

unimpeded is a re-usable library of posterior samples, nested sampling [1, 2] runs and machine learning emulators across a grid of cosmological models for detecting cosmological tensions between datasets from the DiRAC allocation (DP192 & 264). It serves as an analogous grid to the Planck Legacy Archive (PLA), but machine learning enhanced and expanded to enable not only parameter estimation (currently available with the MCMC chains on PLA), but also allowing cosmological model comparison and tension quantification. The emulators are implemented with piecewise normalising flows [3] as part of the package *margarine* [4, 5], though alternative density estimation methods can be used. The combination of nested sampling and density estimation allows us to obtain the same posterior distributions as one would have found from a full nested sampling run over all nuisance parameters, but many orders of magnitude faster. This allows users to use the existing results of cosmological analyses without the need to re-run on supercomputers.

The three pillars of Bayesian inference

Parameter Estimation

What do the data tell us about the parameters of a model? e.g. the size or age of a Λ CDM universe

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}.$$

Model Comparison

How much do the data support a given model? e.g. Λ CDM vs a dynamic dark energy cosmology

$$\text{Posterior} = \frac{\text{Evidence} \times \text{Prior}}{\text{Normalisation}}.$$

Tension quantification

Do different datasets make consistent predictions from the same model?
e.g. Hubble H_0 tension from CMB vs Type 1A supernovae data

$$\mathcal{R} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A \mathcal{Z}_B},$$

Model comparison and tension quantification have become increasingly relevant in recently years because of anomalies, e.g. H_0 and σ_8 tension. Although MCMC chains (currently available on PLA grids) may be used for parameter estimation, model comparison and tension quantification are far more computationally expensive, so more specialist tools like nested sampling are required.

Available Models & Data

unimpeded provides a systematic coverage of cosmological models, datasets and their pairwise comparisons. It will be expanded as new datasets and models become available, e.g. ACT, SPT, DESI, DESY5, Union, Patheon+,.... Requests for new datasets are welcome!

Cosmological datasets

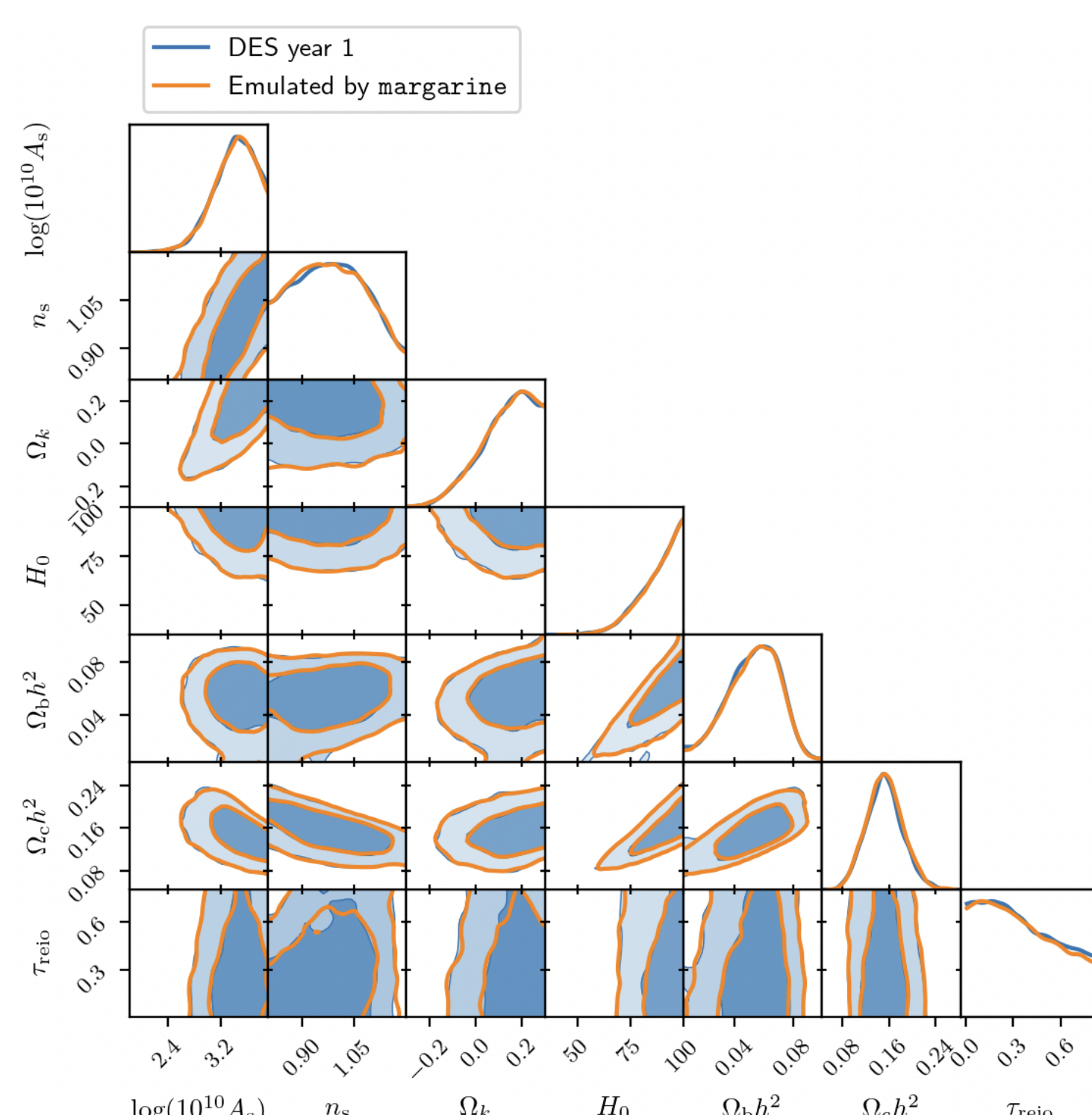
- CMB: (Plik, Camspec, NPIPE, BICEP) \pm CMB lensing
- BAO: SDSS, BOSS, eBOSS, Ly α
- SNe: Pantheon, SH0ES
- WL: DESY1

