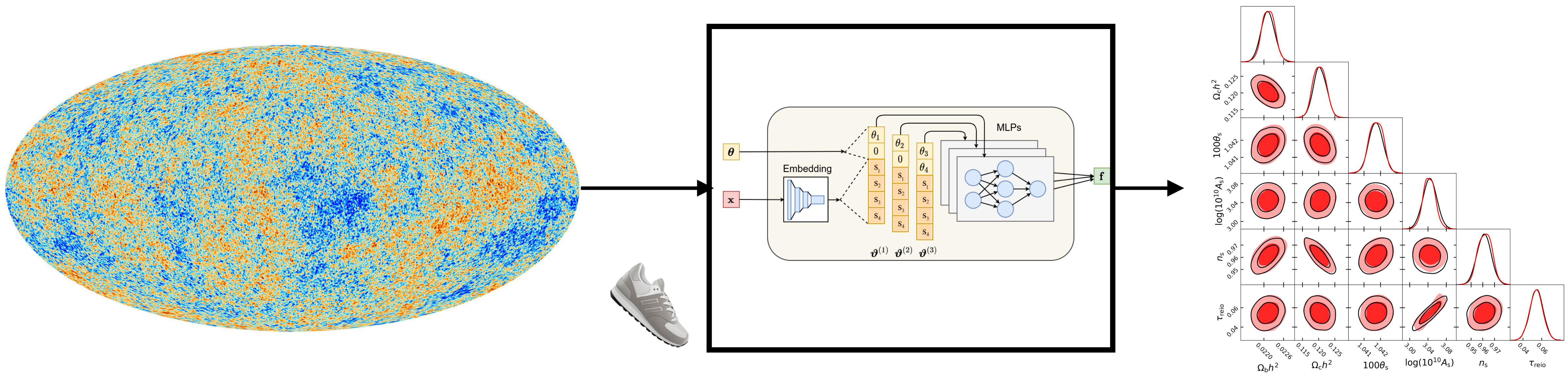


Cosmological applications of

Truncated Marginal Neural Ratio Estimation

Alex Cole (U. of Amsterdam)

11 May 2022

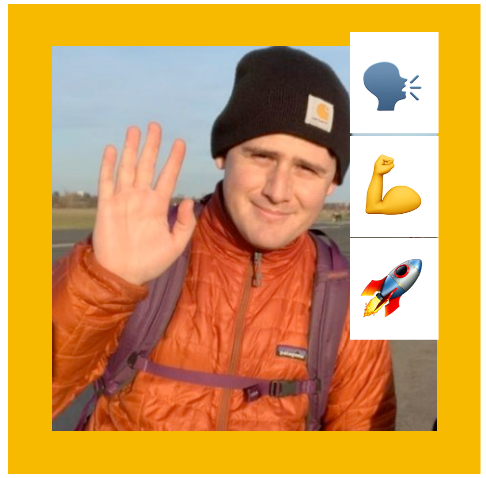
5th IML Workshop @ CERN

a.e.cole@uva.nl

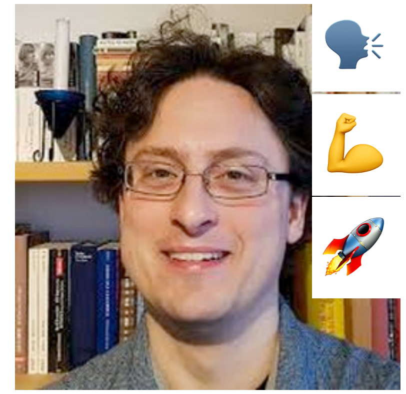
@a_e_cole

primarily based on [2111.08030]

GRAPPA



Ben Miller



Christoph Weniger

Sam Witte

[Miller et al. ’20]

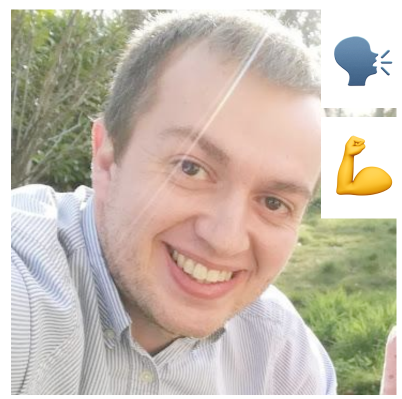
🗣 2011.13591 (NeurIPS ’20 ML4PS)

