

Accelerating Cosmological Inference: A Re-usable Library of JAX-based Machine Learning Emulators for Next-Generation Surveys

Dily Duan Yi Ong

Kavli Institute for Cosmology
Institute of Astronomy

University of Cambridge

26 August 2026



UNIVERSITY OF
CAMBRIDGE



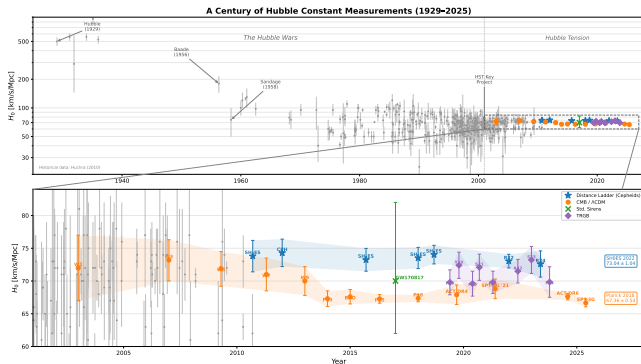
DiRAC
High Performance
Computing Facility

EuCAIFCon 2026
Inference & Uncertainty

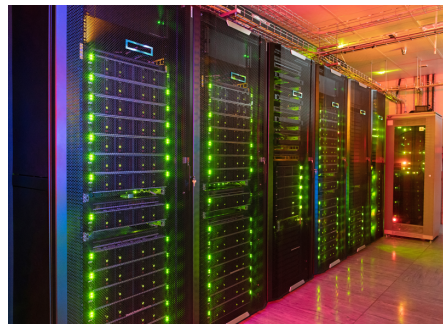
- ▶ Dily Ong
- ▶ Final-year PhD Student at the University of Cambridge
- ▶ Dr. Will Handley's research group
- ▶ Research focuses on Bayesian inference in cosmological model selection and tension quantification
- ▶ Author of the Python package `unimpeded`
- ▶ Principal Investigator (PI) of 2 UKRI-funded DiRAC HPC projects:
 - ▶ JAX-based machine-learning emulators on GPU
 - ▶ Mapping the prior-sensitivity of cosmological Bayesian evidence



- ▶ Precision cosmology $\Rightarrow$ more datasets $\Rightarrow$ more anomalies
- ▶ Major tensions:
 - ▶ H_0 Hubble tension (CMB vs SNe)
 - ▶ σ_8 tension (CMB vs weak lensing)
 - ▶ Ω_K tension (CMB vs lensing vs BAO)
- ▶ **Tension quantification:** do independent datasets, predicted by model to agree, actually agree?
- ▶ **Modern telescopes** bring unprecedented precision $\rightarrow$ joint constraints with legacy data
- ▶ Need robust, systematic comparison across many datasets and models



- ▶ **The problem of scale**
 - ▶ **Tens of instrumental nuisance parameters** per likelihood: calibration, foregrounds, systematics
 - ▶ DES Year 1: **20 of 26** parameters are nuisance; only **6** are cosmology
 - ▶ A joint analysis re-samples *every* dataset's systematics, every time
- ▶ **The problem of cost**
 - ▶ Model comparison & tension quantification need the Bayesian evidence, **orders of magnitude** more computationally expensive than parameter estimation
 - ▶ **HPC resources are limited**
- ▶ **Computationally prohibitive** at next-generation survey volumes



Goal: pay the nuisance cost *once*, then work only in the cosmological subspace.

- ▶ A re-usable library of posterior samples and Nested Sampling chains
- ▶ Systematic coverage of 8 cosmological models and 13 cosmological surveys + combinations
- ▶ Built-in model comparison and tension quantification tools
- ▶ Downloadable posterior samples (MCMC or Nested Sampling)
- ▶ Powered by DiRAC, hosted on Zenodo, pip-installable, available on GitHub
- ▶ **Turns weeks or months of supercomputer time into seconds on your laptop**

Released in papers [2511.04661] [2511.05470]

unimpeded: Universal model comparison & parameter estimation distributed over every dataset

unimpeded:	Universal model comparison & parameter estimation distributed over every dataset
Author:	Dily Ong & Will Handley
Version:	1.0.1
Homepage:	https://github.com/handley-lab/unimpeded
Documentation:	http://unimpeded.readthedocs.io/



