Bayesian optimisation for efficient cosmological model selection

Ameek Malhotra

Work in progress with Nathan Cohen and Jan Hamann

Swansea University
Prifysgol Abertawe

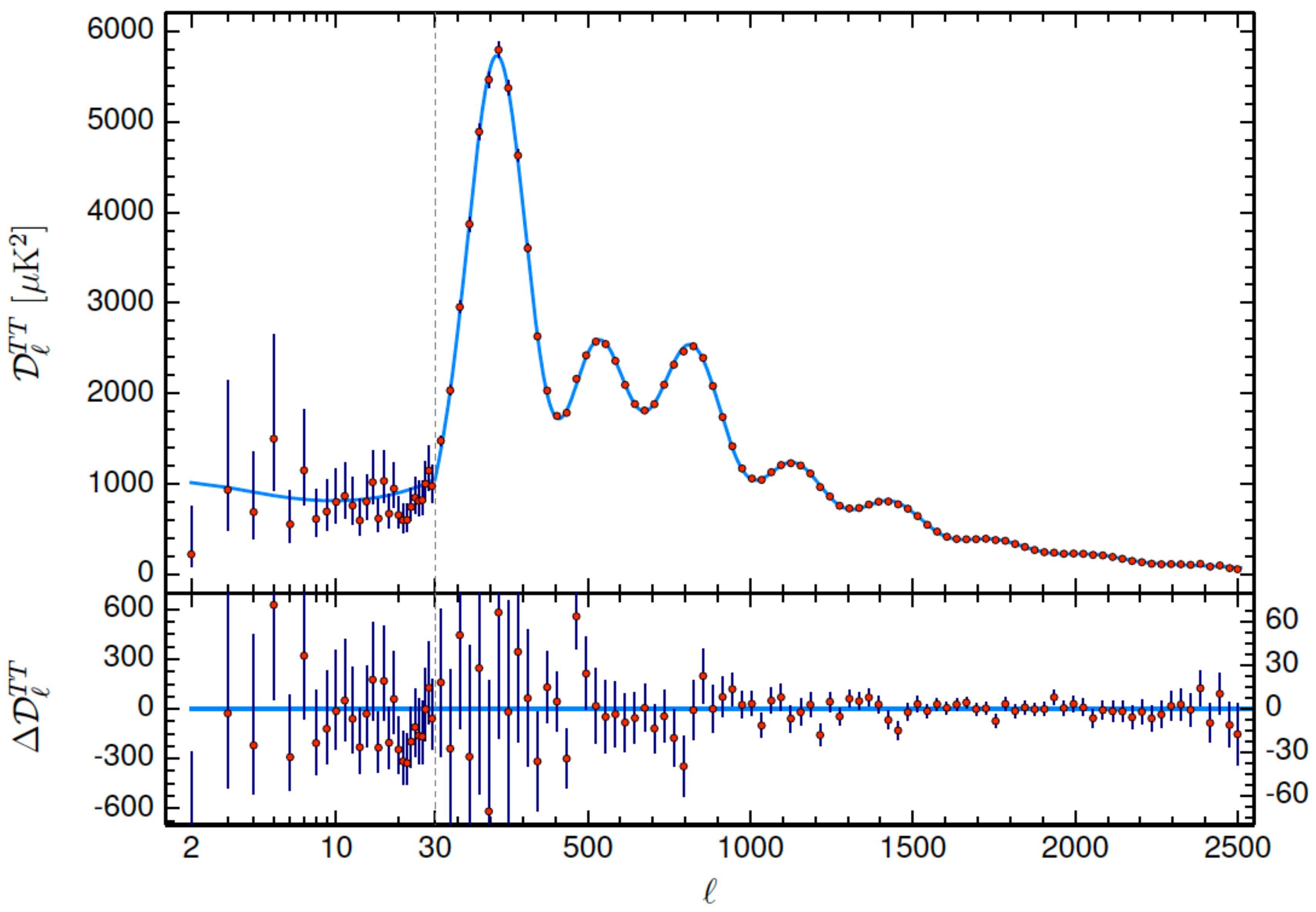
PASCOS 2025, Durham

Bayesian Inference

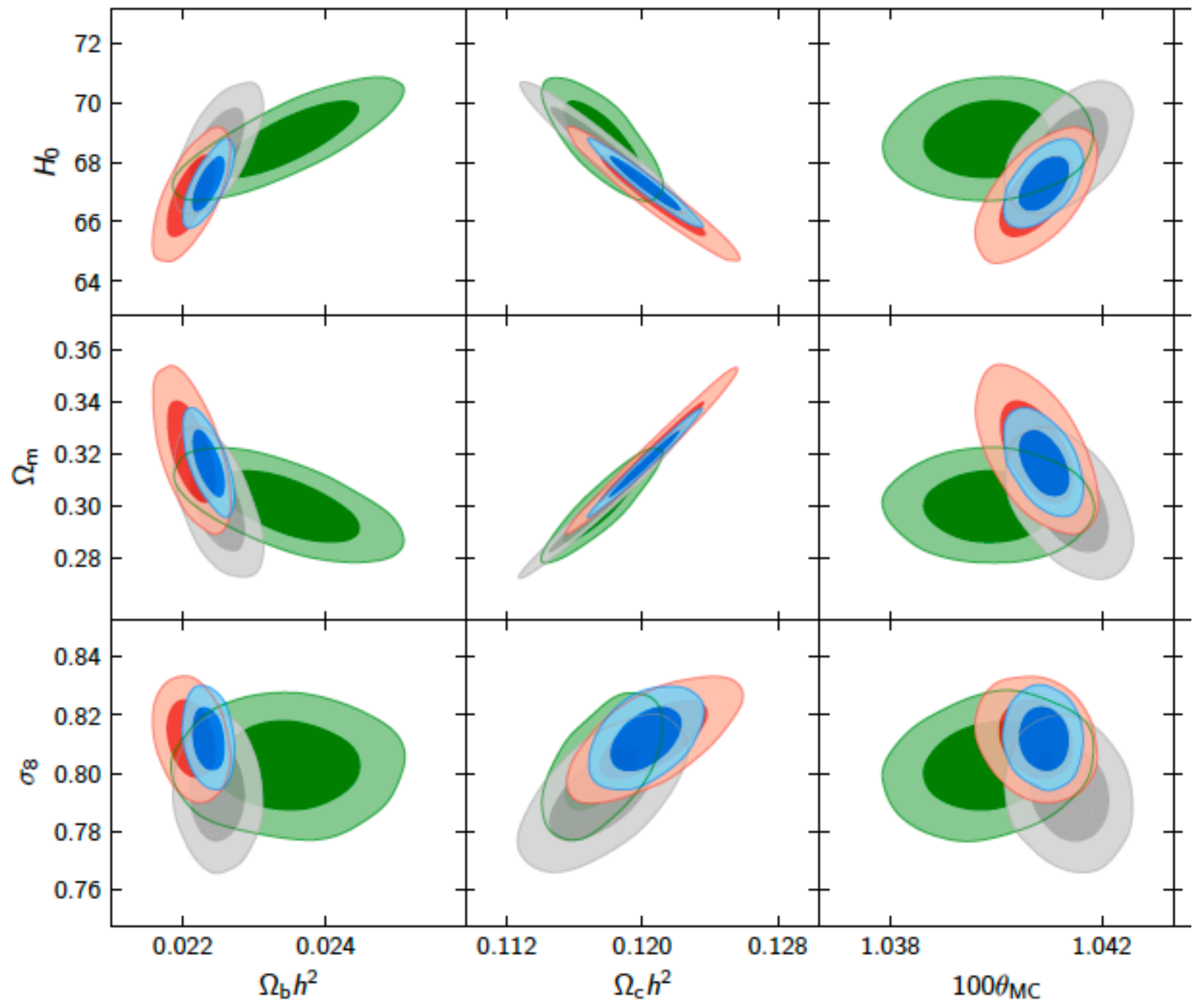
The different levels of inference

