hyperparameter choices. We find consistent results for all these analysis variants, thus demonstrating the robustness of our results.

The flexible and efficient pipeline developed in this work is well-suited for application to additional datasets. A natural extension of this work would be to incorporate CMB polarisation anisotropy data, thus mitigating the dependence of our results on the imposed τ-prior. Our analysis framework also allows for extensions beyond CMB data by incorporating large-scale structure observables such as cosmic shear or galaxy clustering. Therefore, this work lays the groundwork for multi-probe primordial power spectrum reconstruction, which will be crucial for obtaining tight and robust constraints on inflation using current as well as upcoming data. Furthermore, the methodology developed as part of this work can also be employed for future matter power spectrum reconstructions [e.g. 75–78].

A Validating changes in GP hyperparameters

A.1 Effect of changing the number of nodes of the GP

We present the reconstructed PPSa obtained by varying the number of nodes in the GP framework (see Sec. 2.1) while keeping the GP hyperparameters fixed to the values used in our fiducial analysis. As shown in the left panel of Fig. 12, changes in the number of nodes do not significantly affect the reconstruction—the PPS remains largely consistent across different configurations.

However, the choice of the number of GP nodes has important implications. With fixed hyperparameters, there exists a lower limit on the number of nodes required for a stable reconstruction. We find that reducing the number of nodes excessively, without adjusting the GP's amplitude (σ_f) and smoothness (l), can lead to irregular, highly fluctuating spectra. This irregularity can impair the NN emulator's ability to learn the mapping between input parameters and output features effectively.

Conversely, increasing the number of nodes, when combined with well-calibrated hyperparameters, raises the dimensionality of the problem. However, our fully differentiable framework is designed to efficiently manage this increased complexity while preserving robustness and scalability. To evaluate the consistency of our approach under more extreme conditions, we increased the number of nodes to 100 (see right panel of Fig. 12), while keeping all other hyperparameters fixed at their fiducial values; this expansion raises the total number of free parameters to 105. This high number of parameters showcases the potential for gradient-based sampling and differentiable programming to scale the inference to the complexity required by next-generation cosmological surveys [57, 79].
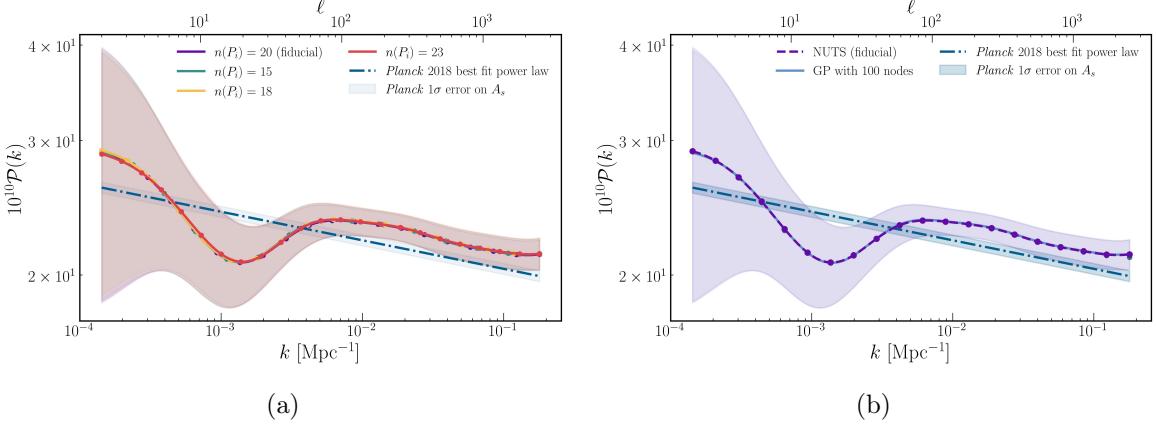
Figure 12: Reconstructed PPS shown with solid lines, and corresponding 68% confidence intervals indicated by shaded regions. (a) PPS reconstructions for different numbers of GP nodes across each k mode. (b) Comparison of the PPS reconstructed using 100 GP nodes with the fiducial reconstruction.

With a higher node count, one might expect the reconstruction to resolve finer features that were previously smoothed out. However, our results show that the reconstructed PPS remains consistent with the fiducial analysis, suggesting that no additional features can be extracted within the constraints of our current GP model and hyperparameter settings. A potential avenue for future investigation is to explore models with correlation lengths smaller than the node separation, which may reveal additional structure.

