Scanning for cosmological tensions

across a DiRAC-enabled grid of models, datasets and samplers (with AI coda)

Will Handley

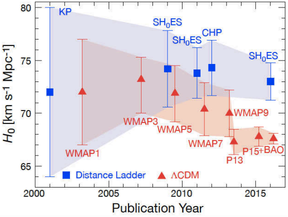
<wh260@cam.ac.uk>

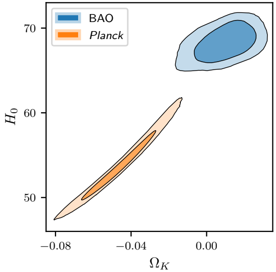
Royal Society University Research Fellow
Institute of Astronomy, University of Cambridge
Kavli Institute for Cosmology, Cambridge
Gonville & Caius College
willhandley.co.uk/talks

21st May 2025

The Challenge: Cosmological Tensions & Analysis Robustness

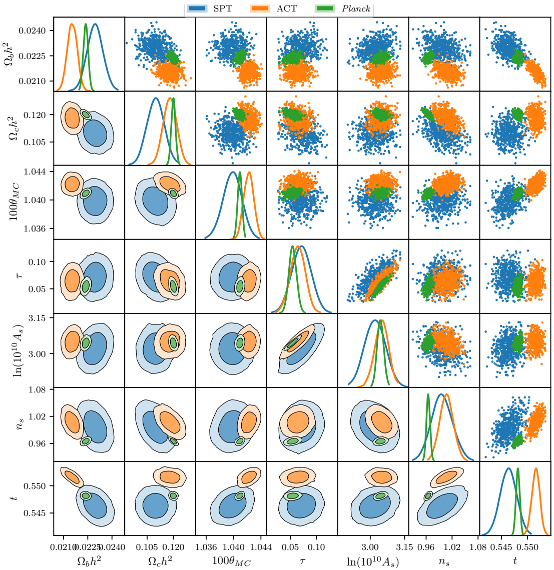
- Precision cosmology has revealed potential discrepancies in key parameters:
  - H_0 (Hubble constant) [1907.10625]
  - Ω_K (spatial curvature) [1908.09139], [1911.02087]
  - σ_8/S_8 (matter clustering) [1610.04606]?
- This underscores the critical importance of robust cosmological analysis.
- Key goals:
  - Accurately **quantifying tensions** between datasets and models.
  - Identifying and **addressing biases**, e.g., ΛCDM bias from fiducial assumptions in likelihoods.
- We need systematic exploration across a wide range of models and datasets to distinguish new physics from systematics.





The importance of global tension metrics

- If you have a well-defined parameter like H_0, this can be a good proxy for assessing the consistency of two fits
- However, in e.g. Planck vs DES we're asking: "In this 6 dimensional space, how well does this 6d grape fit with this 3d banana/blanket"?
- It's easy to make things look worse than they are when there are 6 degrees of freedom
- Inappropriate to pick a single parameter and quote a "sigma"
- Many methods have been developed to quantify "global" tensions [1902.04029].



Bayesian Inference Pillars & The Problem of Scale

Parameter Estimation

What do data tell us about model parameters?

$$\mathcal{P}(\theta|D,M) = \frac{\mathcal{L}(D|\theta,M)\pi(\theta|M)}{\mathcal{Z}(D|M)}$$

Model Comparison

How much do data support a model? (Occam's Razor [2102.11511])

$$P(M|D) = \frac{\mathcal{Z}(D|M)P(M)}{P(D)}$$

Relies on the Bayesian Evidence $\mathcal{Z}$.