1. Parameter constraints, given a model

$$P(\theta|D, M)$$



LCDM



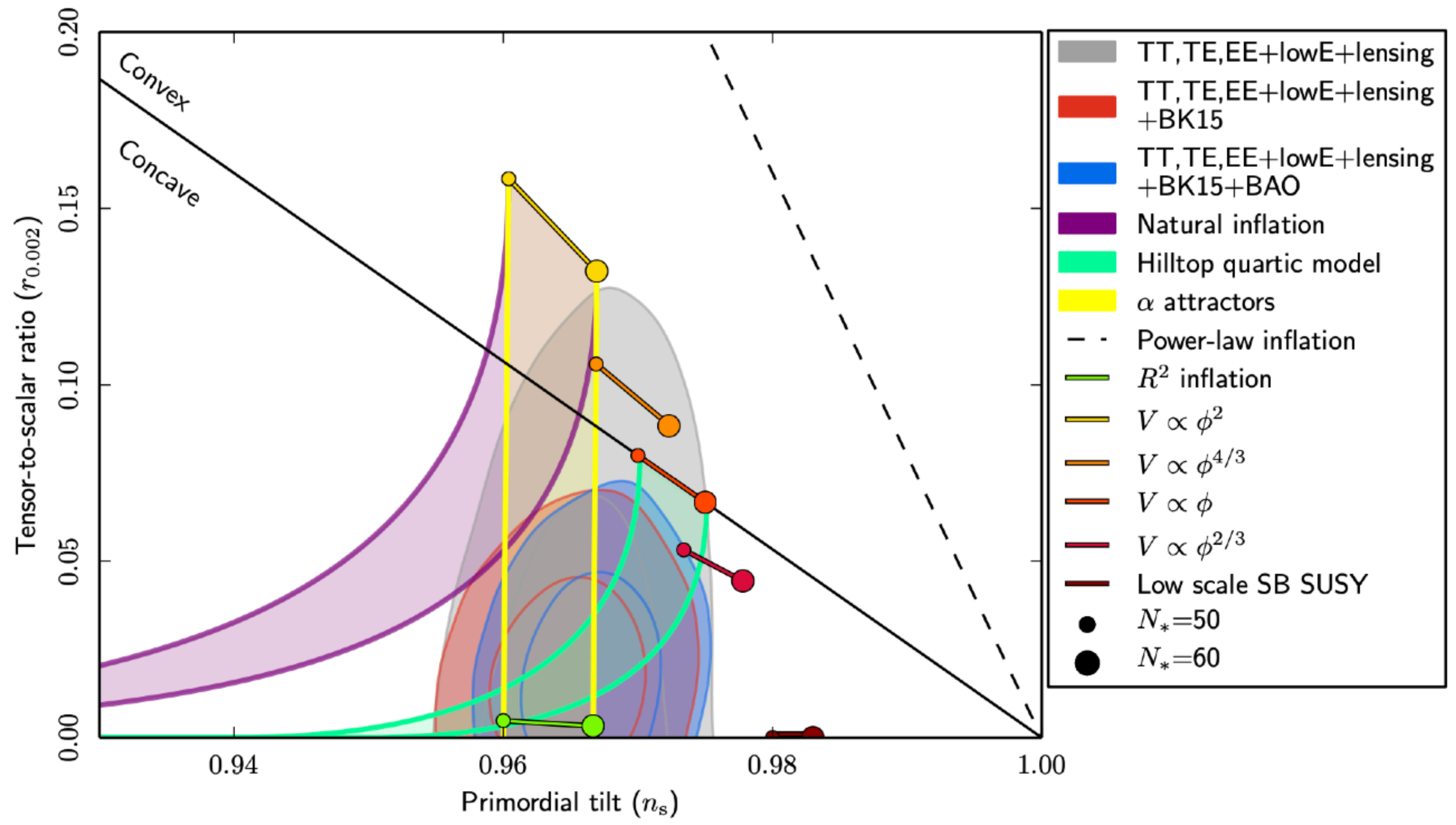
Planck 2018 results

Bayesian Inference

The different levels of inference

2. Model selection — e.g. which inflationary model is favoured

$$Z = P(D|M) = \int \underbrace{P(D|M,\theta)P(\theta|M)}_{\text{Likelihood x Prior}}\,\mathrm{d}\theta$$