- ▶ 8 cosmological models + 13 cosmological surveys
- ▶ Pairwise and triplet combinations
- ▶ Precomputed Cobaya runs with CAMB
- ▶ Supernovae (SNe): Pantheon+, Union3, DES-SN5YR, DES-Dovekie
Other probes: DESI BAO, Planck CMB, DES-Y1 WL

Baseline: Λ CDM

Varying curvature: $\Omega_K \Lambda$ CDM

Constant DE EoS: w CDM

Time-varying DE: $w_0 w_a$ CDM

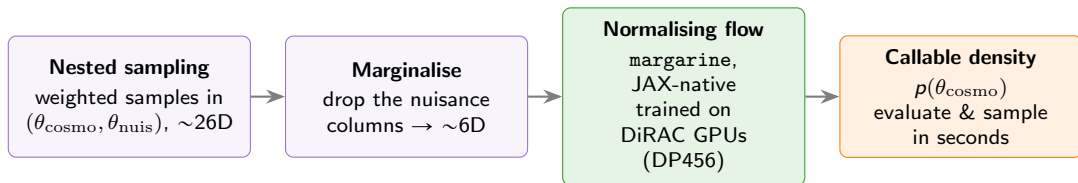
Varying lensing: $A_L \Lambda$ CDM

Neutrino masses: $m_\nu \Lambda$ CDM

Running spectral: $n_{\text{run}} \Lambda$ CDM

Primordial GWs: $r \Lambda$ CDM

Dataset	Likelihood
Cosmic Microwave Background	
Planck [23, 31]	planck_2018_lowl.TT planck_2018_lowl.EE planck_2018_highl.plik.TTTEEE planck_2018_highl.plik.SZ
Planck with CMB lensing [23, 25, 31]	planck_2018_lowl.TT planck_2018_lowl.EE planck_2018_highl.plik.TTTEEE planck_2018_highl.plik.SZ planck_2018_lensing.clik
CamSpec [24]	planck_2018_lowl.TT planck_2018_lowl.EE planck_2018_highl.CamSpec2021.TTTEEE
CamSpec with CMB lensing [24, 25]	planck_2018_lowl.TT planck_2018_lowl.EE planck_2018_highl.CamSpec2021.TTTEEE planck_2018_lensing.clik planck_2018_lensing.clik
CMB Lensing [25]	
Baryon Acoustic Oscillations	
DESI DR1 [26]	bao.desi_2024.bao.all
DESI DR2 [1]	bao.desi_dr2
Type Ia Supernovae	
Pantheon+ [27]	sn.pantheonplus
Union3 [28]	sn.union3
DES-Dovekie [5, 6]	sn.desdovekie
DES-SN5YR [29]	sn.desy5
Weak Lensing	
DES Y1 [30]	des.y1.joint



- ▶ Learn the **marginal posterior directly from the nested sampling chains**, with no new likelihood calls
- ▶ Bijective flow $\Rightarrow$ exact, normalised, differentiable density with a tractable Jacobian
- ▶ margarine [2205.12841] [2207.11457]: JAX-native density estimation

Why this works

- ▶ Orders of magnitude decreased to θ_{cosmo} by marginalising out θ_{nuis}
- ▶ Nested sampling maps out features that traditional MCMC chains cannot

Ultra-fast “nuisance-free likelihood”

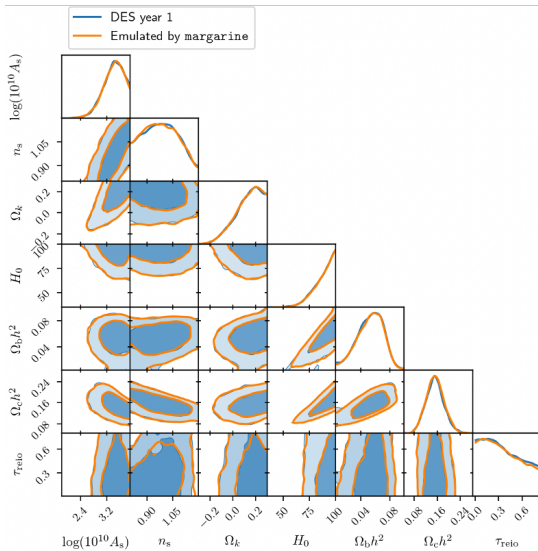
- ▶ Re-use a survey’s constraint in a new joint analysis **without re-sampling its systematics**
- ▶ Calibration and foreground parameters never appear again
- ▶ Multi-survey joint analyses collapse to the **cosmological subspace** ($\sim 6D$, not $\sim 100D$)