Tension Quantification

Are datasets consistent under a model? [1902.04029]

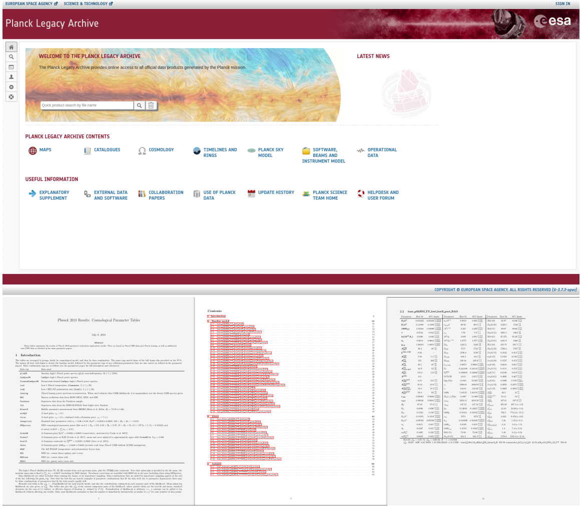
$$\mathcal{R} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A\mathcal{Z}_B}, \quad \log\mathcal{S} = \ldots$$

Also evidence-based.

- Model comparison and tension quantification are computationally hard, especially across many models/datasets.
- This necessitates efficient tools and pre-computed resources.

The Planck legacy archive

- *Planck* collaboration science products
- Distributed cosmology inference results as MCMC chains
- Across a grid of:
  - subsets/combinations of *Planck* data
    - TT, lowl, lowE, lensing
  - ΛCDM extensions
    - base, mnu, nrun, omegak, r
- importance sampling across some other likelihoods (BAO, JLA,...)
- Cannot compute evidences in high dimensions from MCMC chains
  - Only parameter estimation
  - no model comparison



DiRAC Allocations: Powering the Systematic Search

DiRAC 13: Next Gen Cosmo Analysis

- Goal: Create a Planck Legacy Archive (PLA) equivalent, but using nested sampling.
  - Enabling model comparison & tension quantification.
- Systematic scan over models and modern datasets (Planck, DES, SNe).
- A public grid of nested sampling & MCMC chains.

DiRAC 17: New Horizons in Cosmology

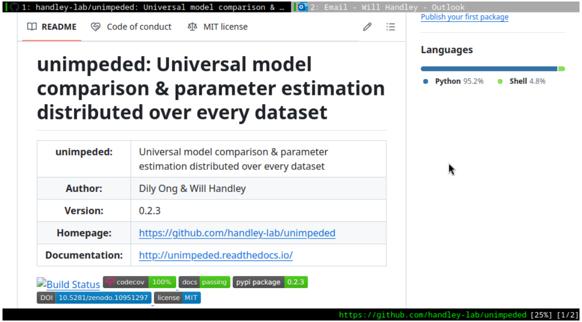
- Extending DiRAC 13 success.
- Incorporating next-gen data: DESI, DES Y5 SNe, Pantheon+, KiDS-1000, HSC, Euclid forecasts.

Introducing `unimpeded`

Universal Model comparison and Parameter Estimation Distributed over Every Dataset

Dily Ong
PhD

- **Core Idea:** A re-usable library of MCMC chains, Nested Sampling runs, and ML emulators. [*Abstract*]
- Built from DiRAC allocations (DP192 & DP264).
- **Systematic Coverage:**
  - Current: $\sim$10 cosmological models (e.g., ΛCDM, $k\Lambda$CDM, $w\Lambda$CDM).
  - $\sim$60 datasets & pairwise combinations (CMB, BAO, SNe, WL).
- Python tool for seamless download, upload, and caching of analysis products.
- Data stored on Zenodo; fast HDF5 local caching.
- Compatible with `anesthetic` for processing