Bayesian evidence (Z) ratio

Planck Collaboration: Constraints on Inflation

Inflationary model	Potential $V(\phi)$	Parameter range	$\Delta\chi^2$	$\ln B$
$R + R^2/(6M^2)$	$\Lambda^4\left(1 - e^{-\sqrt{2/3}\phi/M_{\rm Pl}}\right)^2$	...	...	...
Power-law potential	$\lambda M_{\rm Pl}^{10/3}\phi^{2/3}$	...	4.0	−4.6
Power-law potential	$\lambda M_{\rm Pl}^3\phi$	...	6.8	−3.9
Power-law potential	$\lambda M_{\rm Pl}^{8/3}\phi^{4/3}$	...	12.0	−6.4
Power-law potential	$\lambda M_{\rm Pl}^2\phi^2$	...	21.6	−11.5
Power-law potential	$\lambda M_{\rm Pl}\phi^3$	...	44.7	−13.2
Power-law potential	$\lambda\phi^4$	...	75.3	−56.0
Non-minimal coupling	$\lambda^4\phi^4 + \xi\phi^2 R/2$	$-4 < \log_{10}\xi < 4$	0.4	−2.4
Natural inflation	$\Lambda^4\left[1 + \cos(\phi/f)\right]$	$0.3 < \log_{10}(f/M_{\rm Pl}) < 2.5$	9.9	−6.6
Hilltop quadratic model	$\Lambda^4\left(1 - \phi^2/\mu_2^2 + \ldots\right)$	$0.3 < \log_{10}(\mu_2/M_{\rm Pl}) < 4.85$	1.3	−2.0
Hilltop quartic model	$\Lambda^4\left(1 - \phi^4/\mu_4^4 + \ldots\right)$	$-2 < \log_{10}(\mu_4/M_{\rm Pl}) < 2$	−0.3	−1.4
D-brane inflation ($p = 2$)	$\Lambda^4\left(1 - \mu_{D2}^2/\phi^p + \ldots\right)$	$-6 < \log_{10}(\mu_{D2}/M_{\rm Pl}) < 0.3$	−2.0	0.6
D-brane inflation ($p = 4$)	$\Lambda^4\left(1 - \mu_{D4}^4/\phi^p + \ldots\right)$	$-6 < \log_{10}(\mu_{D4}/M_{\rm Pl}) < 0.3$	−3.5	−0.4
Potential with exponential tails	$\Lambda^4\left[1 - \exp(-q\phi/M_{\rm Pl}) + \ldots\right]$	$-3 < \log_{10} q < 3$	−0.4	−1.0
Spontaneously broken SUSY	$\Lambda^4\left[1 + \alpha_h \log(\phi/M_{\rm Pl}) + \ldots\right]$	$-2.5 < \log_{10}\alpha_h < 1$	6.7	−6.8

$$\frac{Z_1}{Z_2} \propto \frac{P(M_1|D)}{P(M_2|D)}$$

Bayesian Optimisation

When to use it?

Standard methods (MCMC, Nested Sampling) work well for the most part, which is why they are widely used!

However, these methods are based on random sampling and typically require $N > 10^4$ likelihood evaluations for MCMC, even more for Nested Sampling

This can be an issue if the likelihood is expensive to evaluate (high precision calculations, numerical simulations…)

Bayesian Optimisation

The Algorithm

We want to replace the evaluation of the actual slow likelihood with a quick to evaluate surrogate.

Step 1

Guess shape of unknown function from observed function values x_i, y_i using Gaussian Process Regression

