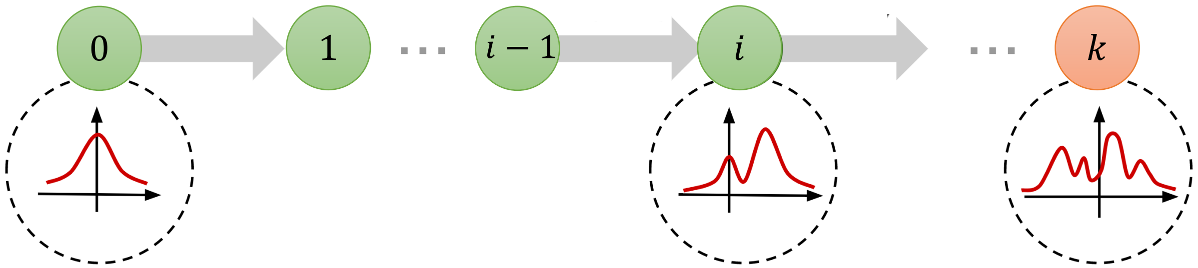
Normalizing flows for machine learning assisted Bayesian model comparison

Alicja Polanska, Matthew A. Price, Davide Piras, Alessio Spurio Mancini and Jason D. McEwen



Learned harmonic mean estimator

Estimator of the Bayesian evidence

Use with any MCMC sampler or on saved down chains

harmonic Python package

github.com/astro-informatics/harmonic

Outline of this talk

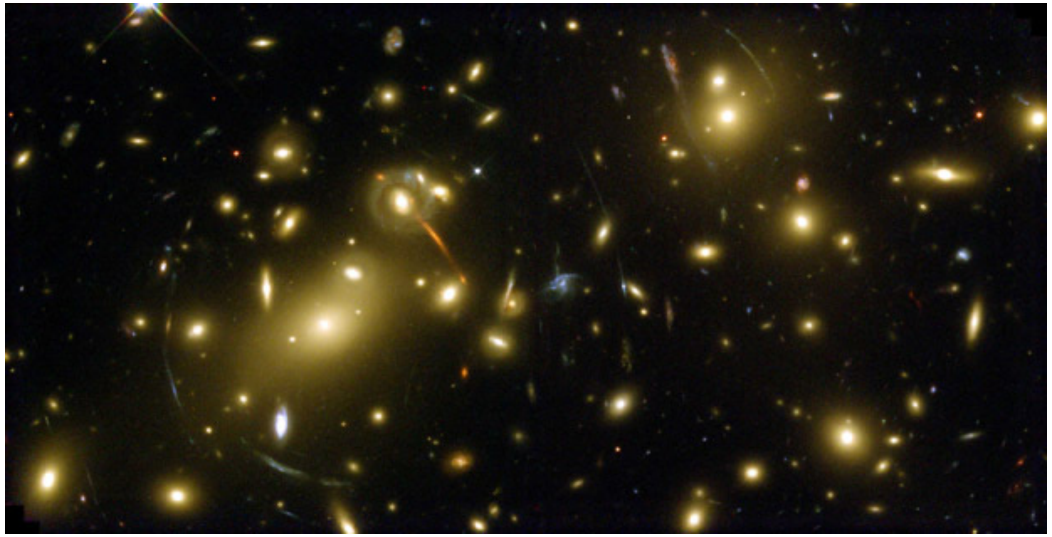
1. Learned harmonic mean estimator
2. High-dimensional model comparison for cosmology
3. Accelerated model comparison for gravitational waves

Outline of this talk

1. Learned harmonic mean estimator
2. High-dimensional model comparison for cosmology
3. Accelerated model comparison for gravitational waves

Model comparison

What model best describes the universe?



ΛCDM or wCDM?

Bayesian model comparison

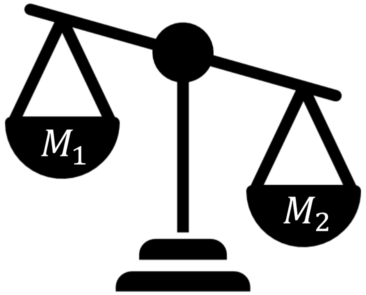
In the Bayesian framework probability distributions provide a quantification of uncertainty.

$$P(\theta|\boldsymbol{y}, M) = \frac{\overbrace{P(\boldsymbol{y}|\theta, M)}^{\text{Likelihood}}\,\overbrace{P(\theta|M)}^{\text{Prior}}}{\underbrace{P(\boldsymbol{y}|M)}_{\text{Bayesian Evidence}}} = \frac{\mathcal{L}(\theta)\pi(\theta)}{z}$$

$$z = \mathrm{P}(\boldsymbol{y}|M) = \int \mathrm{d}\theta\, \mathcal{L}(\theta)\pi(\theta)$$

Which model to choose?