A.2 Effect of changing the GP hyperparameters

In Fig. 13, we present the joint constraints on the cosmological parameters, the PPS values at each k mode, and the nuisance parameter. The constraints on the cosmological parameters remain consistent across all hyperparameter settings. As discussed in the main text and compared with the reconstruction in Fig. 10, we find that the contours for small-scale $\mathcal{P}_i$ remain stable across different hyperparameter choices, aligning with the reconstructed PPS. On large scales, the variations are also consistent with the reconstruction, with the contours shrinking as σ_f is lowered.

B Neural network training and performance

We maintain the CP architecture consistent with the original implementation, with only minor modifications. One notable change is implementing a feature to iteratively save the training and validation losses for each learning rate. We also adjust the training loop to iterate over multiple learning rates, starting with larger updates in the initial stages to facilitate faster convergence and gradually decreasing the learning rate in later epochs to fine-tune the model for improved accuracy. We also increase the `batch_size` to 1024 and the `max_epochs` to 1500, providing the model with sufficient training time if needed. An early stopping mechanism is incorporated in the original framework to terminate training when further improvements are unlikely, ensuring computational efficiency. We trained our NN emulators on an NVIDIA A40 GPU, which significantly improved computational efficiency compared to CPU-based
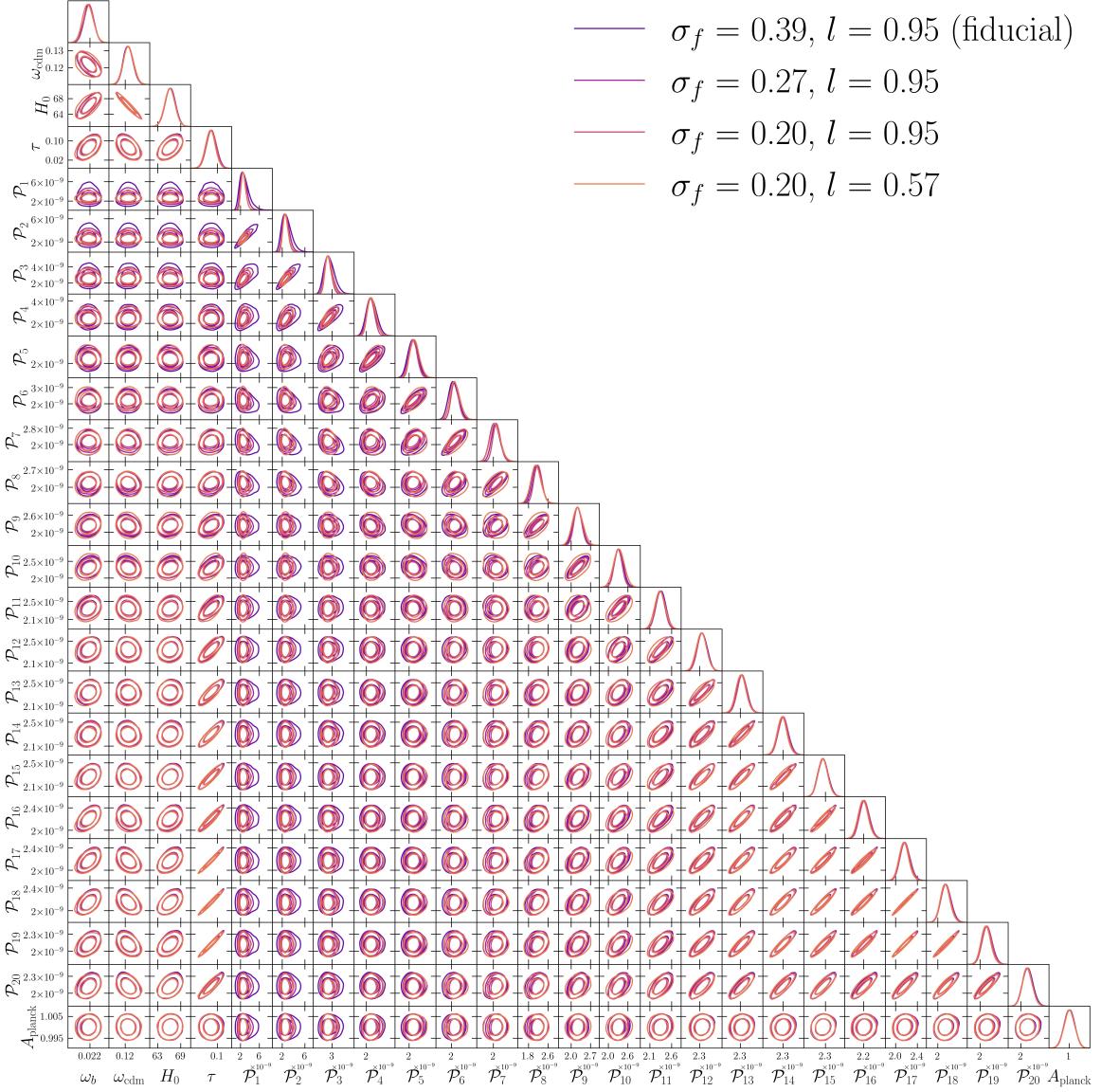
Figure 13: Constraints on cosmological and nuisance parameters, as well as PPS values at each k mode, obtained for different GP hyperparameter settings of l and σ_f, as indicated in the legend. The contours represent the 68% and 95% confidence regions, accompanied by 1-D marginal distributions for each parameter.