Step 2

Find next best point to evaluate (observation) using an acquisition function

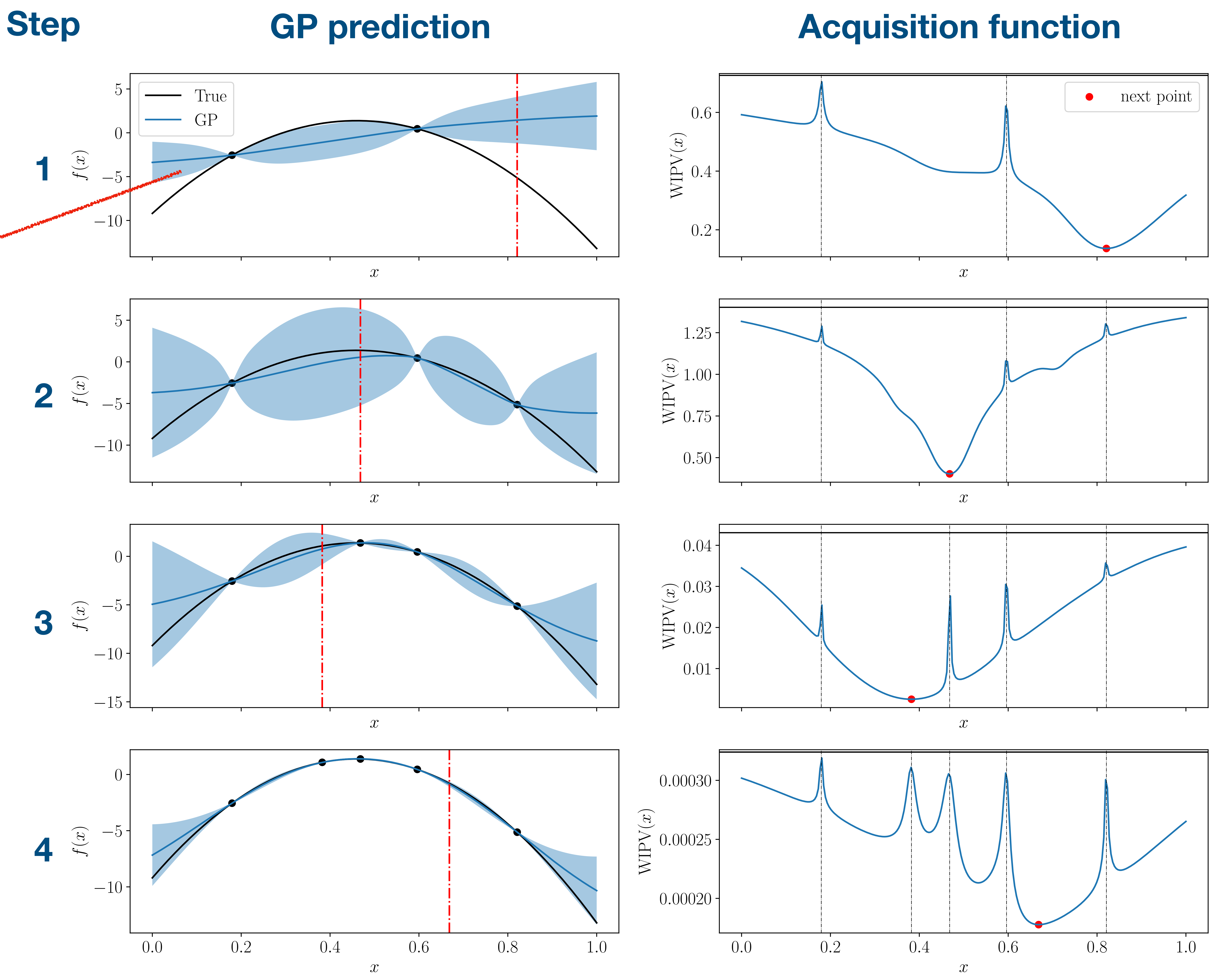
Repeat until convergence

Bayesian Optimisation in action

GP variance $\longrightarrow \Delta \log Z$

Use $\Delta \log Z$ for convergence check

Once trained, can run MCMC/HMC or Nested Sampler using GP surrogate.



Bayesian Optimisation

Pros and Cons

Converges to true likelihood shape (or maximum) in far fewer evaluations (10-100x) compared to traditional methods — uses all available information to obtain new points

Works for arbitrary shaped likelihoods (multimodal, non-Gaussian…)