Cosmological models

- Λ CDM : $H_0, \tau_{\text{reio}}, \Omega_b h^2, \Omega_c h^2, A_s, n_s$
- $K\Lambda$ CDM : Λ CDM + Ω_K (varying curvature)
- $N\Lambda$ CDM : Varying N_{eff} and total mass of 3 degenerate ν 's
- $n\Lambda$ CDM : Varying total mass of 3 degenerate ν 's with $N_{\text{eff}}=3.044$
- $m\Lambda$ CDM : Varying N_{eff} with two massless ν and one with $m=0.06$
- $n_{\text{run}}\Lambda$ CDM : Λ CDM + n_{run} (running of spectral index $dn_s/d\ln k$)
- w CDM : Λ CDM + w (constant cosmological equation of state)
- $w_0 w_a \Lambda$ CDM : Λ CDM + $w_0 + w_a$ (varying dark energy equation of state, CLP)
- $r\Lambda$ CDM : Λ CDM + r (varying scalar-to-tensor ratio)

Results - Emulators

unimpeded provides a library of trained bijectors to be used as priors or emulators [6] or nuisance marginalised likelihoods [5]. Piecewise normalising flows are used with margarine to model complex probability densities through bijective transformations between a base distribution and the target distribution. Density estimators, such as Kernel Density Estimators [7, 8] and Masked Autoregressive Flows [9] are used to rapidly calculate reliable and reusable marginal probability densities and marginal Bayesian summary statistics for key signal or cosmological parameters.



This enables the access to the nuisance-free likelihood functions, which greatly reduces computational cost in combining parameters constraints from different datasets. It allows users to use a real ‘planck prior’ rather than a Gaussian approximation for future cosmological analyses. Here is an example of the DES year 1 dataset (blue filled contours) emulated by margarine (orange line) using nested sampling chains, with a p-value of 0.752.

Using unimpeded

unimpeded is a Python tool for seamlessly downloading and cacheing chains (API in 'alpha'). It provides both MCMC and nested sampling chains. Data are stored on zenodo, with hdf5 storage for fast and reliable download and storage. unimpeded is a pip-installable package.

```
pip install unimpeded
```

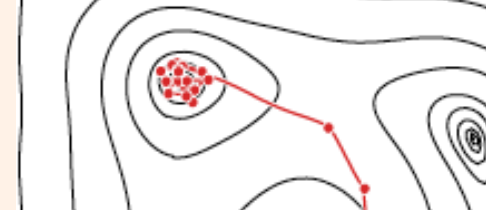
To access any combination of these models and datasets (also the pairwise combinations of datasets, e.g. Planck 2018 CamSpec+DES year 1), simply call the function `unimpeded.get(data, model, method)` and specify the dataset, model and sampling method (`ns` = nested sampling or `mcmc` = Metropolis–Hastings MCMC methods).