training. Although we experimented with training on a CPU, the GPU's parallel processing capabilities dramatically reduced training time and enhanced overall performance.

In Figure 14, we illustrate the performance of our trained emulator in terms of relative error. Across the entire multipole range, the relative error in emulation accuracy remains below 1.5%, demonstrating the emulator's capability to accurately reproduce any spectra within the training prior range specified in Table 1. When using CAMB to generate the corresponding angular power spectrum, we set the parameters `lmax` and `lens_potential_accuracy` to 2509 and 1, respectively. The latter ensures *Planck*-level accuracy for the lensing potential. As we
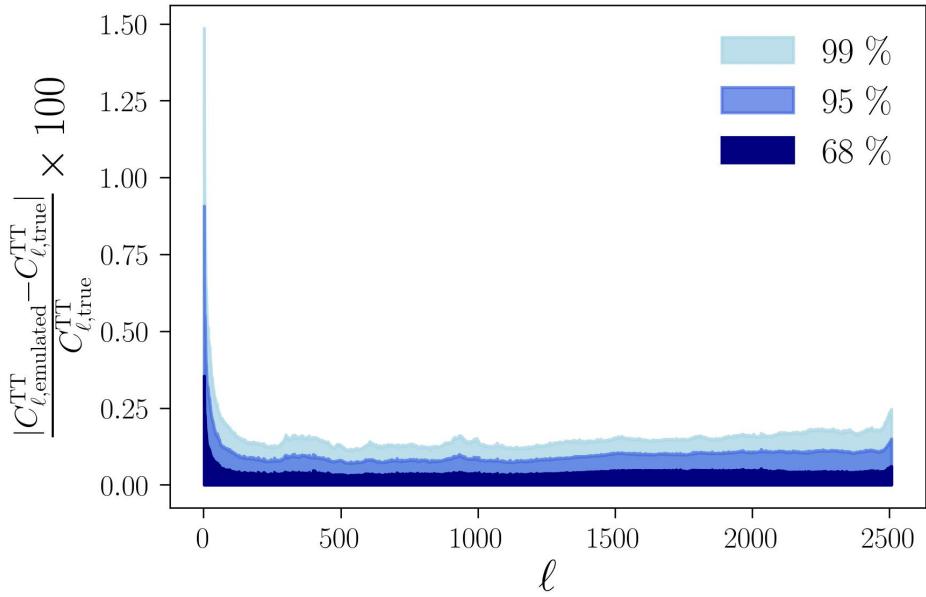
Figure 14: Emulation accuracy of the CMB power spectra emulator, calculated in terms of relative error. The different colours indicate regions corresponding to the 68th, 95th, and 99th percentiles of the test set as evaluated by the emulator.

are using *Planck* 2018 data, we do not need to adjust additional accuracy parameters. We have checked the effect of using higher accuracy settings for our analysis and we did not find any significant improvements.

C Sampling

C.1 Effect of imposed τ prior

In Fig. 15, we present the results of our joint constraints on cosmological parameters, PPS values at specific k modes, and the nuisance parameter for the updated τ prior, based on *Planck* 2018 low-ℓ polarisation data, compared to the fiducial results. Except for τ, all cosmological parameters remain consistent within 1σ. As discussed in the main text, the observed shift in τ arises from the updated prior informed by *Planck* 2018 [5]. Comparing this figure with Fig. 8, we find that constraints on $\mathcal{P}_i$ at large scales remain unchanged, as seen in the overlapping contours in Fig. 15. This is expected since large-scale constraints are cosmic variance-limited, and *Planck* 2018 did not significantly improve them over *Planck* 2015. At smaller scales, however, the contours for $\mathcal{P}_i$ become tighter, reflecting a closer match to the *Planck* 2018 best-fit power-law model. This trend is also apparent in Fig. 8 in the main text.