Informative true prior

- ▶ A *real* “Planck prior”: the actual posterior shape, **not a Gaussian approximation**
- ▶ Captures curvature, skew and degeneracy directions that a covariance matrix throws away
- ▶ Portable: download the flow, use it as a prior anywhere

Same object, two uses, and a natural bridge to **simulation-based inference**.

- ▶ DES Year 1, $\Omega_K \Lambda$ CDM: full nuisance-marginalised nested sampling run (**blue**) vs the margarine flow (**orange**)
- ▶ Contours lie **on top of each other**, including the non-Gaussian $\Omega_K - H_0$ degeneracy
- ▶ The **exact posterior** of a full run, many orders of magnitude faster
- ▶ 20 nuisance parameters marginalised away, permanently



- ▶ A re-usable library of **JAX-based normalising-flow emulators**, trained on a grid of nested sampling chains
- ▶ Ultra-fast **nuisance-free likelihoods** and **informative true priors**
- ▶ Exact posteriors of a full nuisance-marginalised run, many orders of magnitude faster
- ▶ Collapses multi-survey joint analyses to the **cosmological subspace**
- ▶ Versatile, efficient and **equitable**: rigorous inference without an HPC allocation
- ▶ Expanding to Euclid, Rubin/LSST, ACT, SPT; seeking collaborators & α -testers
- ▶ Final-year PhD, on the postdoc market; happy to chat

Thank you!



```
pip install unimpeded
```

[[2511.04661](#)] [[2511.05470](#)] [[2205.12841](#)]

A library is only re-usable if it trains *reliably* across every model–dataset cell, so the hyperparameters were swept systematically, not hand-tuned.

Network architecture

Number of flow layers, hidden units per layer, and the conditioner width, traded off against the dimensionality of each model's cosmological subspace.

Learning schedule

Learning rate, decay schedule, batch size and early stopping, chosen so that sparse, unevenly weighted nested sampling samples do not overfit.

Activation functions

Compared across the grid for stability and tail behaviour; the tails matter, because these densities are used as priors.

- ▶ Trained on DiRAC GPUs under **DP456** (10k GPU hours, as PI)
- ▶ Validated against the parent chain: KL divergence and contour-level agreement, per cell

Outcome

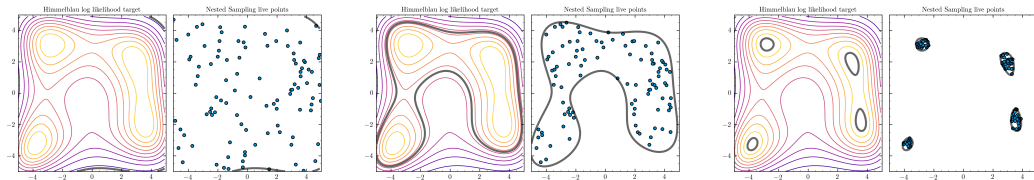
One configuration that trains robustly across all 8 models $\times$ 13 surveys, shipped with the package.

MCMC

- ▶ Single “walker” explores the posterior
- ▶ **Cannot** calculate $\mathcal{Z}$ (it cancels in the ratio)
- ▶ Parameter estimation only

Nested Sampling (Skilling 2006)

- ▶ Ensemble of “live points” scans prior $\rightarrow$ peak
- ▶ $\mathcal{Z} = \int \mathcal{L}(\theta) \pi(\theta) d\theta$ as **primary output**
- ▶ Unlocks **model comparison & tension quantification**
- ▶ Bonus: weighted posterior samples for free



Credit: David Yallup

Parameter Estimation

Using the observed data to estimate the values of cosmological parameters.