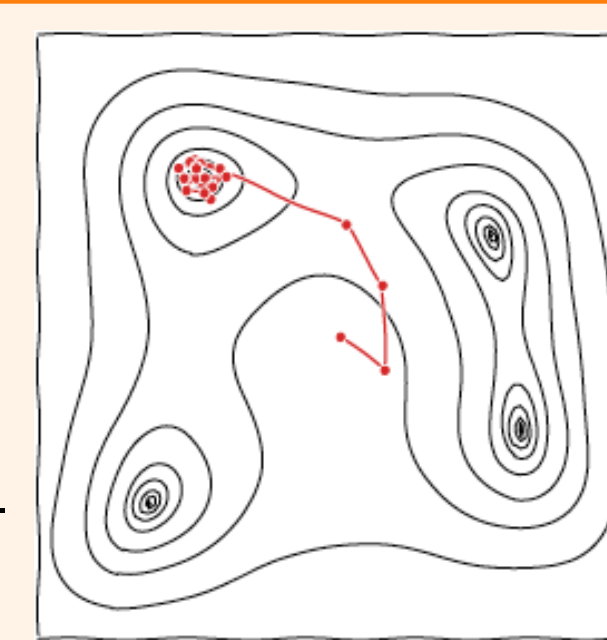
```
samples = unimpeded.get(data='planck_2018_CamSpec', model='lcdm',
method='ns')
```

Samplers


Samples are the fundamental building block of numerical Bayesian inference, encapsulating high-dimensional posterior probability distributions. The samplers used in `unimpeded` are Metropolis Hastings MCMC and Nested Sampling, using Cobaya [10].

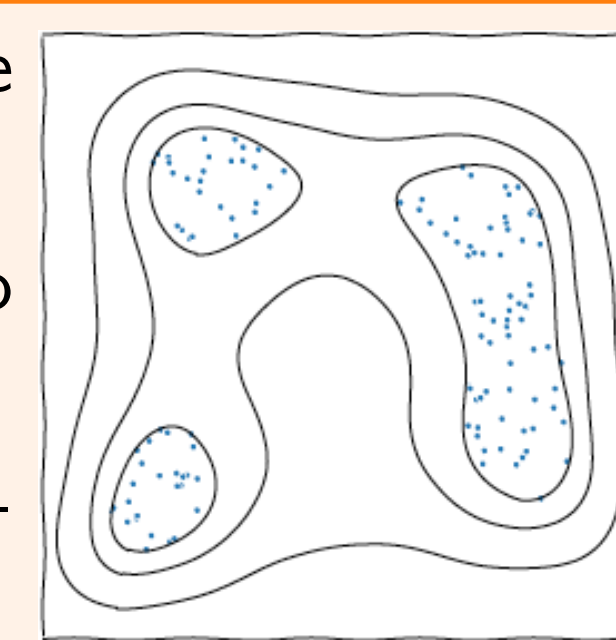
Metropolis–Hastings MCMC

- Single “walker”
 - Explores posterior
 - Fast, if proposal matrix is tuned
 - Parameter estimation, suspiciousness calculation
 - Channel capacity optimised for generating posterior samples
- 