C.2 Testing robustness of sampling techniques

To ensure the robustness of our INS-based results, we conduct a series of tests by varying key hyperparameters in the NAUTILUS implementation. Specifically, we explore different values for the number of live points, n_{live}, and the effective sample size, N_{eff}, which control the precision and efficiency of the sampling process. Across all hyperparameter settings
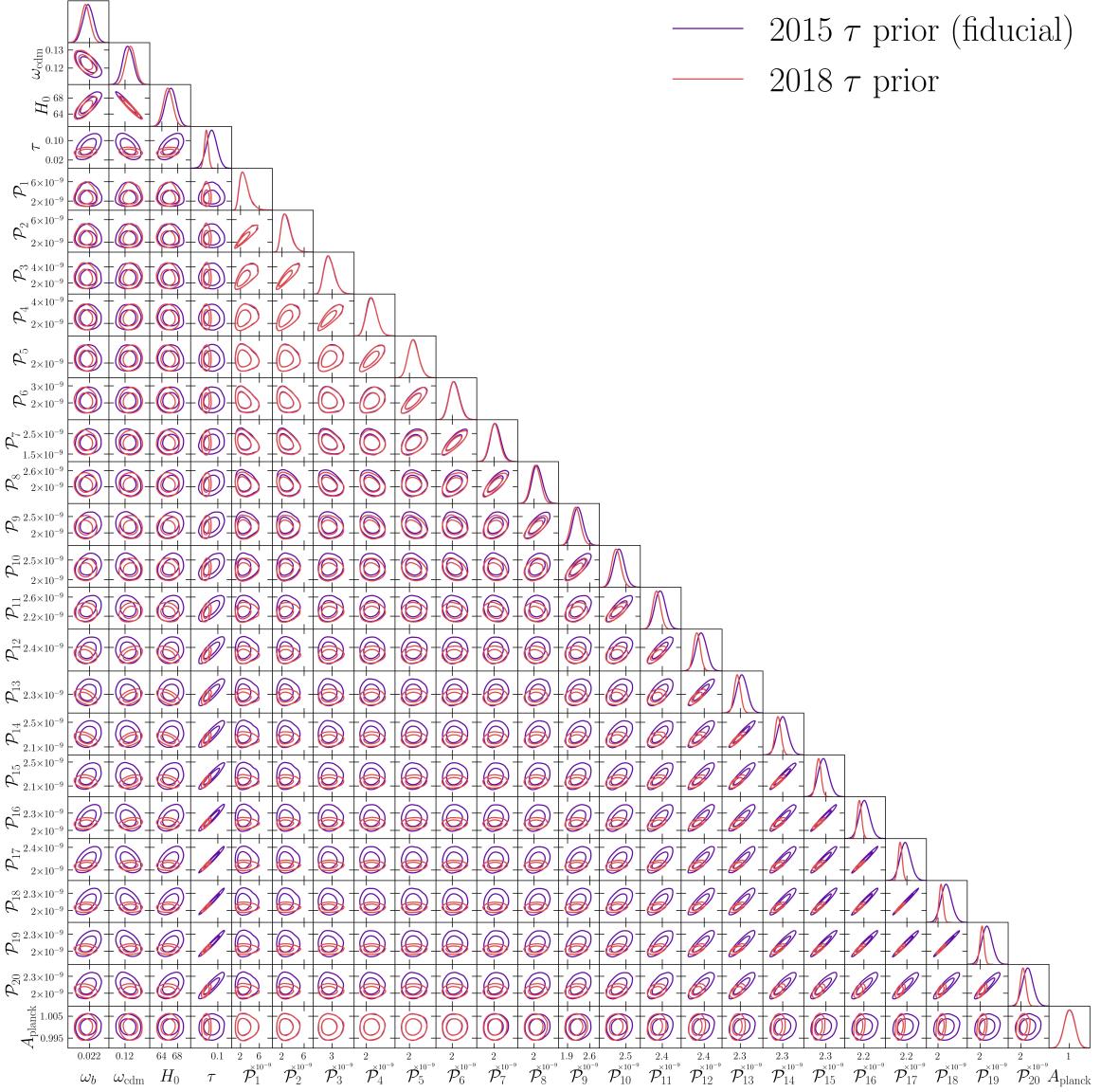
Figure 15: The joint constraints on 20 nodes of the GP used to generate the PPS values at each k mode, the cosmological parameters and the nuisance parameter using different τ priors. The contours represent the 68% and 95% confidence regions, accompanied by 1-D marginal distributions for each parameter.

considered, we obtain consistent results, demonstrating that constraints are robust and not artifacts of specific hyperparameter choices.