E.g. $H_0, \sigma_8, \tau, \omega_b h^2, \omega_c h^2 \dots$

Commonly uses the MCMC methods

Model Comparison

How much does the data support a particular model?
Comparison between two models.

E.g. Λ CDM model vs dynamic dark energy cosmology

Relies on the Bayesian evidence $\mathcal{Z}$

Tension Quantification

Do different datasets make consistent predictions from the same model? Study of discrepancies between different datasets.

E.g. CMB vs Type Ia supernovae data $\rightarrow$ Hubble tension

Also needs evidence $\mathcal{Z}$

Core Concept: Traditional MCMC is designed to measure *relative* probabilities. It bypasses the Evidence because it does not need to know the *absolute* volume of the probability space.

► The Math: Evidence Cancels Out

- Calculating the Evidence requires a hard multi-dimensional integral.
- Metropolis-Hastings avoids this by deciding jumps (from state A to B) using a **ratio**. Because Evidence is a constant, it cancels out:

$$\frac{\text{Posterior}(B)}{\text{Posterior}(A)} = \frac{\text{Likelihood}(B) \times \text{Prior}(B) / \text{Evidence}}{\text{Likelihood}(A) \times \text{Prior}(A) / \text{Evidence}} = \frac{\text{Likelihood}(B) \times \text{Prior}(B)}{\text{Likelihood}(A) \times \text{Prior}(A)}$$

$$\alpha(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}') = \min \left(1, \underbrace{\frac{P(\boldsymbol{\theta}'|\mathcal{D})}{P(\boldsymbol{\theta}|\mathcal{D})}}_{\text{Posterior Ratio}} \times \underbrace{\frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})}}_{\text{Hastings Factor}} \right)$$

- ▶ Pip-installable
 - ▶ `pip install unimpeded`
- ▶ Open-source
 - ▶ unimpeded on GitHub
 - ▶ Nested Sampling chains on Zenodo
- ▶ Simple API in Python for seamless download
- ▶ Compatible with plotting tools, anesthetic



```
from unimpeded.database import DatabaseExplorer

# Initialise DatabaseExplorer
dbe = DatabaseExplorer()

# Get a list of currently available models and datasets
models_list = dbe.models
datasets_list = dbe.datasets

# Choose model, dataset and sampling method
method = 'ns' # 'ns' for nested sampling, 'mcmc' for MCMC
model = 'klcdm' # from models_list
dataset = 'des_y1.joint+planck_2018_CamSpec' # from datasets_list

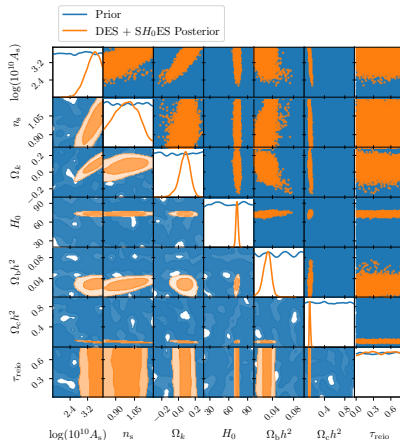
# Download samples chain
samples = dbe.download_samples(method, model, dataset)

# Download Cobaya and PolyChord run settings
info = dbe.download_info(method, model, dataset)
```

```
from unimpeded.tension import tension_calculator

tension_samples = tension_calculator(method='ns',
                                       model='lcdm',
                                       datasetA='planck_2018_CamSpec',
                                       datasetB='des_y1.joint',
                                       nsamples=1000)
```

- ▶ Posteriors from nested sampling and MCMC runs
- ▶ Pre-computed by HPC, accessible on your laptop in **seconds**
- ▶ Example: Planck + CMB lensing + SDSS using w_0w_a CDM
- ▶ Posteriors clearly more constrained than priors



- ▶ $\mathcal{R}$ statistics: $\mathcal{R} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A \mathcal{Z}_B} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$ (measure confidence in combining datasets)
- ▶ Information Ratio: $\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB}$ (KL-divergence-based)
- ▶ Suspiciousness: $\log \mathcal{S} = \log \mathcal{R} - \log I$ (prior-independent tension)
- ▶ p -value sigma: Approximate “number of sigma” separation
- ▶ Bayesian Model Dimensionality: $\frac{d}{2} = \int \mathcal{P}(\theta) \left(\log \frac{\mathcal{P}(\theta)}{\pi(\theta)} - \mathcal{D} \right)^2 d\theta$ (effective constrained parameter count)