`github.com/handley-lab/unimpeded`

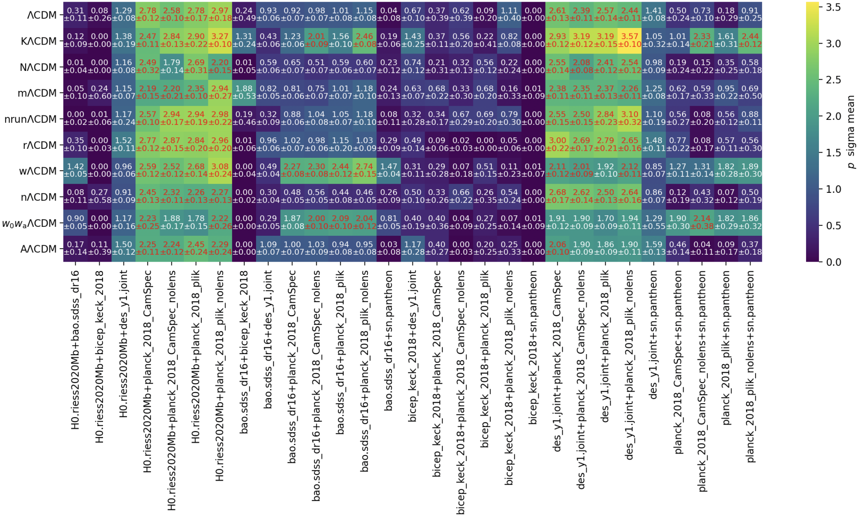
unimpeded in Action: Easy Access to Complex Results

Key Features & Benefits:

- Easily access pre-computed results with a few lines of Python.
- Facilitates:
  - Parameter estimation.
  - Model comparison (via nested sampling evidences).
  - Tension quantification.
  - Pairwise dataset comparisons.
- Removes computational barriers for many common analyses.
- Enables robust science by allowing checks over many models and data combinations.
- Foundation for building more complex, multi-probe analyses.

```python
from unimpeded.database import DatabaseExplorer
dbe = DatabaseExplorer()
planck = dbe.download_samples(
        method='ns', model='walcdm',
        dataset='planck_2018_CamSpec')
sdss = dbe.download_samples(
        method='ns', model='walcdm',
        dataset='bao.sdss_dr16')
planck_sdss = dbe.download_samples(
        method='ns', model='walcdm',
        dataset='bao.sdss_dr16+planck_2018_CamSpec')
```

- Models on y-axis
- Dataset combinations on x axis
- numbers and colours refer to σ-values for their tension
- Lots to unpack here (a very expensive plot)!

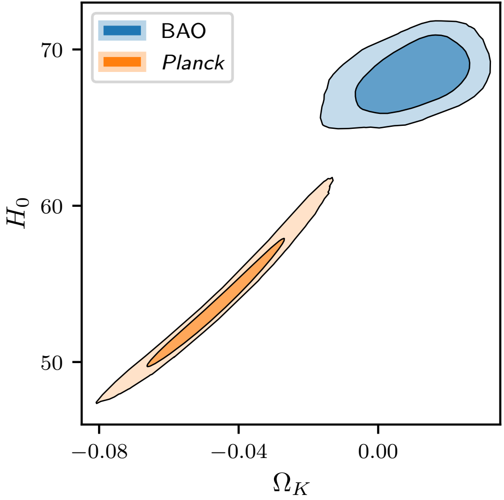


Addressing Biases: Beyond ΛCDM Fiducial Assumptions

- Many cosmological likelihoods (e.g., CMB lensing, BAO reconstruction) rely on fiducial ΛCDM assumptions for:
  - Simulation calibrations.
  - Theoretical templates.
  - Data correction algorithms.
- This can introduce an inherent bias towards ΛCDM or affect parameter inferences when testing extended models.
- Example: Curvature constraints from CMB lensing or BAO are derived assuming an underlying flat fiducial cosmology. This can mask or create artificial tensions.
- Robustly testing beyond-ΛCDM physics requires assessing the impact of these assumptions.
- `unimpeded`'s grid, with varied models and datasets, helps explore these effects without being locked into one fiducial for all analyses.