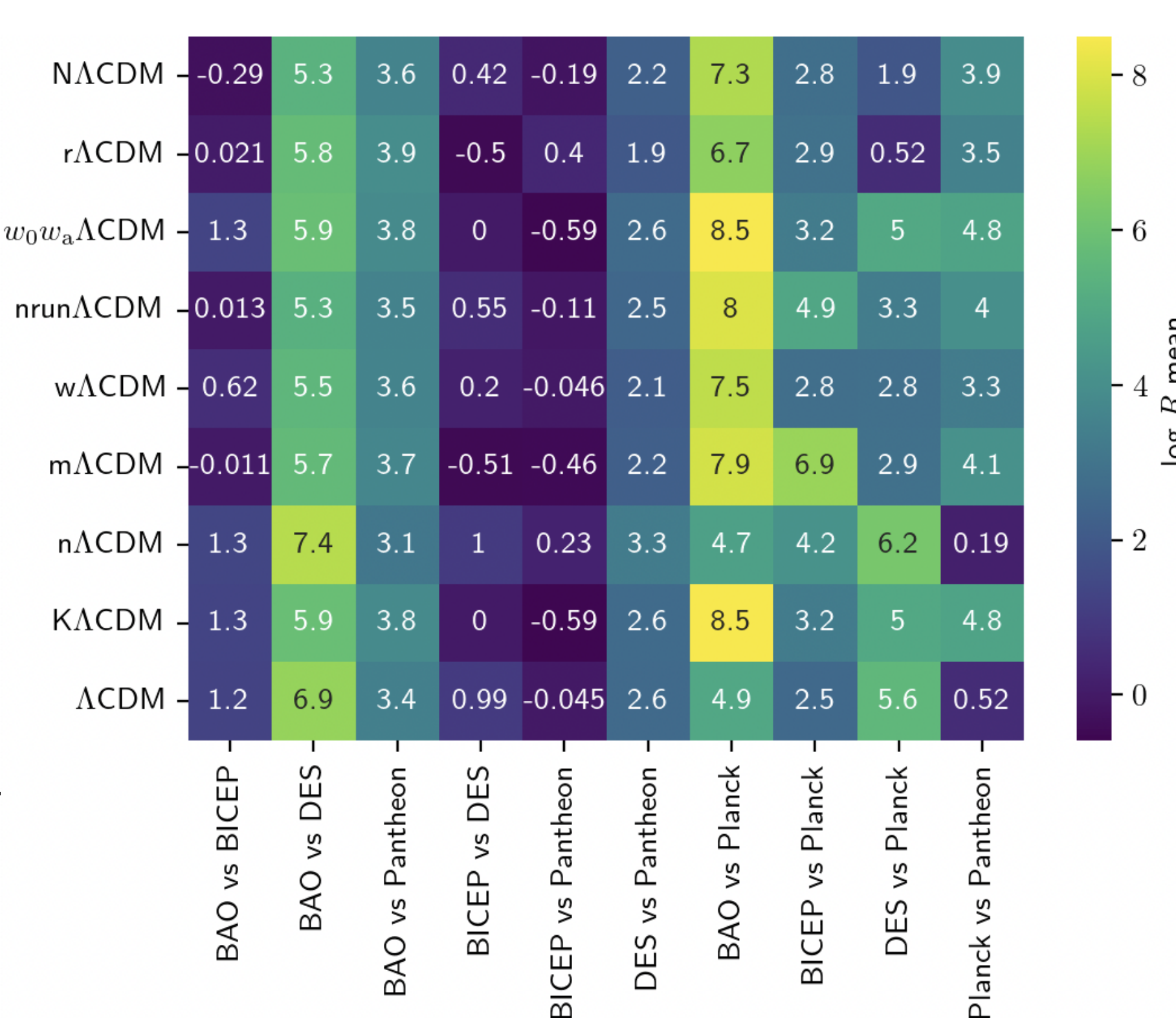
Nested Sampling

- Ensemble of “live points”
 - Scans from prior to peak of likelihood
 - Slower, no tuning required
 - Parameter estimation, model comparison, tension quantification
 - Channel capacity optimised for computing partition function
- 



Preliminary Results - Tension Statistics

unimpeded acts as a tool for convenient and systematic measurement of tension statistics, and quantifying the Bayesian degree of confidence in comparing and combining datasets across different models. Available tension statistics include the p -value, R statistics [11], the suspiciousness [12], the information ratio from the Kullback-Leibler divergences and the Bayesian model dimensionality. Here is an example of one of the tension statistics ($\log R$) compared between 10 pairwise datasets across 9 cosmological models.



References

Github repository
for unimpeded
(top) and this
poster (bottom).

This work was performed using the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. DiRAC is part of the National e-Infrastructure.

- [1] John Skilling, Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006.
- [2] George Box and A. E. Jenkins, Time series analysis: concepts and techniques. John Wiley & Sons, 1976.
- [3] Andrew Gelman, John Hill, James Buchner, John Hill, and Matthew Griffiths, Will Handley, Michael Beach, Edward Hagger, Michael Hobson, Anthony Lessley, David Parkinson, Livia B. Pártay, Matthew Pitkin, Doris Schneider, Joshua S. Speagle, Leah Stott, John Veitch, Philipp Wagner, David J. Wales, and David Yallup, Nested sampling for physical scientists. *Nature Reviews Physics*, 2:39, May 2022.
- [4] Will Handley, Will Handley, and Pablo Lemos, Nested sampling for cosmological parameters. *arXiv preprint arXiv:2304.05029*, May 2023.
- [5] Harry T. J. Beavis, William J. Handley, Pablo Lemos, Peter Sims, Eloy de La Rosa Accoto, Anastasia Fialkov, and Justin Alsing, Removing the tail in your posterior samples with margarine. *arXiv e-prints*, page arXiv:2205.12841, May 2022.
- [6] Harry Beavis, Will Handley, Pablo Lemos, Peter Sims, Eloy de La Rosa, and Anastasia Fialkov, Marginal Bayesian Statistics Using Censored Exponential Flows and Kernel Density Estimators with Examples from Cosmology. *arXiv preprint arXiv:2304.05029*, May 2023.
- [7] Justin Alsing and Will Handley, Nested sampling with any prior you like. *MNRAS*, 505(1):L05–L9, 2021.
- [8] Emanuel Parzen, On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.
- [9] George Box and A. E. Jenkins, Time series analysis: concepts and techniques. The Annals of Statistics, 27(3):837–856, 1999.
- [10] George Papamakarios, Theo Pavlakou, and Iain Murray, Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, page arXiv:1705.07057, May 2017.
- [11] Jesús Torrado and Antony Lewis, Query costs for Bayesian analysis of hierarchical physical models. *JCAP*, 2021(5):057, May 2021.
- [12] Will Handley and Pablo Lemos, Cosmological parameters: Nested sampling as a tool for comparing datasets. *Physical Review D*, 103(12):123501, June 2021.
- [13] Will Handley and Pablo Lemos, Quantifying tensions in cosmological parameters: Interpreting the DES detection rate. *PRD*, 100(4):043504, August 2019.