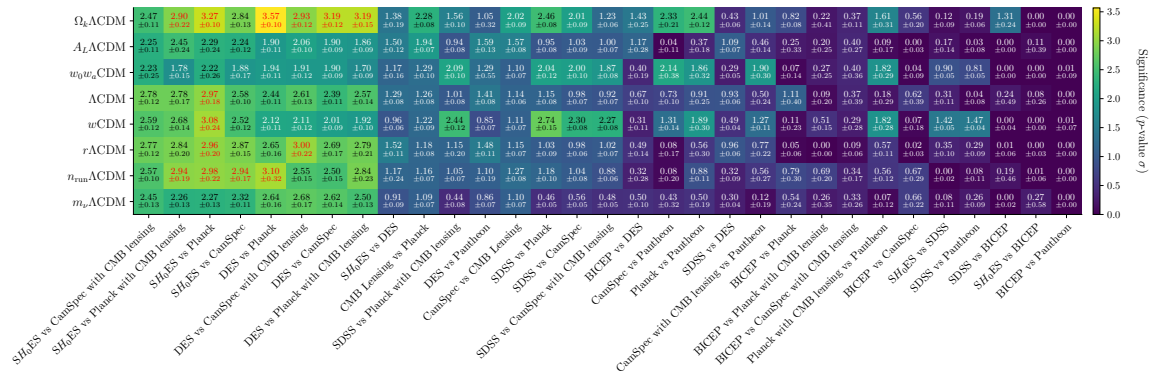
unimpeded provides a built-in calculator for all five statistics, operating directly on its library of nested sampling chains.

Backup: A Global View of Tensions

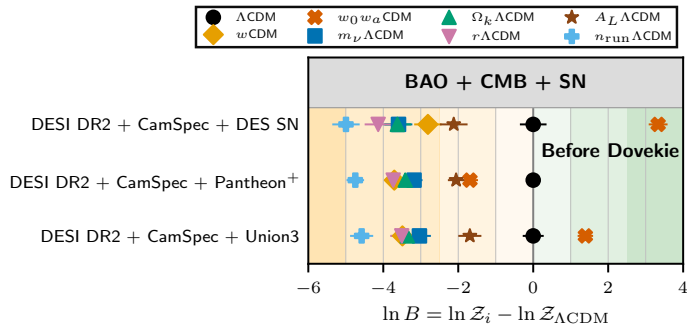


Tension statistics across 8 models and 31 pairwise datasets. Red: $\sigma > 2.88$ (look-elsewhere corrected).

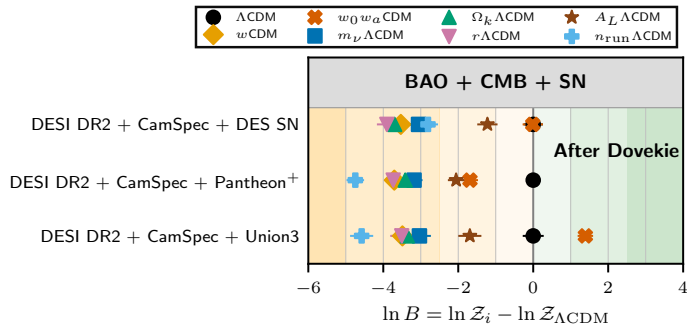
$$\sigma_{\text{threshold}} = \sqrt{2} \operatorname{Erfc}^{-1}\left(\frac{1}{N}\right)$$