Bayesian evidence tells us which scientific model is more plausible



Very useful but hard to compute!

Harmonic mean estimator

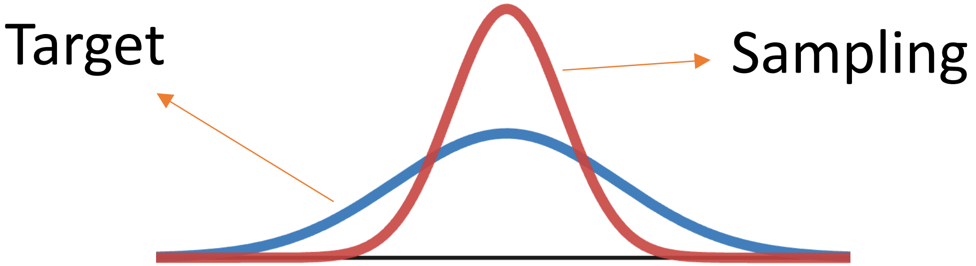
Estimator of evidence (Newton and Raftery, 1994)

$$\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{1}{\mathcal{L}(\theta)}\right] = \frac{1}{z}$$

☺ It's agnostic to sampling method $\longrightarrow$ It's flexible

☹ But…. fails catastrophically

Why does it fail?



Can be interpreted as importance sampling

$$\rho = \int \mathrm{d}\theta \, \frac{1}{\mathcal{L}(\theta)} \, \mathrm{P}(\theta \,|\, y)$$

$$= \int \mathrm{d}\theta \, \frac{1}{z} \, \frac{\pi(\theta)}{\mathrm{P}(\theta \,|\, y)} \, \mathrm{P}(\theta \,|\, y)$$

Target density

Sampling density

Target density has fatter tails than sampling density

Harmonic mean estimator fails

Learned harmonic mean estimator

Introduce arbitrary normalized target density $\varphi(\theta)$ (Gelfand and Dey, 1994)

$$\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{1}{\mathcal{L}(\theta)}\right] = \frac{1}{z} \quad \Rightarrow \quad \boxed{\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{\varphi(\theta)}{\mathcal{L}(\theta)\pi(\theta)}\right]}$$

Learned harmonic mean estimator

Introduce **learned harmonic mean estimator** (McEwen et al., 2021) :

$\varphi(\theta)$ is learned from posterior samples

$$\psi^{\mathrm{ML}} \approx \psi^{\mathrm{optimal}}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{z}$$

$$\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{\varphi(\theta)}{\mathcal{L}(\theta)\pi(\theta)}\right]$$

Learned harmonic mean estimator

Introduce **learned harmonic mean estimator** (McEwen et al., 2021) :

$\varphi(\theta)$ is learned from posterior samples

$$\psi^{\mathrm{ML}} \approx \psi^{\mathrm{optimal}}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{z}$$

$$\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{\varphi(\theta)}{\mathcal{L}(\theta)\pi(\theta)}\right]$$

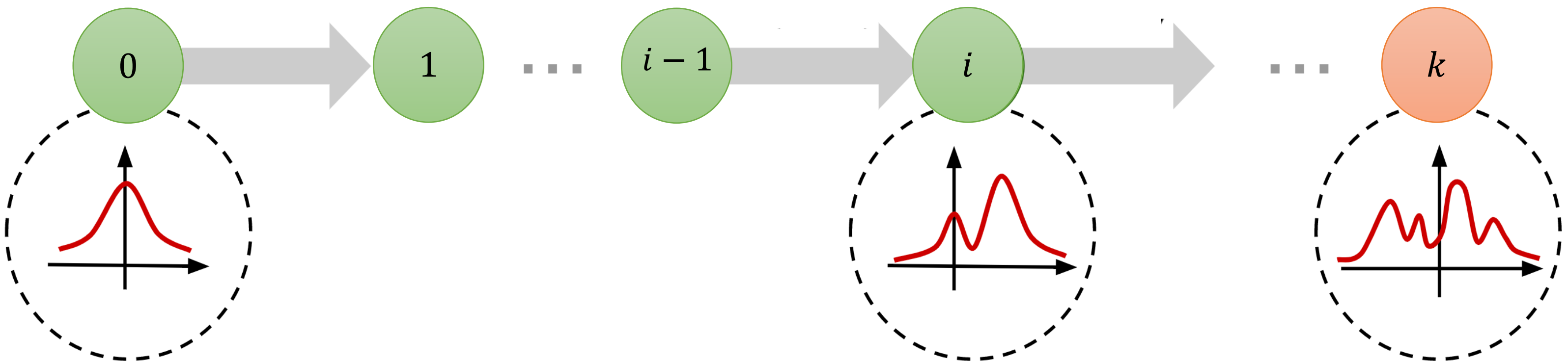
☹ Requires bespoke training approach and fine-tuning