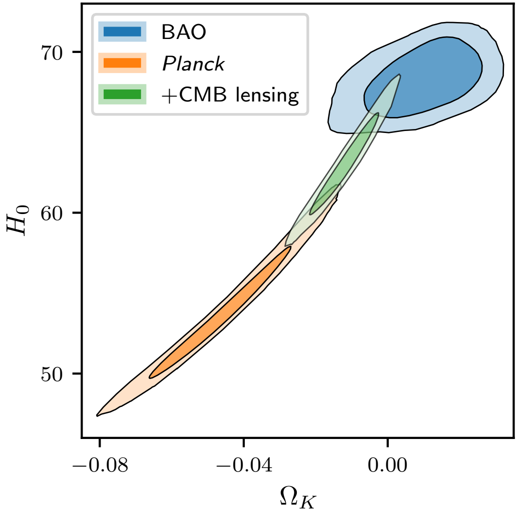
Curvature tension Ω_K

- ΛCDM assumes the universe is flat
- If you allow $\Omega_K \neq 0$, *Planck* (plikTTTEEE) has a moderate preference for closed universes (50:1)
- *Planck*+CMB lensing +BAO strongly prefer a flat universe
- But, *Planck* vs lensing is 2.5σ in tension, and Planck vs BAO is 3σ.
- This is reduced if plik $\to$ camspec
  - Di Valentino et al [1911.02087]
  - Handley [1908.09139]
  - Efsthathiou & Gratton [2002.06892]
- BAO & lensing summary stats and compression assume ΛCDM [2205.05892].



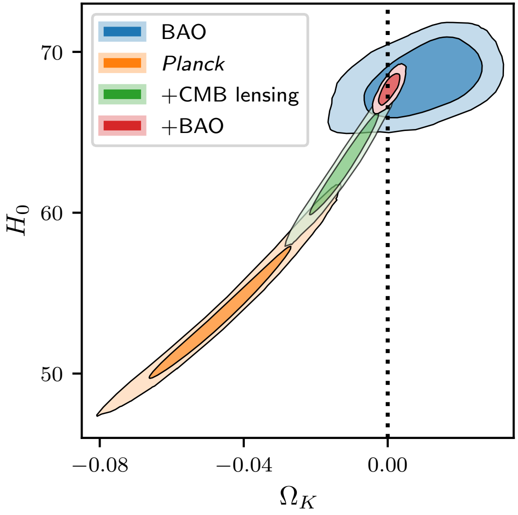
Curvature tension Ω_K

- ΛCDM assumes the universe is flat
- If you allow $\Omega_K \neq 0$, *Planck* (`plikTTTEEE`) has a moderate preference for closed universes (50:1)
- *Planck*+CMB lensing +BAO strongly prefer a flat universe
- But, *Planck* vs lensing is 2.5σ in tension, and Planck vs BAO is 3σ.
- This is reduced if `plik` $\rightarrow$ `camspec`
  - Di Valentino et al [1911.02087]
  - Handley [1908.09139]
  - Efsthathiou & Gratton [2002.06892]
- BAO & lensing summary stats and compression assume ΛCDM [2205.05892].



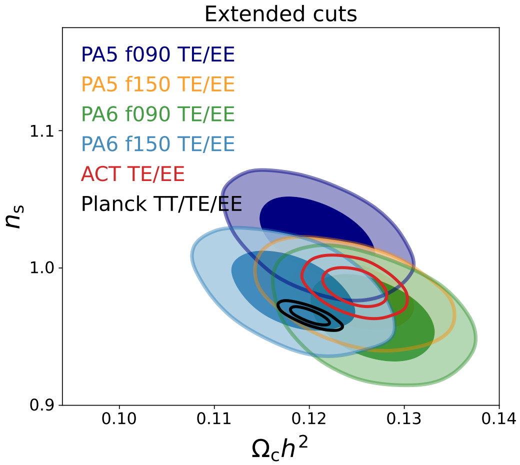
Curvature tension Ω_K

- ΛCDM assumes the universe is flat
- If you allow $\Omega_K \neq 0$, *Planck* (`plikTTTEEE`) has a moderate preference for closed universes (50:1)
- *Planck*+CMB lensing +BAO strongly prefer a flat universe
- But, *Planck* vs lensing is 2.5σ in tension, and Planck vs BAO is 3σ.
- This is reduced if `plik` $\rightarrow$ `camspec`
  - Di Valentino et al [1911.02087]
  - Handley [1908.09139]
  - Efsthathiou & Gratton [2002.06892]
- BAO & lensing summary stats and compression assume ΛCDM [2205.05892].