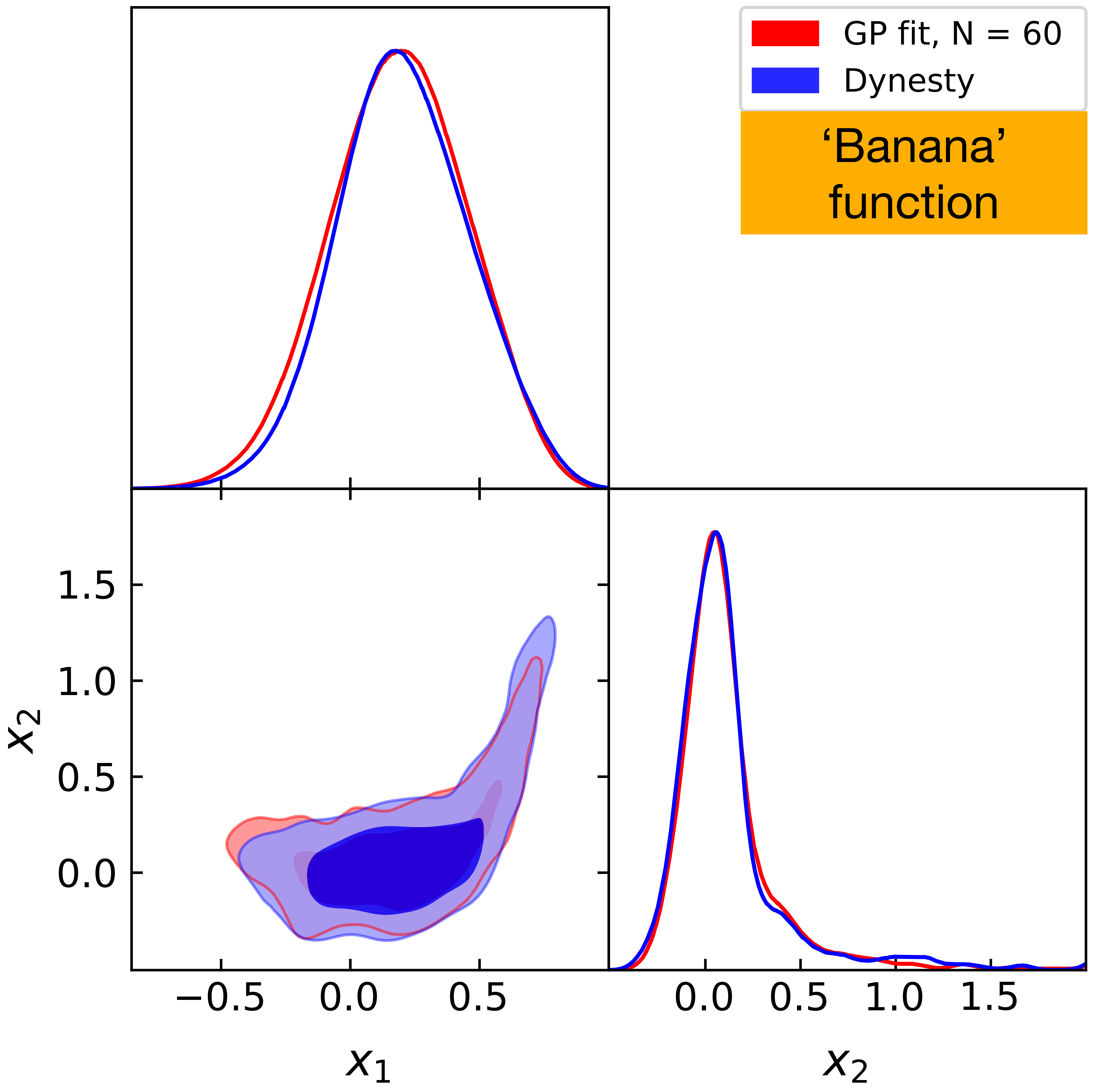
However, this doesn't come for free…

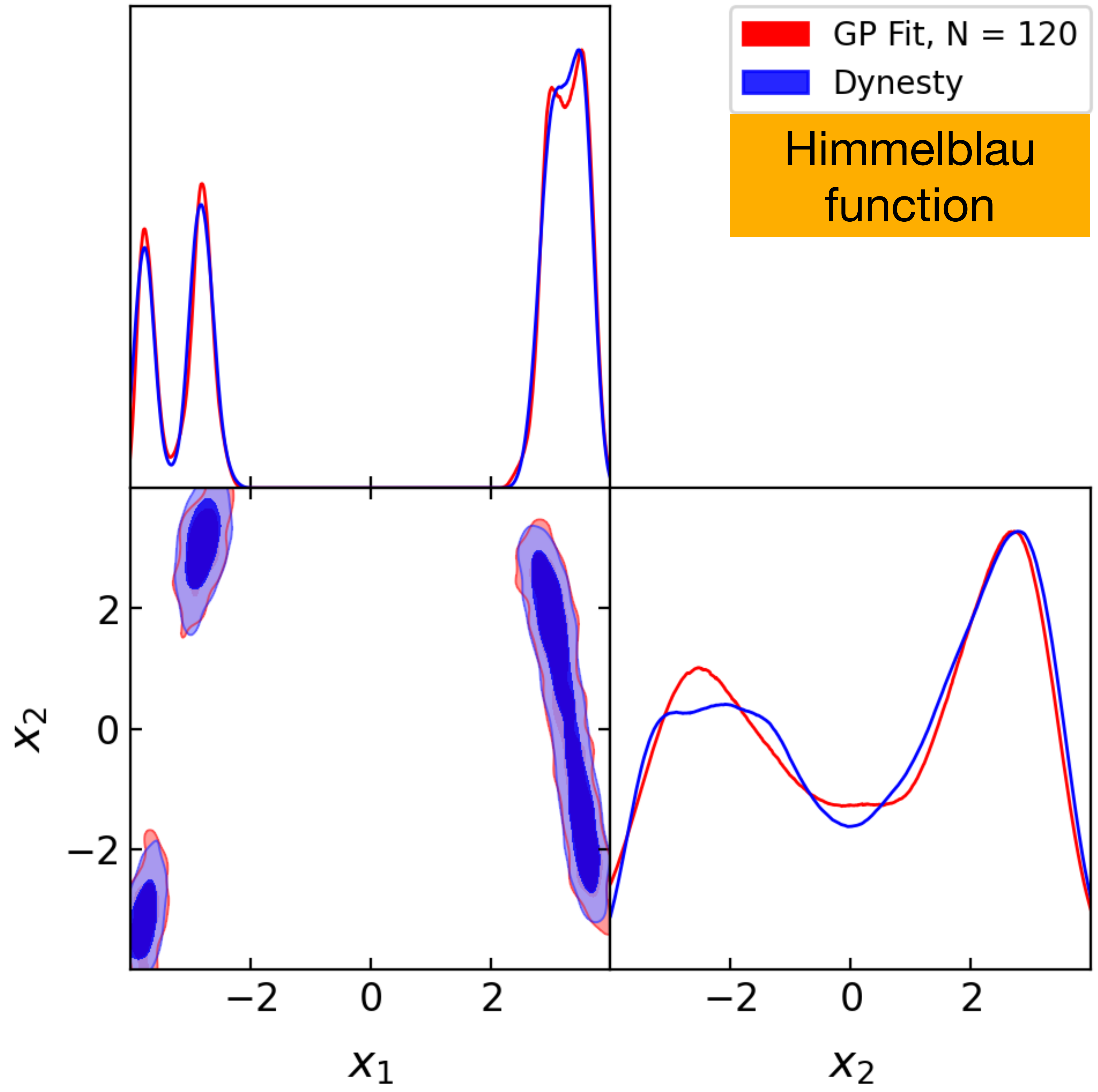
Greater computational overhead

Unfavourable scaling with dimensions, may be limited to $D \lesssim 15 - 20$, unlike MCMC or nested sampling

Bayesian Optimisation

Some examples





Bayesian Optimisation

Cosmology from CMB

Runtime < 15 min on laptop with ~ 200 likelihood evaluations — 6D Planck lite (no nuisance params)

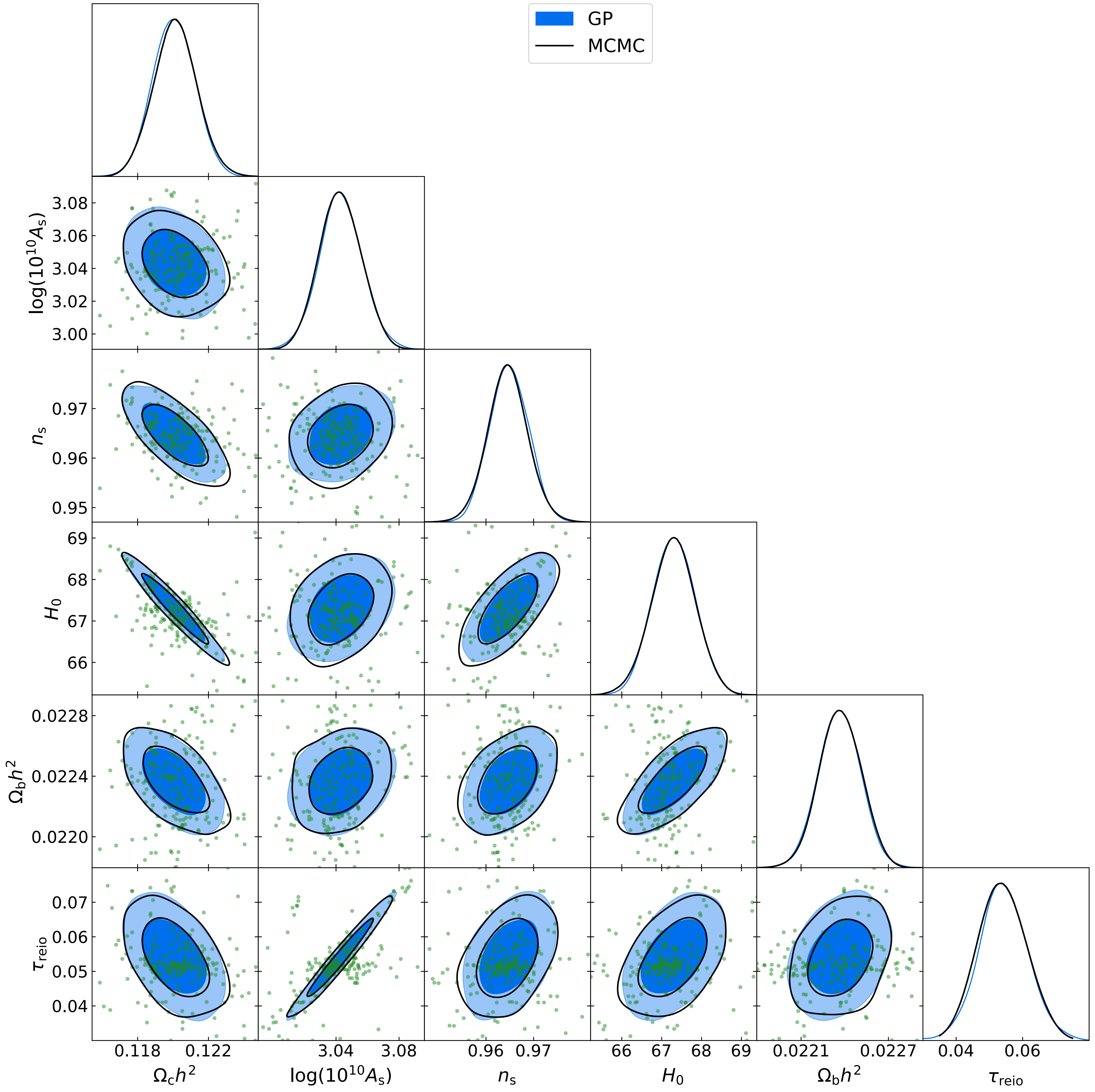
Evidence estimate from GP

```
[NS]:Final LogZ info : mean: = -502.2473,
upper: = -501.4937, lower: = -502.9939,
dlogz sampler: = 0.0884
```