➡ Use normalizing flows to solve these issues!

(Polanska et al., 2024) arXiv:2405.05969

Normalizing flows

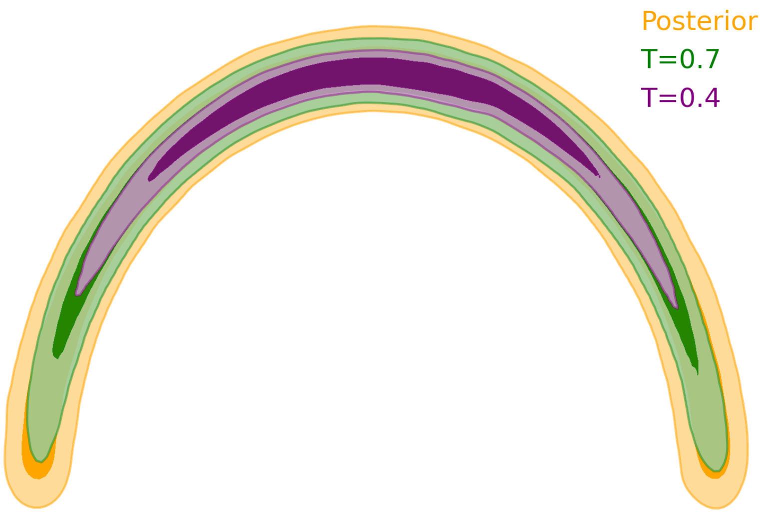
Normalizing flows take a simple base distribution through a series of reversible transformations to approximate a complex distribution



Adapted from lilianweng.github.io/posts/2018-10-13-flow-models

We use real non-volume preserving and rational-quadratic spline flows

Concentrating the target distribution



We train a flow on samples from the posterior and introduce temperature parameter T to concentrate the probability density

The base distribution variance is scaled by

$$0 < T < 1$$

Learned harmonic mean estimator

Train normalizing flow on posterior samples

Concentrate probability density

Use concentrated flow as $\varphi(\theta)$

Evidence estimate

$$\rho = \mathbb{E}_{P(\theta|\boldsymbol{y})}\left[\frac{\varphi(\theta)}{\mathcal{L}(\theta)\pi(\theta)}\right]$$

Learned harmonic mean estimator

Our method provides a tool for Bayesian model comparison that is:

- ☑ Accurate
- ☑ Scalable
- ☑ Robust
- ☑ Flexible

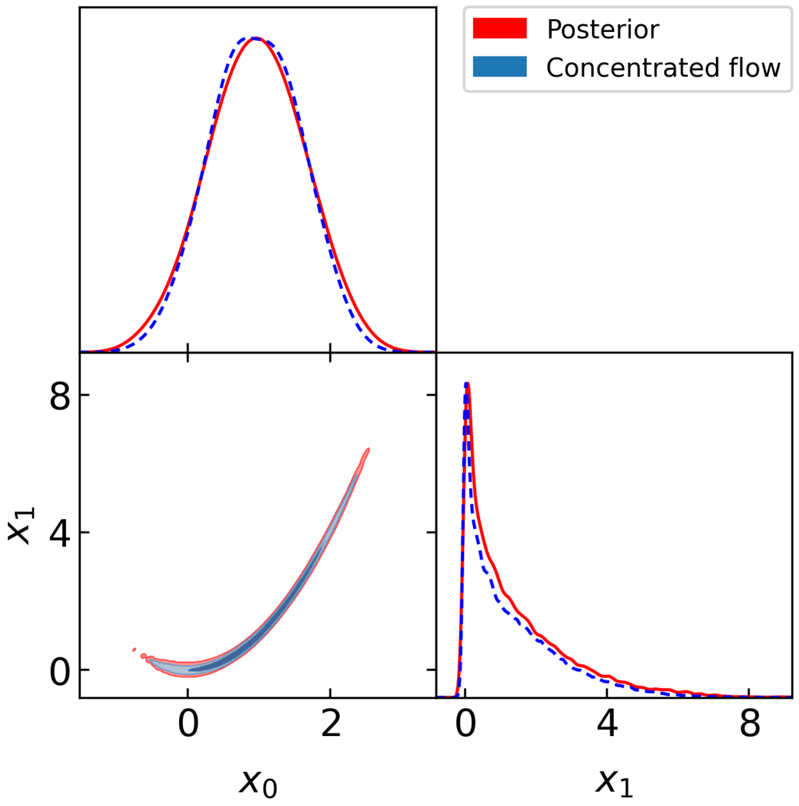
harmonic software

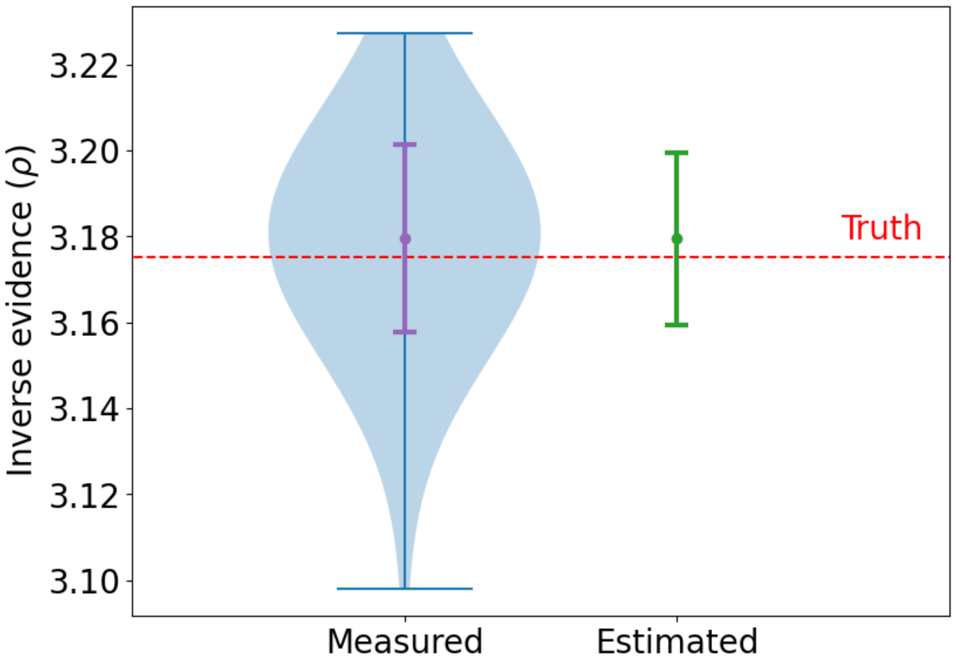
harmonic Python package[1] has been made available in the new release of harmonic on PyPi and GitHub

SCAN ME

[1]github.com/astro-informatics/harmonic

Rosenbrock example





Outline of this talk

1. Learned harmonic mean estimator
2. High-dimensional model comparison for cosmology
3. Accelerated model comparison for gravitational waves

SciAI UCL

High-dimensional model comparison for cosmology

Piras and Spurio Mancini, 2023

Emulation (CosmoPower-JAX)

+

Differentiable and probabilistic programming

+

Scalable sampling (NUTS)

+

Decoupled and scalable evidence (*harmonic*)