💪 2107.01214 **[Miller et al. ’21]** (NeurIPS ’21 conference)

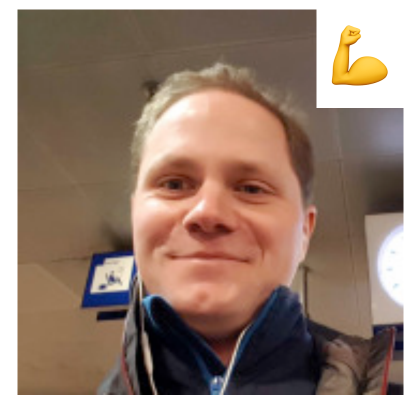
🚀 2111.08030 (applications to cosmology)

[AC et al. ’21]

U. Liège



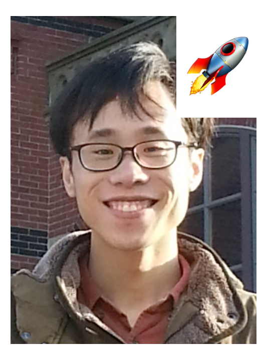
Gilles Louppe



Patrick Forré

UvA AMLab

Netherlands eScience center/SURF



Maxwell Cai

Meiert Grootes

Francesco Nattino

- We just heard from Ben about TMNRE, and the results seems nice! But before we grab our wallets:
    - let's examine performance, nice features, etc. in a **cosmological context.**

Outline

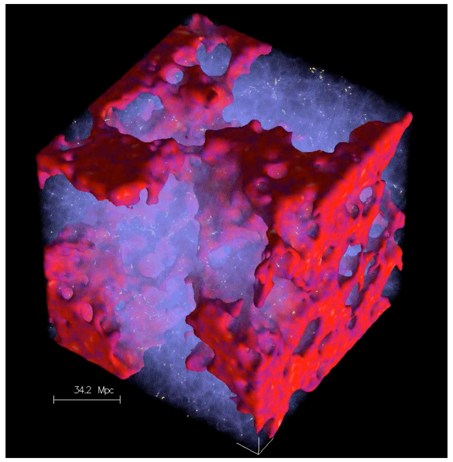
1. Motivation
2. Our playground: CMB power spectra, results
3. Discussion

1. Motivation

Problem 1

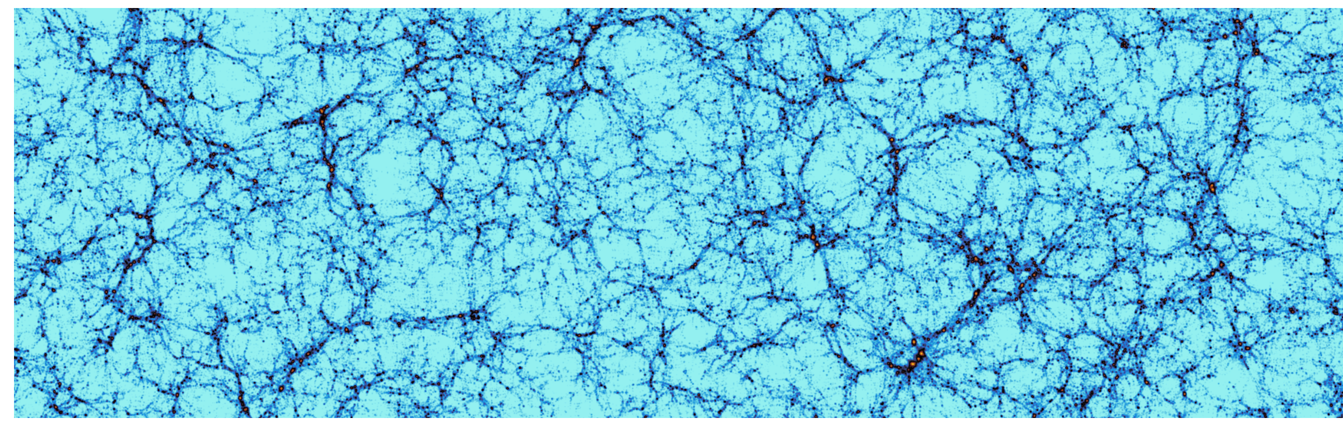
- For most simulators, we **cannot evaluate the full likelihood**.
  - In cosmology: large-scale structure, 21-cm field, most late-time observations…
- Practitioners often **restrict** to theoretically controlled summary statistics such as the power spectrum at large scales.
  - We should worry that we're throwing the baby out with the bathwater.