We also test the robustness of our NUTS-based results by varying the `target_accept_prob` hyperparameter, which controls the target acceptance probability for step size adaptation. Increasing this parameter leads to a smaller step size, making the sampling process slower but more robust by reducing the likelihood of divergences.

The default value for `target_accept_prob` is 0.8. We incrementally increase this value in our tests and observed no significant changes in the results, indicating that our analysis is

robust to this hyperparameter choice and unaffected by divergences. We specifically focused on this hyperparameter because divergences in HMC can signal poor exploration of the posterior distribution. The consistency of our results across different `target_accept_prob` values reassures us that our sampling accurately captures the posterior without being influenced by numerical artifacts or parameter tuning.

C.3   Details of sampler specific hyperparameters for fiducial results

The joint constraints in Fig. 5 result from two different sampling approaches: the primary sampling algorithm, NUTS, within the differentiable pipeline, and with the non-differentiable framework using INS.

For the NUTS sampler, we perform the analysis on a high-performance computing system, utilizing four NVIDIA A40 GPUs. The MCMC setup with a NUTS kernel is configured with 1000 warm-up iterations, 4000 samples, and four chains, initializing the parameters to the median of the prior distribution. We did not specify any value for the `inverse_mass_matrix` and thus it was initialised to the identity matrix by default. This configuration allows us to complete the sampling process in approximately 2 hours.

In comparison, we conduct the same analysis using the non-differentiable INS framework. Here, we configure the sampler with 3000 live points (n_{live}) and a pool of 48 CPU cores. The sampler runs with 8000 effective samples (N_{eff}), discarding exploration to improve convergence. This setup takes significantly longer, approximately 1 day and 17 hours.

By comparing the results, we highlight the efficiency of the differentiable NUTS pipeline, which benefits from GPU acceleration and gradient-based optimization methods, in contrast to the more computationally intensive INS approach that relies solely on CPU cores.

Acknowledgments

References

[1] A.H. Guth, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, *Phys. Rev. D* **23** (1981) 347.

[2] A. Linde, *A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems*, *Physics Letters B* **108** (1982) 389.

[3] A. Starobinsky, *A new type of isotropic cosmological models without singularity*, *Physics Letters B* **91** (1980) 99.

[4] Planck Collaboration, N. Aghanim, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont et al., *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *A&A* **641** (2020) A1 [1807.06205].

[5] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. V. CMB power spectra and likelihoods*, *A&A* **641** (2020) A5 [1907.12875].

[6] A.A. Starobinskij, *Spectrum of adiabatic perturbations in the universe when there are singularities in the inflationary potential.*, *Soviet Journal of Experimental and Theoretical Physics Letters* **55** (1992) 489.

[7] J. Adams, B. Cresswell and R. Easther, *Inflationary perturbations from a potential with a step*, *Phys. Rev. D* **64** (2001) 123514 [`astro-ph/0102236`].

[8] M. Joy, V. Sahni and A.A. Starobinsky, *New universal local feature in the inflationary perturbation spectrum*, Phys. Rev. D **77** (2008) 023514 [`0711.1585`].

[9] M. Joy, A. Shafieloo, V. Sahni and A.A. Starobinsky, *Is a step in the primordial spectral index favoured by CMB data?*, J. Cosmology Astropart. Phys. **2009** (2009) 028 [`0807.3334`].

[10] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, A&A **641** (2020) A6 [`1807.06209`].

[11] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. X. Constraints on inflation*, A&A **641** (2020) A10 [`1807.06211`].

[12] S.L. Bridle, A.M. Lewis, J. Weller and G. Efstathiou, *Reconstructing the primordial power spectrum*, MNRAS **342** (2003) L72 [`astro-ph/0302306`].

[13] G. Ballesteros, J.A. Casas and J.R. Espinosa, *The running spectral index as a probe of physics at high scales*, J. Cosmology Astropart. Phys. **2006** (2006) 001 [`hep-ph/0601134`].