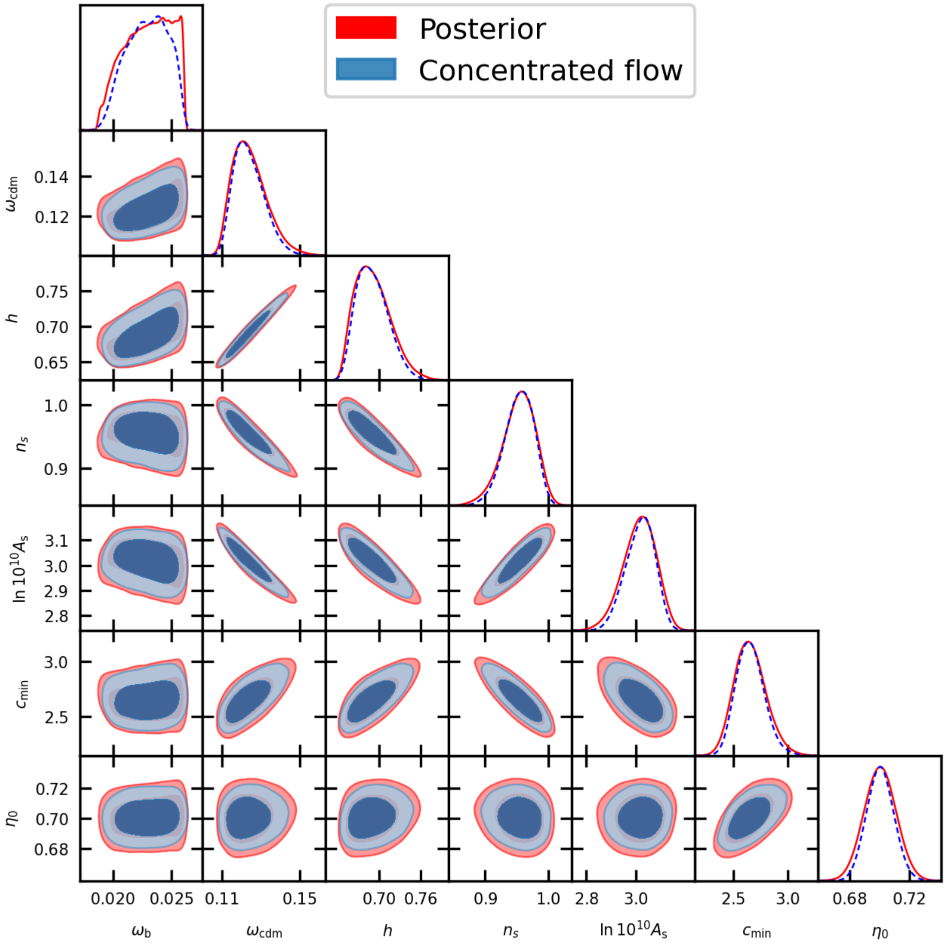
=

HARMONIC

The future of cosmological likelihood-based inference…

(Piras et al., 2024) arXiv:2405.12965

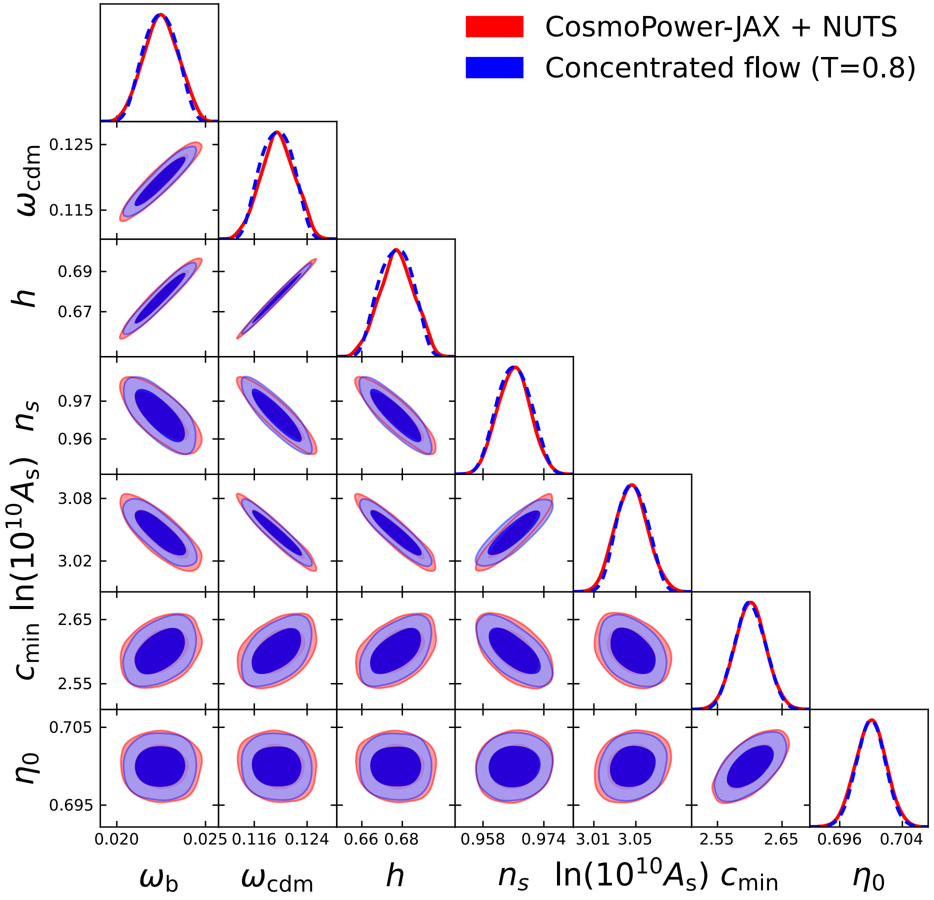
High-dimensional model comparison for cosmology