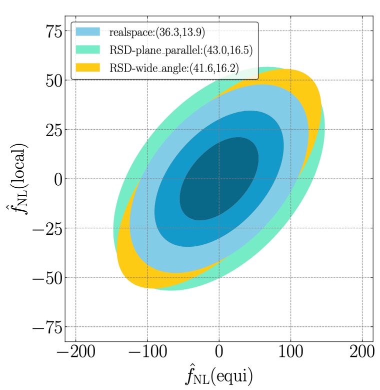
21cm field,
[SKA white paper 1210.0197]



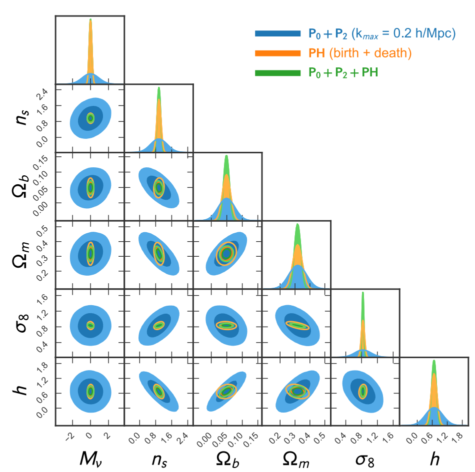
These problems clearly **demand more refined summary statistics**. One option is hand-crafted summaries, e.g. persistent homology for large-scale structure, whose **likelihoods can be approximated**. Would prefer more knobs to optimize, theoretical guarantees about saturating information content.

[Biagetti, AC, Shiu (JCAP) ’20;
AC, Biagetti, Shiu (NeurIPS wksp ’20)]



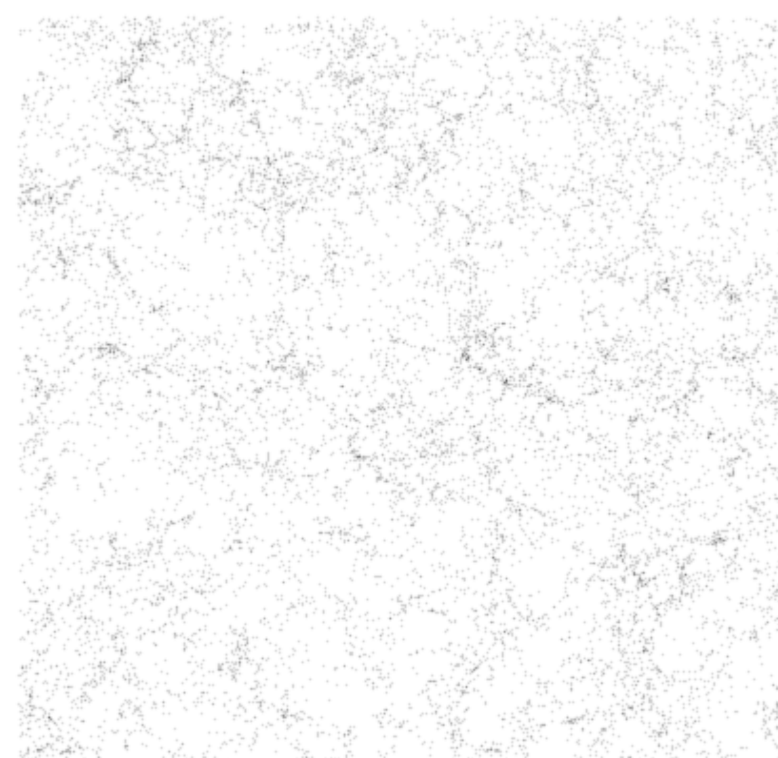


[Equilateral NG, 2203.08262]