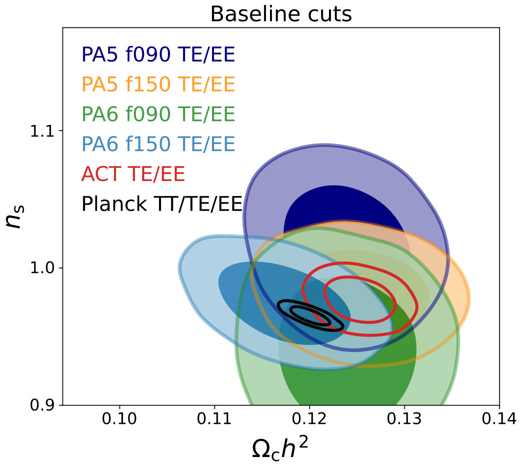
A Note on Blinding: Aspirations vs. Reality

- Blinding is a common strategy to mitigate confirmation bias in cosmological analyses.
- However, the ideal of a pure "one-shot" unblinding is often challenged in practice.
- Examples from recent large analyses (e.g., DES Y3, ACT DR4/DR6) show that post-unblinding adjustments or re-evaluations do occur.
  - These might be necessary due to unforeseen complexities or subtle effects.
  - But they invalidate the intended benefits of the blinding protocol.
- Openly available chains and analysis tools (like `unimpeded`) can aid in community scrutiny and reproducibility, whether analyses are blinded or not.



A Note on Blinding: Aspirations vs. Reality

- Blinding is a common strategy to mitigate confirmation bias in cosmological analyses.
- However, the ideal of a pure "one-shot" unblinding is often challenged in practice.
- Examples from recent large analyses (e.g., DES Y3, ACT DR4/DR6) show that post-unblinding adjustments or re-evaluations do occur.
  - These might be necessary due to unforeseen complexities or subtle effects.
  - But they invalidate the intended benefits of the blinding protocol.
- Openly available chains and analysis tools (like `unimpeded`) can aid in community scrutiny and reproducibility, whether analyses are blinded or not.



Machine Learning Enhancement: Emulators

Harry Bevins

JRF

- The `unimpeded` framework includes providing machine learning emulators. [*Abstract*]
- These emulate marginalised likelihoods or posteriors.
- Based on techniques like Masked Autoregressive Flows (MAFs) (e.g., as in `margarine` [2205.12841], [2207.11457]).
- **Benefits:**
  - Dramatically speeds up inference for re-use in new analyses.
  - Provides a fast and flexible alternative to full MCMC/NS runs for many applications.
  - Enables easy application of, e.g., a "Planck prior" without running Planck likelihoods.
- Part of the DiRAC 17 vision: leveraging advanced SBI and ML.

The Future: Next-Gen Tools AI in Cosmological Robustness

Simulation-Based Inference (SBI)

- Shifts focus from explicit likelihoods to forward simulations. [*ERC Grant*, *DiRAC 17 CfS*]
- Powerful for complex, non-linear probes and intractable likelihoods.

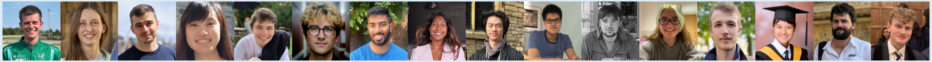
BlackJAX Nested Sampling

- New JAX-based nested sampling algorithm.
- Designed for modern hardware (GPUs).
- Aims for high performance and accessibility (open source).
  github.com/williamjameshandley/talks/blob/cosmoverse_2025/blackjax_ns.py

Large Language Models (LLMs)

- LLMs (like the one assisting with this talk) show increasing potential:
  - Rapid code generation for analysis pipelines (e.g., custom Boltzmann solvers).
  - Synthesis of complex information, literature review.
  - Structuring analyses and drafting papers/talks.
- A capable prompt engineer could significantly accelerate research workflows.