[14] A. Kosowsky and M.S. Turner, *CBR anisotropy and the running of the scalar spectral index*, Phys. Rev. D **52** (1995) R1739 [`astro-ph/9504071`].

[15] M. Zarei, *On the running of the spectral index to all orders: a new model-dependent approach to constrain inflationary models*, *Classical and Quantum Gravity* **33** (2016) 115008 [`1408.6467`].

[16] R. Easther and H.V. Peiris, *Implications of a running spectral index for slow roll inflation*, J. Cosmology Astropart. Phys. **2006** (2006) 010 [`astro-ph/0604214`].

[17] P.D. Meerburg, R.A.M.J. Wijers and J.P. van der Schaar, *WMAP7 constraints on oscillations in the primordial power spectrum*, MNRAS **421** (2012) 369 [`1109.5264`].

[18] P.D. Meerburg, D.N. Spergel and B.D. Wandelt, *Searching for oscillations in the primordial power spectrum. I. Perturbative approach*, Phys. Rev. D **89** (2014) 063536 [`1308.3704`].

[19] P.D. Meerburg and D.N. Spergel, *Searching for oscillations in the primordial power spectrum. II. Constraints from Planck data*, *Phys. Rev. D* **89** (2014) 063537 [`1308.3705`].

[20] Y. Wang and G.J. Mathews, *Model-independent Primordial Power Spectrum from MAXIMA, BOOMERANG, and DASI Data*, ApJ **573** (2002) 1 [`astro-ph/0011351`].

[21] S. Hannestad, *Reconstructing the primordial power spectrum—a new algorithm*, J. Cosmology Astropart. Phys. **2004** (2004) 002 [`astro-ph/0311491`].

[22] R. Hlozek, J. Dunkley, G. Addison, J.W. Appel, J.R. Bond, C. Sofia Carvalho et al., *The Atacama Cosmology Telescope: A Measurement of the Primordial Power Spectrum*, ApJ **749** (2012) 90 [`1105.4887`].

[23] P. Mukherjee and Y. Wang, *Direct Wavelet Expansion of the Primordial Power Spectrum*, ApJ **598** (2003) 779 [`astro-ph/0301562`].

[24] P. Mukherjee and Y. Wang, *Model-independent Reconstruction of the Primordial Power Spectrum from Wilkinson Microwave Anistropy Probe Data*, ApJ **599** (2003) 1 [`astro-ph/0303211`].

[25] C. Sealfon, L. Verde and R. Jimenez, *Smoothing spline primordial power spectrum reconstruction*, Phys. Rev. D **72** (2005) 103520 [`astro-ph/0506707`].

[26] L. Verde and H. Peiris, *On minimally parametric primordial power spectrum reconstruction and the evidence for a red tilt*, J. Cosmology Astropart. Phys. **2008** (2008) 009 [`0802.1219`].

[27] A. Ravenni, L. Verde and A.J. Cuesta, *Red, Straight, no bends: primordial power spectrum reconstruction from CMB and large-scale structure*, J. Cosmology Astropart. Phys. **2016** (2016) 028 [`1605.06637`].

[28] M. Matsumiya, M. Sasaki and J. Yokoyama, *Cosmic inversion: Reconstructing the primordial spectrum from CMB anisotropy*, Phys. Rev. D **65** (2002) 083007 [`astro-ph/0111549`].

[29] N. Kogo, M. Matsumiya, M. Sasaki and J. Yokoyama, *Reconstructing the Primordial Spectrum from WMAP Data by the Cosmic Inversion Method*, ApJ **607** (2004) 32 [`astro-ph/0309662`].

[30] F. Lanusse, P. Paykari, J.L. Starck, F. Sureau, J. Bobin and A. Rassat, *PRISM: Recovery of the primordial spectrum from Planck data*, A&A **571** (2014) L1 [`1410.2571`].

[31] J.A. Vázquez, M. Bridges, M.P. Hobson and A.N. Lasenby, *Model selection applied to reconstruction of the Primordial Power Spectrum*, J. Cosmology Astropart. Phys. **2012** (2012) 006 [`1203.1252`].

[32] C. Gauthier and M. Bucher, *Reconstructing the primordial power spectrum from the CMB*, J. Cosmology Astropart. Phys. **2012** (2012) 050 [`1209.2147`].

[33] W.J. Handley, A.N. Lasenby, H.V. Peiris and M.P. Hobson, *Bayesian inflationary reconstructions from Planck 2018 data*, Phys. Rev. D **100** (2019) 103511 [`1908.00906`].