ΛCDM vs $w_0 w_a$CDM in 37/39D

Method	$\Delta \log z$	Computation time
CAMB + Nested sampling	0.78 ± 0.79	8 months on 48 CPUs
CosmoPower-JAX + NUTS + harmonic	$1.53^{0.07}_{-0.07}$	2 days on 12 GPUs

High-dimensional model comparison for cosmology



ΛCDM vs w_0w_aCDM in 157/159D

Method	$\Delta \log z$	Computation time
CAMB + Nested sampling	Not feasible	Estimated 12 years on 48 CPUs
CosmoPower-JAX + NUTS + harmonic	$1.9^{0.7}_{-0.5}$	8 days on 24 GPUs

Outline of this talk

1. Learned harmonic mean estimator
2. High-dimensional model comparison for cosmology
3. Accelerated model comparison for gravitational waves

Accelerated model comparison for gravitational waves

Wong et al., 2023a,b

Accelerated differentiable gravitational waveform models (Jim)

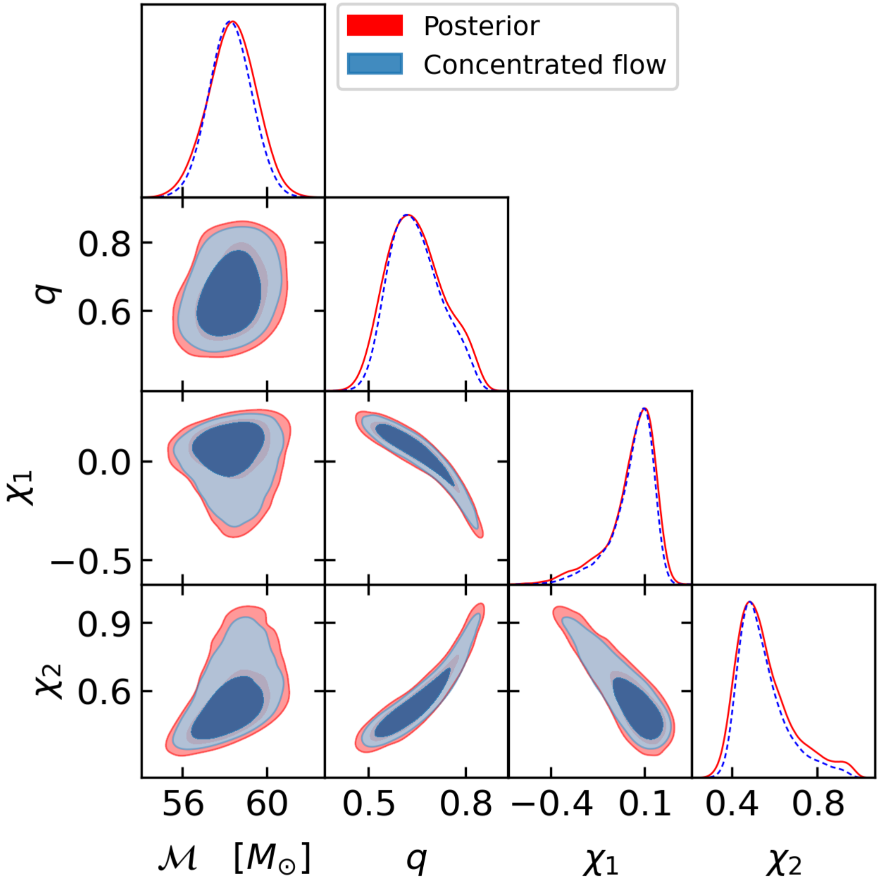
+

Normalizing-flow assisted sampling (flowMC)