Estimate from Nested Sampler

```
log(Z) = -502.65 +/-0.32 |
```

MCMC ~ 10000 likelihood evals and Nested Sampler ~ 40000



Bayesian Optimisation

Dynamical DE from CMB + BAO + SN

6 LCDM + Ω_K + $w_0 w_a$

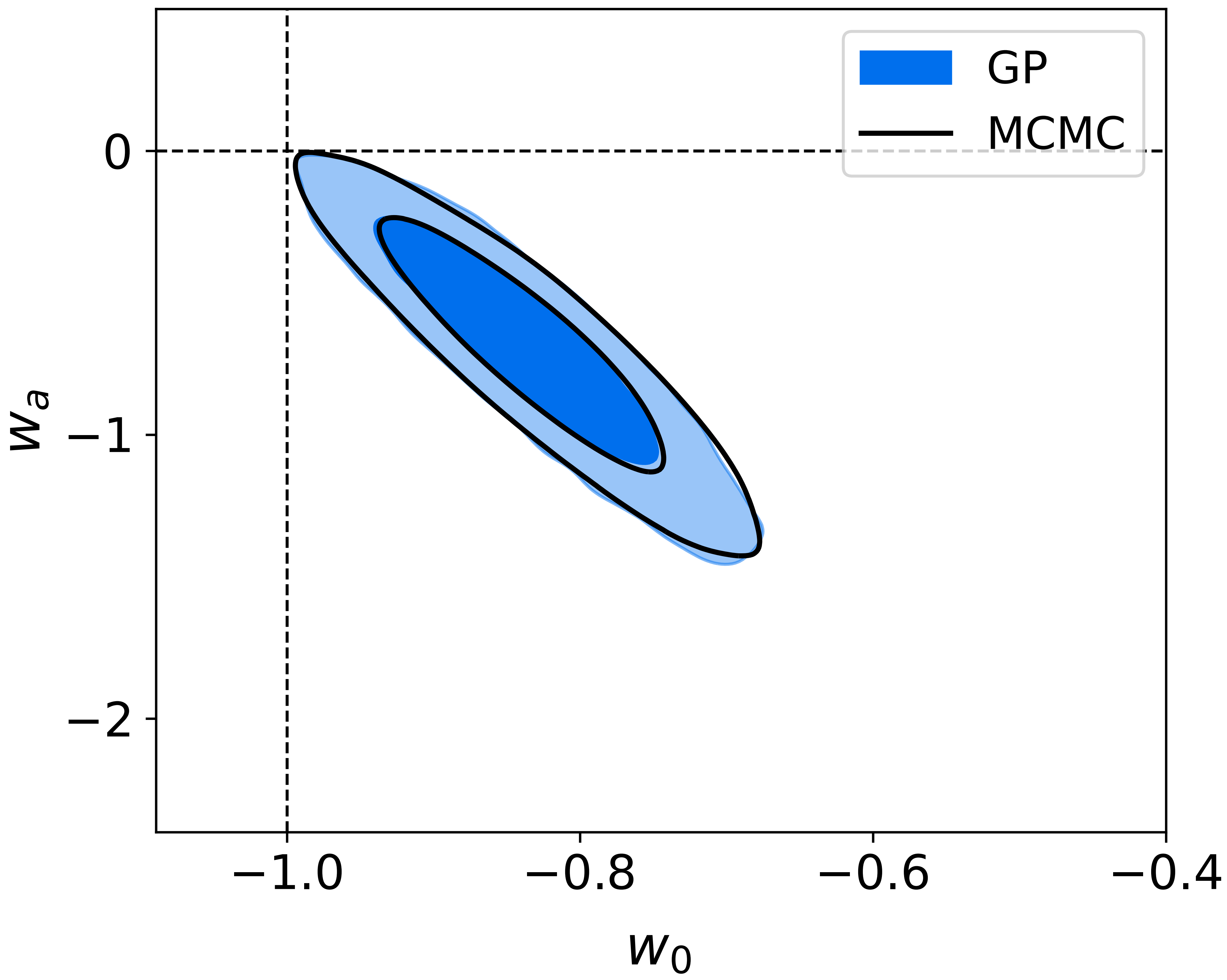
Planck lite + BAO + PantheonPlus with ~ 800 evaluations, runtime ~ 6 hours. With curvature, each likelihood eval takes ~2-3 seconds.