[34] G. Martínez-Somonte, A. Marcos-Caballero, E. Martínez-González and G. Cañas-Herrera, *Bayesian inference methodology for primordial power spectrum reconstructions from Large Scale Structure*, J. Cosmology Astropart. Phys. **2024** (2024) 072 [`2306.16866`].

[35] A. Shafieloo and T. Souradeep, *Primordial power spectrum from WMAP*, Phys. Rev. D **70** (2004) 043523 [`astro-ph/0312174`].

[36] G. Nicholson and C.R. Contaldi, *Reconstruction of the primordial power spectrum using temperature and polarisation data from multiple experiments*, J. Cosmology Astropart. Phys. **2009** (2009) 011 [`0903.1106`].

[37] R.S. Chandra and T. Souradeep, *Primordial Power Spectrum reconstruction from CMB Weak Lensing Power Spectrum*, J. Cosmology Astropart. Phys. **2021** (2021) 081 [`2104.12253`].

[38] W. Sohn, A. Shafieloo and D.K. Hazra, *Deblurring the early Universe: reconstruction of primordial power spectrum from Planck CMB using image analysis techniques*, J. Cosmology Astropart. Phys. **2024** (2024) 056 [`2211.15139`].

[39] C.K. Williams and C.E. Rasmussen, *Gaussian processes for machine learning*, vol. 2, MIT press Cambridge, MA (2006).

[40] G. Cañas-Herrera, J. Torrado and A. Achúcarro, *Bayesian reconstruction of the inflaton's speed of sound using CMB data*, Phys. Rev. D **103** (2021) 123531 [`2012.04640`].

[41] A. Shafieloo, A.G. Kim and E.V. Linder, *Gaussian process cosmography*, Phys. Rev. D **85** (2012) 123530 [`1204.2272`].

[42] J. Ruiz-Zapatero, D. Alonso, C. García-García, A. Nicola, A. Mootoovaloo, J.M. Sullivan et al., *LimberJack.jl: auto-differentiable methods for angular power spectra analyses*, The Open Journal of Astrophysics **7** (2024) 11 [`2310.08306`].

[43] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, Physics Letters B **195** (1987) 216.

[44] R. Neal, *MCMC Using Hamiltonian Dynamics* (2011), 10.1201/b10905.

[45] S. Dodelson and F. Schmidt, *Modern Cosmology* (2020), 10.1016/C2017-0-01943-2.

[46] H.V. Peiris and L. Verde, *The shape of the primordial power spectrum: A last stand before Planck data*, Phys. Rev. D **81** (2010) 021302 [`0912.0268`].

[47] J. Ruiz-Zapatero, C. García-García, D. Alonso, P.G. Ferreira and R.D.P. Grumitt, *Model-independent constraints on Ω_m and $H(z)$ from the link between geometry and growth*, MNRAS **512** (2022) 1967 [`2201.07025`].

[48] J. Ruiz-Zapatero, D. Alonso, P.G. Ferreira and C. Garcia-Garcia, *Impact of the Universe's expansion rate on constraints on modified growth of structure*, Phys. Rev. D **106** (2022) 083523 [2207.09896].

[49] S.-g. Hwang, B. L'Huillier, R.E. Keeley, M.J. Jee and A. Shafieloo, *How to use GP: effects of the mean function and hyperparameter selection on Gaussian process regression*, J. Cosmology Astropart. Phys. **2023** (2023) 014 [2206.15081].

[50] R. Jimenez, L. Verde, H. Peiris and A. Kosowsky, *Fast cosmological parameter estimation from microwave background temperature and polarization power spectra*, *Phys. Rev. D* **70** (2004) 023005 [astro-ph/0404237].

[51] T. Auld, M. Bridges and M.P. Hobson, *COSMONET: fast cosmological parameter estimation in non-flat models using neural networks*, MNRAS **387** (2008) 1575 [astro-ph/0703445].

[52] K. Heitmann, E. Lawrence, J. Kwan, S. Habib and D. Higdon, *The Coyote Universe Extended: Precision Emulation of the Matter Power Spectrum*, ApJ **780** (2014) 111 [1304.7849].

[53] A. Mootoovaloo, A.F. Heavens, A.H. Jaffe and F. Leclercq, *Parameter inference for weak lensing using Gaussian Processes and MOPED*, MNRAS **497** (2020) 2213 [2005.06551].