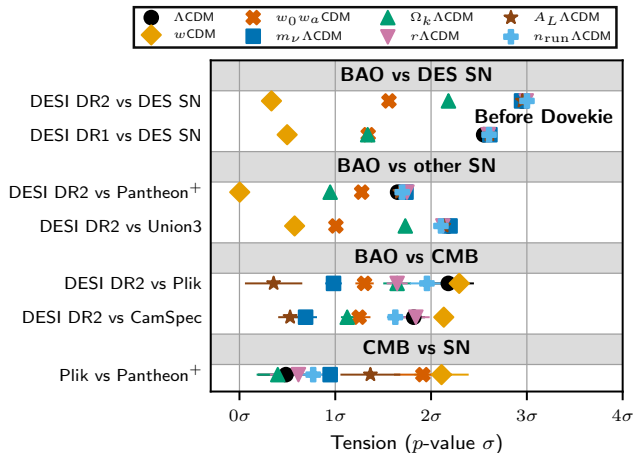
- ▶ DESI DR2 reported up to 4.2σ frequentist evidence for evolving dark energy
- ▶ Full Bayesian reanalysis across **248** dataset-model combinations
- ▶ Only the original DES-SN5YR drives the preference; Pantheon+ prefers Λ CDM
- ▶ **[click]** With the DES-Dovekie recalibration, the preference pulls back to Λ CDM



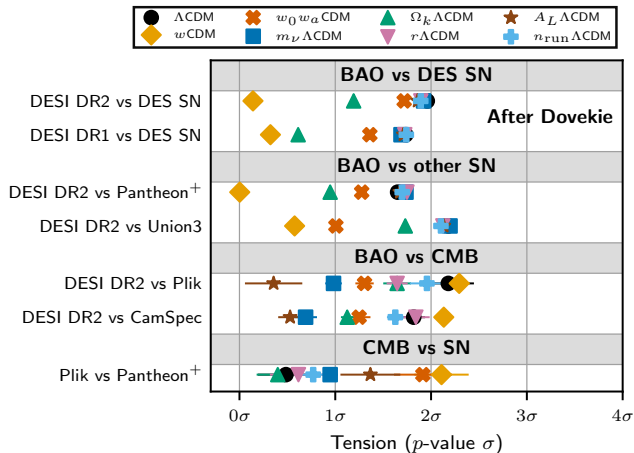
- ▶ DESI DR2 reported up to 4.2σ frequentist evidence for evolving dark energy
- ▶ Full Bayesian reanalysis across **248** dataset-model combinations
- ▶ Only the original DES-SN5YR drives the preference; Pantheon+ prefers Λ CDM
- ▶ **[click]** With the DES-Dovekie recalibration, the preference pulls back to Λ CDM



- ▶ **2.95 σ** conflict between DESI DR2 and DES-SN5YR under Λ CDM
- ▶ Bayesian model dimensionality $\approx 1 \Rightarrow$ localised, not a broad model failure
- ▶ The $w_0 w_a$ parameters simply absorb that tension
- ▶ **[click]** With DES-Dovekie: tension drops to **1.96 σ** and the dark-energy evidence largely vanishes



- ▶ **2.95 σ** conflict between DESI DR2 and DES-SN5YR under Λ CDM
- ▶ Bayesian model dimensionality $\approx 1 \Rightarrow$ localised, not a broad model failure
- ▶ The $w_0 w_a$ parameters simply absorb that tension
- ▶ **[click]** With DES-Dovekie: tension drops to **1.96 σ** and the dark-energy evidence largely vanishes



- ▶ **Principal Investigator**, an STFC-funded UKRI DiRAC seedcorn project **DP465 (100k CPU hours, CSD3/Cambridge)**
- ▶ *Mapping the Prior-Sensitivity of Cosmological Bayesian Evidence via Importance Sampling*
- ▶ Bayesian evidence, and hence model comparison, depends on the **choice of prior bounds** (e.g. the w_0, w_a ranges)
- ▶ **Importance sampling**: reweight pre-computed nested-sampling chains onto systematically narrowed priors to recompute $\mathcal{Z}$, with **no re-running**
- ▶ Quantifies how robust a verdict (“ Λ CDM vs $w_0 w_a$ CDM”) is to subjective prior choices

Reweight, don't rerun

$$\mathcal{Z}' = \mathcal{Z} \left\langle \frac{\pi'(\theta)}{\pi(\theta)} \right\rangle_{\mathcal{P}}$$

Evidence under a new prior π' from the *existing* posterior samples $\mathcal{P}$

Turns “does the answer depend on priors?” into a fast, systematic map.