Evidence estimate from GP

```
Final LogZ info: mean = -1232.9086,
upper=-1228.3018, lower=-1233.1163,
dlogz sampler=0.1282,
```

Estimate from Nested Sampler

```
log(Z) = -1232.6624 +/-0.41
```



Bayesian Optimisation

Cosmology from CMB

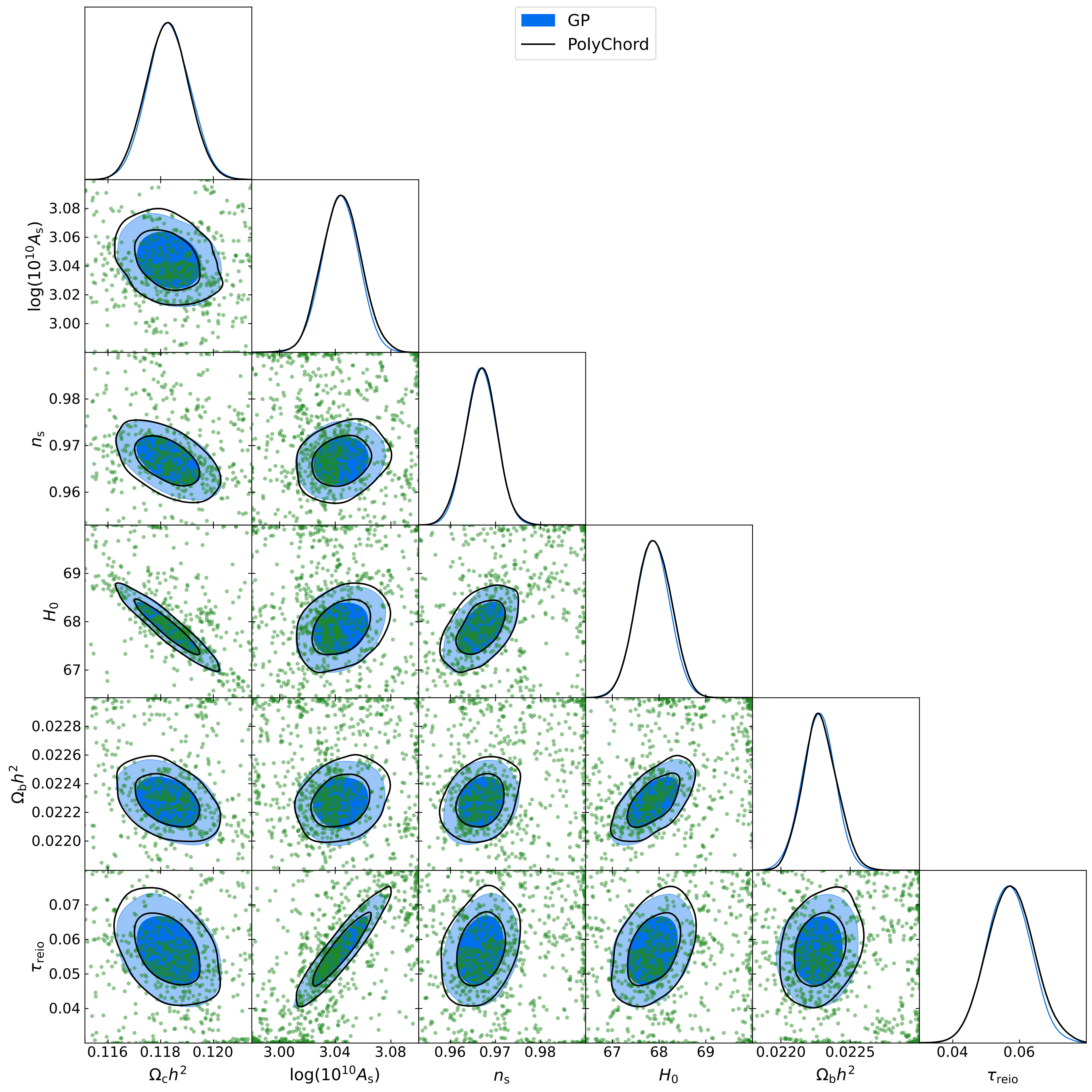
6 LCDM + 9 nuisance parameters, `Planck CamSpec` likelihood with ~ 1000 evaluations, runtime ~ 12 hours

Evidence estimate from GP

```
Final LogZ info: mean = -5529.4664,
upper = -5527.8009, lower = -5529.9712,
dlogz sampler = 0.1680,
```

Estimate from Nested Sampler (runtime >1 day)

```
 log(Z) = -5529.6521 +/-0.45
```