[54] A. Manrique-Yus and E. Sellentin, *Euclid-era cosmology for everyone: neural net assisted MCMC sampling for the joint 3 × 2 likelihood*, MNRAS **491** (2020) 2655 [1907.05881].

[55] G. Aricò, R.E. Angulo and M. Zennaro, *Accelerating Large-Scale-Structure data analyses by emulating Boltzmann solvers and Lagrangian Perturbation Theory*, *arXiv e-prints* (2021) arXiv:2104.14568 [2104.14568].

[56] A. Spurio Mancini, D. Piras, J. Alsing, B. Joachimi and M.P. Hobson, *COSMOPOWER: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys*, MNRAS **511** (2022) 1771 [2106.03846].

[57] D. Piras and A. Spurio Mancini, *CosmoPower-JAX: high-dimensional Bayesian inference with differentiable cosmological emulators*, *The Open Journal of Astrophysics* **6** (2023) 20 [2305.06347].

[58] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, ApJ **538** (2000) 473 [astro-ph/9911177].

[59] P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso et al., *The Simons Observatory: science goals and forecasts*, J. Cosmology Astropart. Phys. **2019** (2019) 056 [1808.07445].

[60] H. Prince and J. Dunkley, *Data compression in cosmology: A compressed likelihood for Planck data*, Phys. Rev. D **100** (2019) 083502 [1909.05869].

[61] F. Feroz, M.P. Hobson, E. Cameron and A.N. Pettitt, *Importance Nested Sampling and the MultiNest Algorithm*, *The Open Journal of Astrophysics* **2** (2019) 10 [1306.2144].

[62] J.U. Lange, *NAUTILUS: boosting Bayesian importance nested sampling with deep learning*, MNRAS **525** (2023) 3181 [2306.16923].

[63] J. Skilling, *Nested Sampling*, vol. 735 of *American Institute of Physics Conference Series*, AIP (Nov., 2004), 10.1063/1.1835238.

[64] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, PASP **125** (2013) 306 [1202.3665].

[65] M. Betancourt, *A Conceptual Introduction to Hamiltonian Monte Carlo*, *arXiv e-prints* (2017) arXiv:1701.02434 [1701.02434].

[66] M.D. Hoffman and A. Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, *arXiv e-prints* (2011) arXiv:1111.4246 [1111.4246].

[67] J. Bradbury, R. Frostig, P. Hawkins, M.J. Johnson, C. Leary, D. Maclaurin et al., *JAX: composable transformations of Python+NumPy programs*, 2018.

[68] R. Frostig, M.J. Johnson and C. Leary, *Compiling machine learning programs via high-level tracing*, Systems for Machine Learning **4** (2018) .

[69] E. Bingham, J.P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos et al., *Pyro: Deep universal probabilistic programming*, J. Mach. Learn. Res. **20** (2019) 28:1.

[70] D. Phan, N. Pradhan and M. Jankowiak, *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro*, arXiv e-prints (2019) arXiv:1912.11554 [1912.11554].

[71] Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters*, A&A **594** (2016) A11 [1507.02704].

[72] C.L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon et al., *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, ApJS **148** (2003) 1 [astro-ph/0302207].

[73] Planck Collaboration, P.A.R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown et al., *Planck 2015 results. XX. Constraints on inflation*, A&A **594** (2016) A20 [1502.02114].

[74] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. III. High Frequency Instrument data processing and frequency maps*, A&A **641** (2020) A3 [1807.06207].

[75] C. Preston, A. Amon and G. Efstathiou, *Reconstructing the matter power spectrum with future cosmic shear surveys*, Mon. Not. Roy. Astron. Soc. **533** (2024) 621 [2404.18240].

[76] J.C. Broxterman and K. Kuijken, *First step toward matter power spectrum reconstruction with Stage III weak gravitational lensing surveys*, Astron. Astrophys. **692** (2024) A201 [2409.13404].

[77] G. Ye, J.-Q. Jiang and A. Silvestri, *A model-independent reconstruction of the matter power spectrum*, 2411.07082.

[78] P. Simon, L. Porth, P. Burger and K. Kuijken, *Kids-1000: Detection of deviations from a purely cold dark-matter power spectrum with tomographic weak gravitational lensing*, 2025.

[79] D. Piras, A. Polanska, A.S. Mancini, M.A. Price and J.D. McEwen, *The future of cosmological likelihood-based inference: accelerated high-dimensional parameter estimation and model comparison*, The Open Journal of Astrophysics **7** (2024) 73 [2405.12965].