+

Decoupled and scalable evidence (*harmonic)*

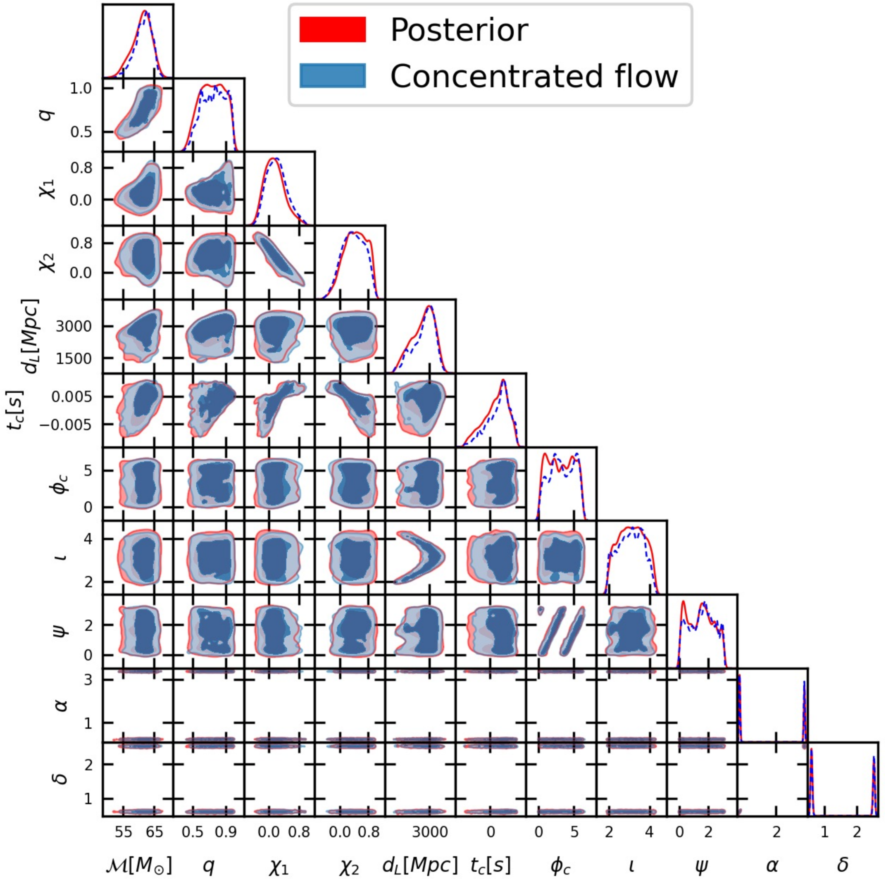
Accelerated model comparison for gravitational waves



Evidence for simulated GW event in 4D

Method	$\log z$	Computation time
Nested sampling	390.33 ± 0.11	31.3 min on 16 CPUs
Jim + harmonic	$390.360\,^{0.006}_{-0.006}$	5.3 min on 1 GPU

Accelerated model comparison for gravitational waves



Evidence for simulated GW event in 11D

Method	$\log z$	Computation time
Nested sampling	378.29 ± 0.15	3.5 h on 16 CPUs
Jim + harmonic	$378.420\ ^{0.09}_{-0.08}$	14.2 min on 1 GPU

Summary: Learned harmonic mean

Method to estimate the evidence that is

SCAN ME

- ☑ **Accurate:** based on a principled statistical framework
- ☑ **Robust:** no fine-tuning
- ☑ **Scalable:** analysis in 159 dimensions
- ☑ **Flexible:** use with any MCMC sampler, saved down chains, or any variational inference approach…

Summary: Learned harmonic mean

Method to estimate the evidence that is

- **Accurate:** based on a principled statistical framework
- **Robust:** no fine-tuning
- **Scalable:** analysis in 159 dimensions
- **Flexible:** use with any MCMC sampler, saved down chains, or any variational inference approach…

… or your application!

SCAN ME

References

Jason D. McEwen, Christopher G. R. Wallis, Matthew A. Price, and Alessio Spurio Mancini. Machine learning assisted Bayesian model comparison: learnt harmonic mean estimator, 2023 arXiv:2111.12720

Michael A. Newton and Adrian E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):3–26, 1994.

Davide Piras and Alessio Spurio Mancini. CosmoPower-JAX: high-dimensional Bayesian inference with differentiable cosmological emulators. *The Open Journal of Astrophysics*, 6, June 2023. ISSN 2565-6120.

Davide Piras and Alicja Polanska and Alessio Spurio Mancini and Matthew A. Price and Jason D. McEwen. 'The future of cosmological likelihood-based inference: accelerated high-dimensional parameter estimation and model comparison', https://arxiv.org/abs/2405.12965