[ΛCDM, to appear]

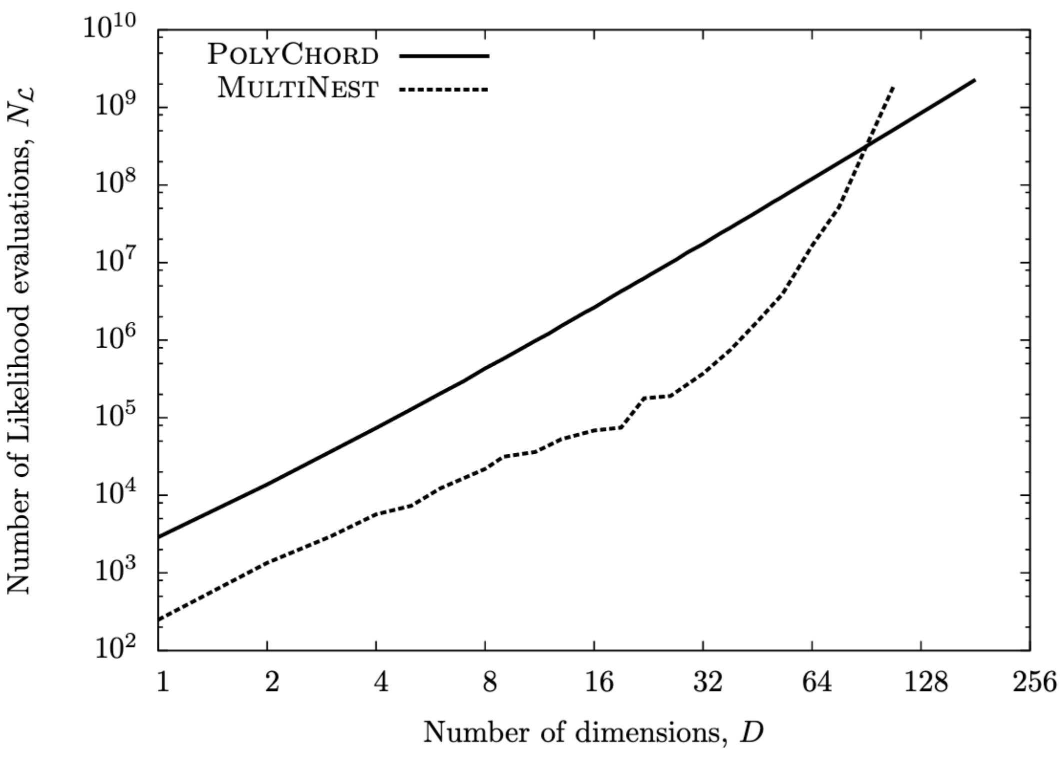
k=20, p=10, q=1, $\nu = +0.00$



Problem 2

- Even if likelihood is known/ tractable:
  - For realistic inference, one must vary over instrumental calibration parameters, foreground residuals, latent variables … ≡ **nuisance parameters**
  - **Sampling the joint** posterior scales poorly with parameter space dimension.

classical inference cost w/ dimension



[Handley et al. 1506.00171]