Summary



github.com/handley-lab/group

- Cosmological tensions (e.g., H_0, S_8, Ω_K) demand systematic and robust analysis frameworks.
- `unimpeded`, powered by DiRAC allocations (DP192 & DP264), provides a growing, publicly accessible library of:
  - MCMC chains & Nested Sampling runs.
  - ML-emulators for fast likelihood/posterior evaluations.
- Covers $\sim$10 models and $\sim$60 dataset combinations (and growing).
- Facilitates parameter estimation, model comparison, and critical tension quantification.
- Key for addressing fiducial biases and rigorously exploring beyond-ΛCDM physics.
- **Future:**
  - Expansion with DiRAC 17: Next-gen datasets (DESI, Euclid, etc.).
  - Integration of advanced SBI and AI/LLM techniques.
  - Continued community-focused development.
- We are seeking α-testers and collaborators!

Bonus Slide: The Role of LLMs in This Talk

This presentation was drafted with the assistance of a Large Language Model. LLMs can be powerful tools for:

- Synthesizing complex information from multiple lengthy documents.
- Structuring scientific narratives and identifying key talking points.
- Generating initial drafts for talks, papers, and code.
- Assisting with literature reviews.
- Brainstorming slide content and visual aids.

The future of scientific research, communication, and analysis will likely involve closer and more sophisticated human-AI collaboration.

Bonus Slide: The Role of LLMs in This Talk

This presentation was drafted with the assistance of a Large Language Model. LLMs can be powerful tools for:

- Synthesizing complex information from multiple lengthy documents.
- Structuring scientific narratives and identifying key talking points.
- Generating initial drafts for talks, papers, and code.
- Assisting with literature reviews.
- Brainstorming slide content and visual aids.

The future of scientific research, communication, and analysis will likely involve closer and more sophisticated human-AI collaboration.

```
131                    0:/tmp/vLgnzZ6-3-aichat*            will@maxwell 21 May 07:48
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%07:46:25 [442/36382]
% The Challenge: Cosmological Tensions & Analysis Robustness
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{frame}
    \frametitle{The Challenge: Cosmological Tensions \& Analysis Robustness}
    \begin{columns}[T]
        \column{0.55\textwidth}
        \begin{itemize}
            \item Precision cosmology has revealed persistent discrepancies in key parameters:
                \begin{itemize}
                    \item $H_0$ (Hubble constant) [\arxiv{1907.10625}]
                    \item $\sigma_8$/$S_8$ (matter clustering) [\arxiv{1610.04606}]
                    \item $\Omega_K$ (spatial curvature) [\arxiv{1908.09139}, \arxiv{1911.02087}
]
                \end{itemize}
            \item This underscores the critical importance of robust cosmological analysis.
            \item Key goals:
                \begin{itemize}
                    \item Accurately \textbf{quantifying tensions} between datasets and models.
                    \item Identifying and \textbf{addressing biases}, e.g., $\Lambda$CDM bias fr
om fiducial assumptions in likelihoods.
                \end{itemize}
            \item We need systematic exploration across a wide range of models and datasets to d
istinguish new physics from systematics.
        \end{itemize}

        \column{0.45\textwidth}
        \begin{figure}
            \centering
            % Figure from ERC Grant, Fig 1
            \includegraphics[width=\columnwidth]{placeholder_COSMOLOGICAL_TENSIONS_H0_S8_OMEGAK.
png}
            \caption{Discrepancies in $H_0$, $S_8$ and $\Omega_K$ highlight the challenge. (Figu
re adapted from \arxiv{1908.09139} and others)}
        \end{figure}
    \end{columns}
\end{frame}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Bayesian Inference Pillars & The Problem of Scale
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
\begin{frame}
    \frametitle{Bayesian Inference Pillars \& The Problem of Scale}
    \begin{columns}[T]
        \column{0.33\textwidth}
        \begin{block}{Parameter Estimation}
            What do data tell us about model parameters?
            \[ \mathcal{P}(\theta|D,M) = \frac{\mathcal{L}(D|\theta,M) \pi(\theta|M)}{\mathcal{Z
}(D|M)} \]
```