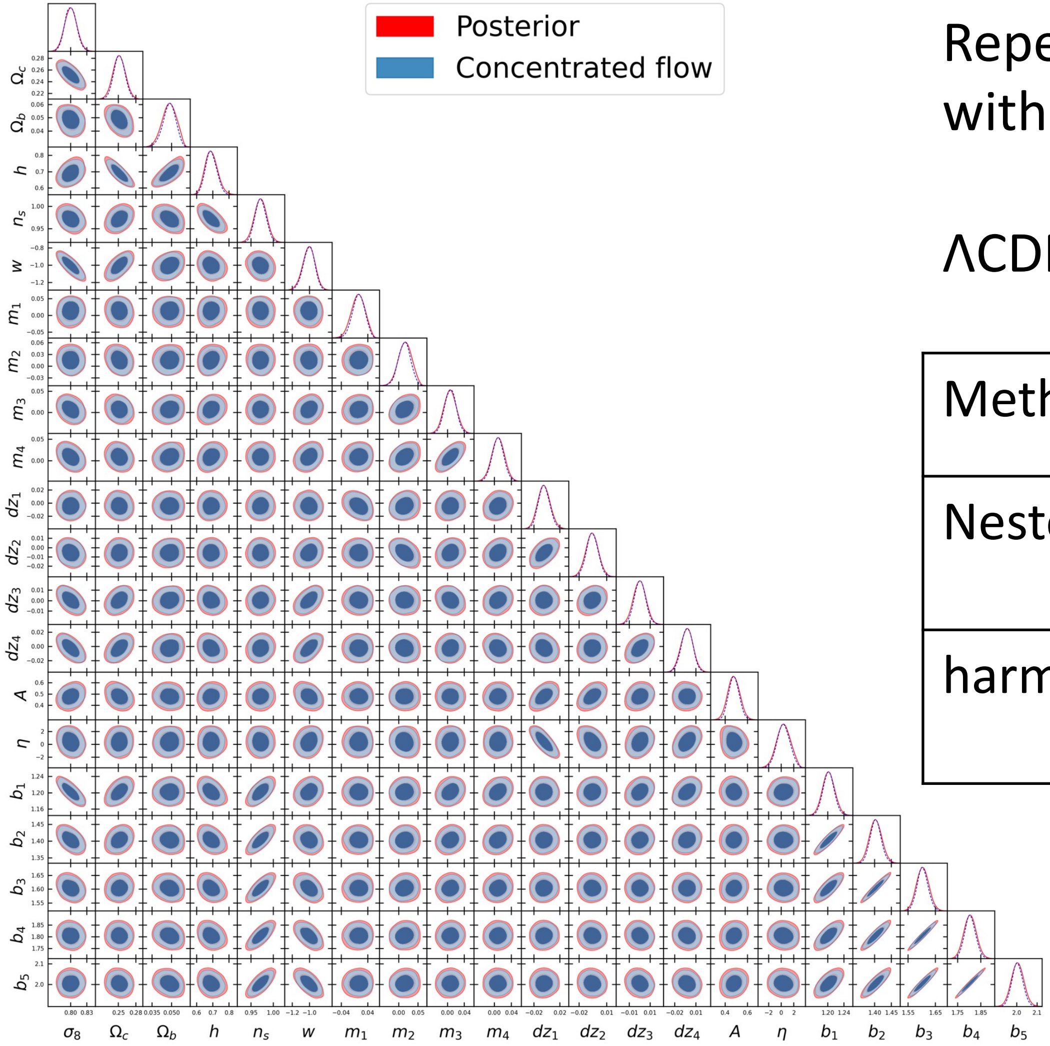
Alicja Polanska, Matthew A. Price, Davide Piras, Alessio Spurio Mancini, and Jason D. McEwen. 'Learned Harmonic Mean Estimation of the Bayesian Evidence with Normalizing Flows'. arXiv [Astro-Ph.IM], 9 May 2024. arXiv. http://arxiv.org/abs/2405.05969.

Kaze W. K. Wong, Maximiliano Isi, and Thomas D. P. Edwards. Fast Gravitational-wave Parameter 183 Estimation without Compromises. Astrophys. J., 958(2):129, 2023.

Kaze W. K. Wong, Marylou Gabri., and Daniel Foreman-Mackey. flowMC: Normalizing flow enhanced sampling package for probabilistic inference in JAX. J. Open Source Softw., 8(83):5021, 2023.

A. Spurio Mancini, M. M. Docherty, M. A. Price, and J. D. McEwen. Bayesian model comparison for simulation-based inference. RASTI, submitted, Jul 2022.

DES Y1 Example



Repeat DES Y1 3x2pt analysis from (Campagne et al., 2023) with harmonic

ΛCDM vs wCDM in 20D

Method	$\Delta \log z$	Computation time
Nested sampling	2.23 ± 0.64	94h on 64 CPU
harmonic	2.15 ± 0.01	16h on 64CPU + 16min