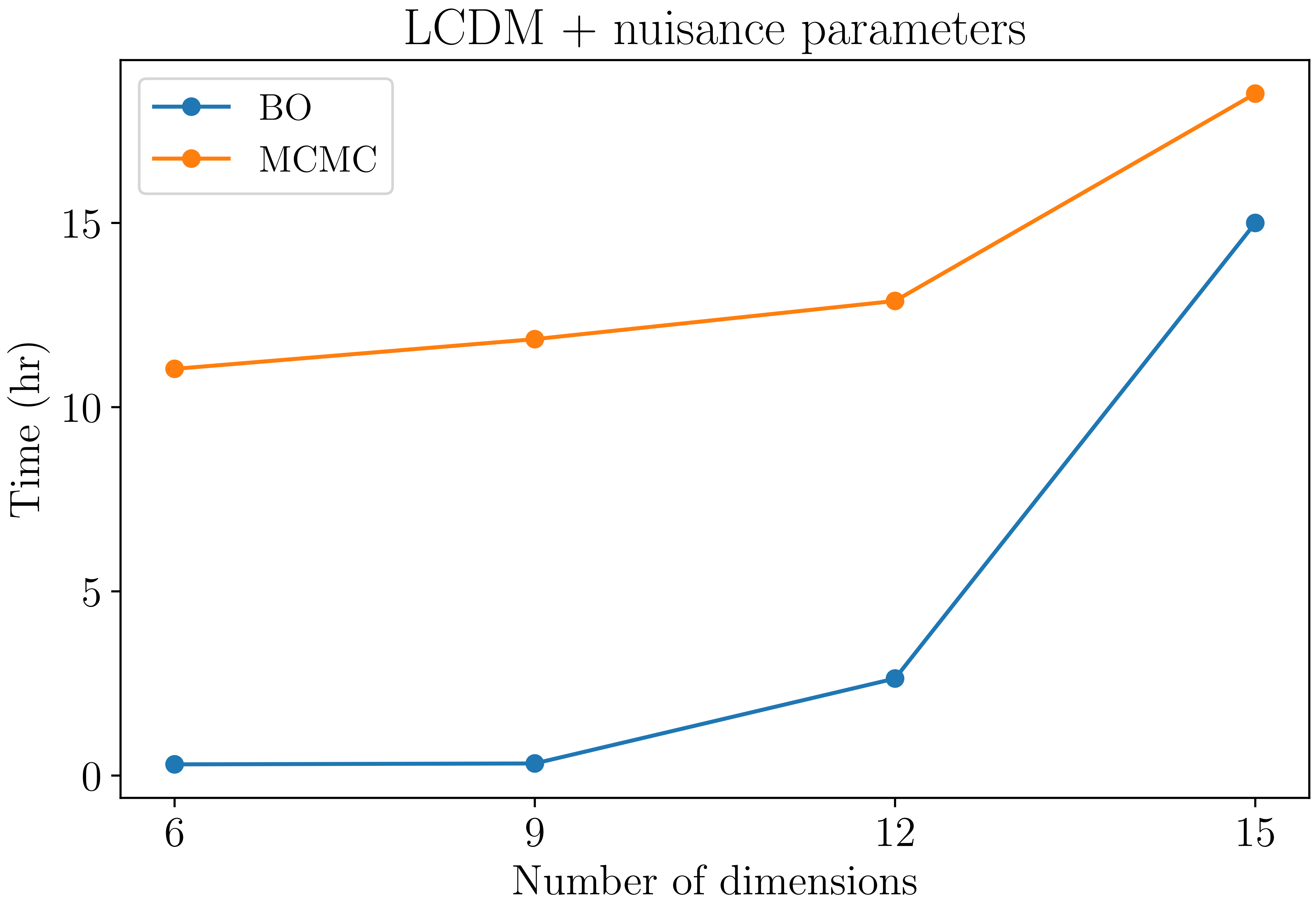
Bayesian Optimisation

Comparison vs traditional methods

Faster for slow likelihoods but runtime increases more steeply…

MCMC runtime dominated by likelihood evaluation time, BO by the other steps

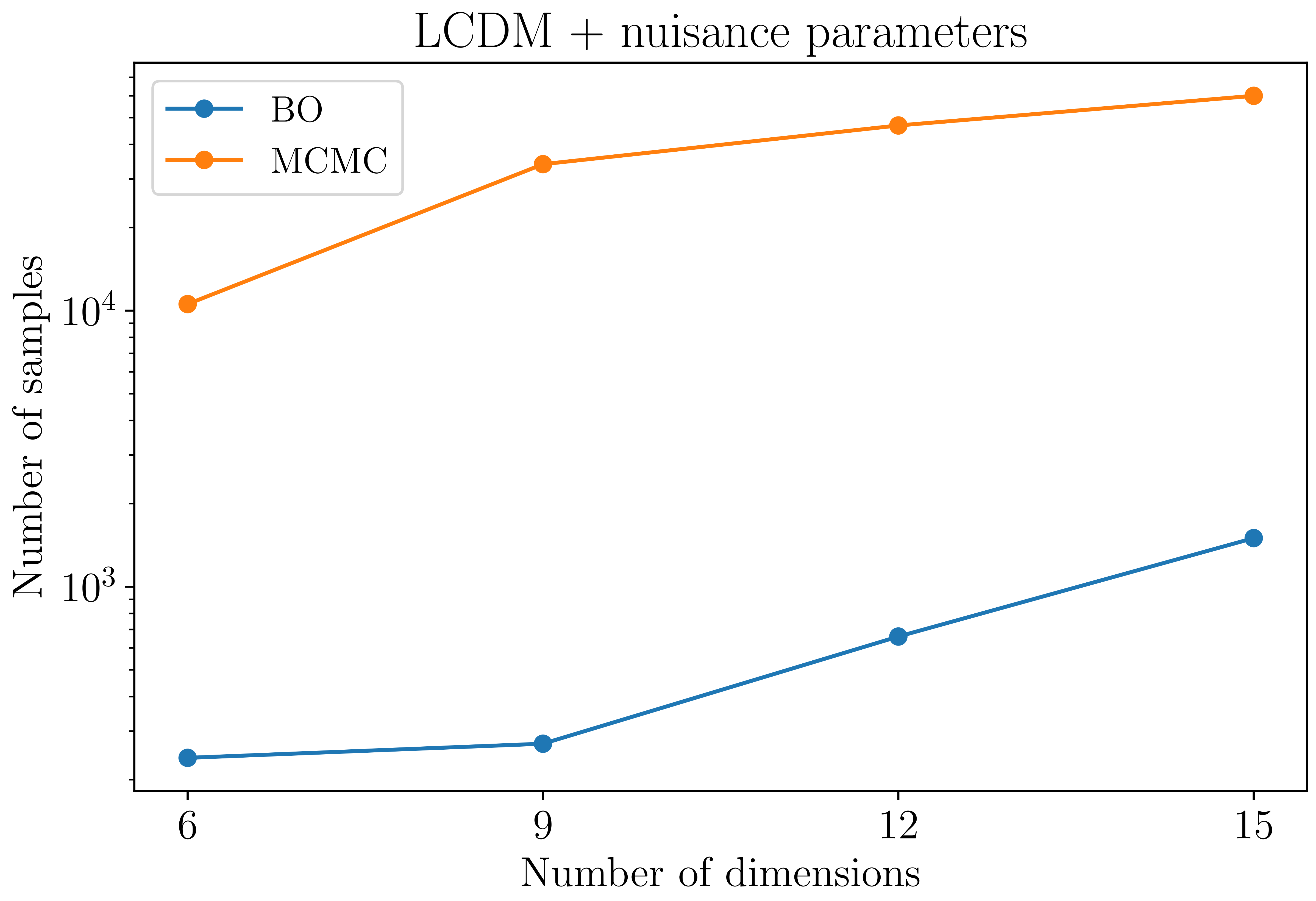
~1s likelihood evaluation time



Bayesian Optimisation

Comparison vs traditional methods

We need about 10–100x fewer likelihood evaluations throughout!



Bayesian Optimisation

Summary

Gaussian Process regression + active acquisition strategy can be used to perform Bayesian inference for cosmology

High efficiency of the algorithm compared to traditional methods based on random sampling

Can obtain parameter posteriors + evidence for **complicated distributions** and **computationally expensive likelihoods**

Stay tuned for paper and code (JAX based) + more applications to cosmology

Thank you!

Contact: ameek.malhotra@swansea.ac.uk