Problem 2

- Decompose $\theta = (\vartheta, \eta) = (\text{relevant}, \text{nuisance})$

marginal posterior

ϑ: **relevant** parameters

likelihood

η: nuisance parameters

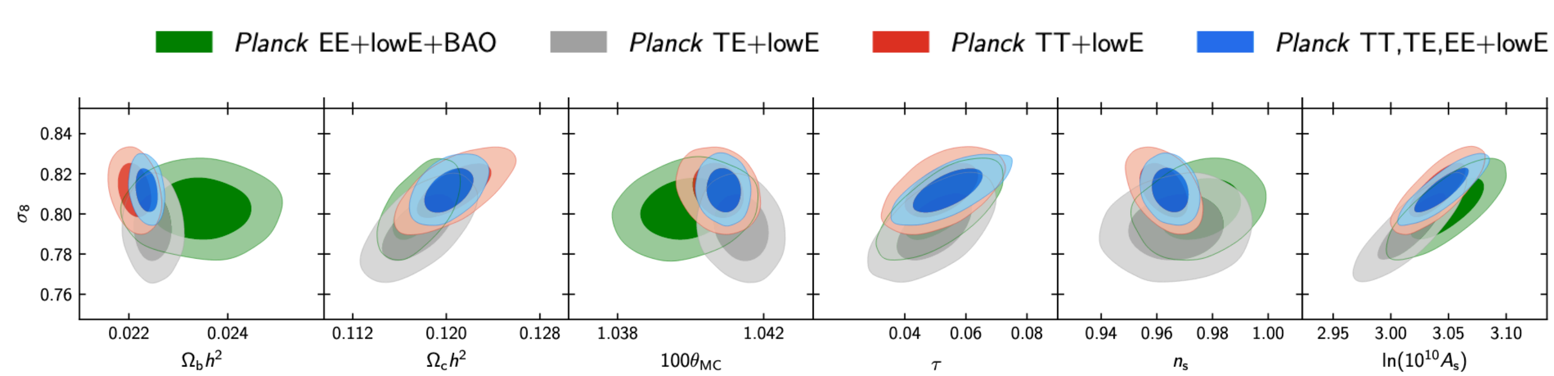
prior

$$p(\vartheta \mid x) = \frac{\int p\left(x \mid \vartheta, \eta\right) p(\eta) d\eta \; p(\vartheta)}{p(x)}$$

x: data

evidence

- *Many* scientific insights are derived from plots where ϑ is 0- or 1-dimensional.



Planck 2018

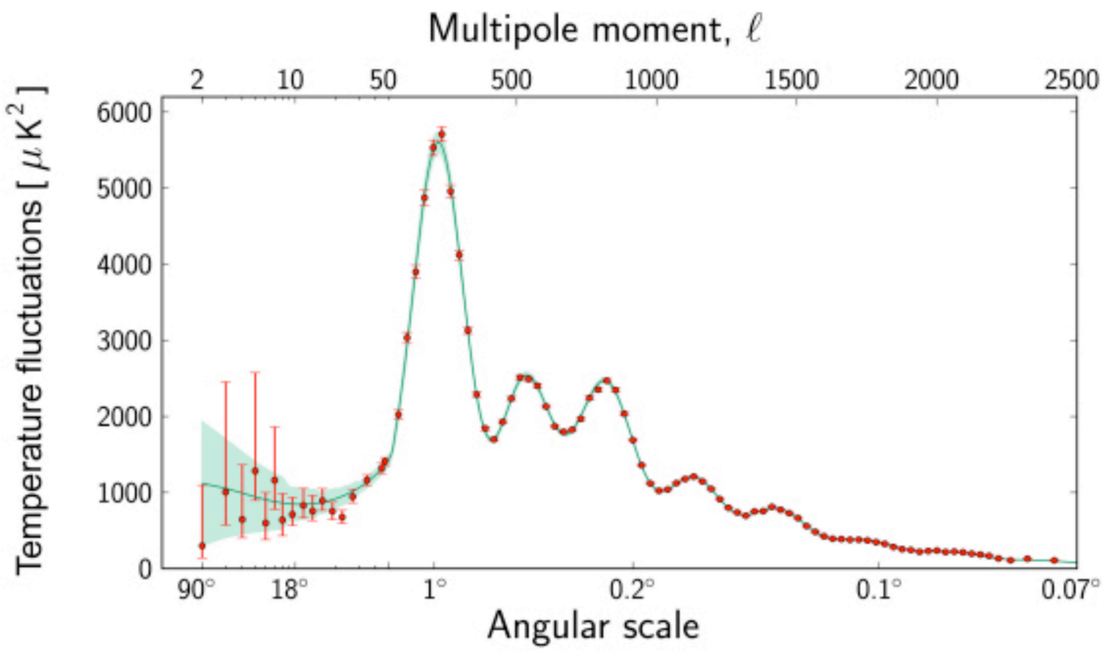
Vapnik's principle: "When solving a problem of interest, do not solve a more general problem as an intermediate step."

Motivates both NRE and directly targeting marginals.

- Trust, but verify.
  - We'd especially like to ***avoid overconfidence***. False detections/etc. are embarrassing!
  - Need tools to rigorously assess consistency of results!

2. CMB Power Spectra

CMB forecasting

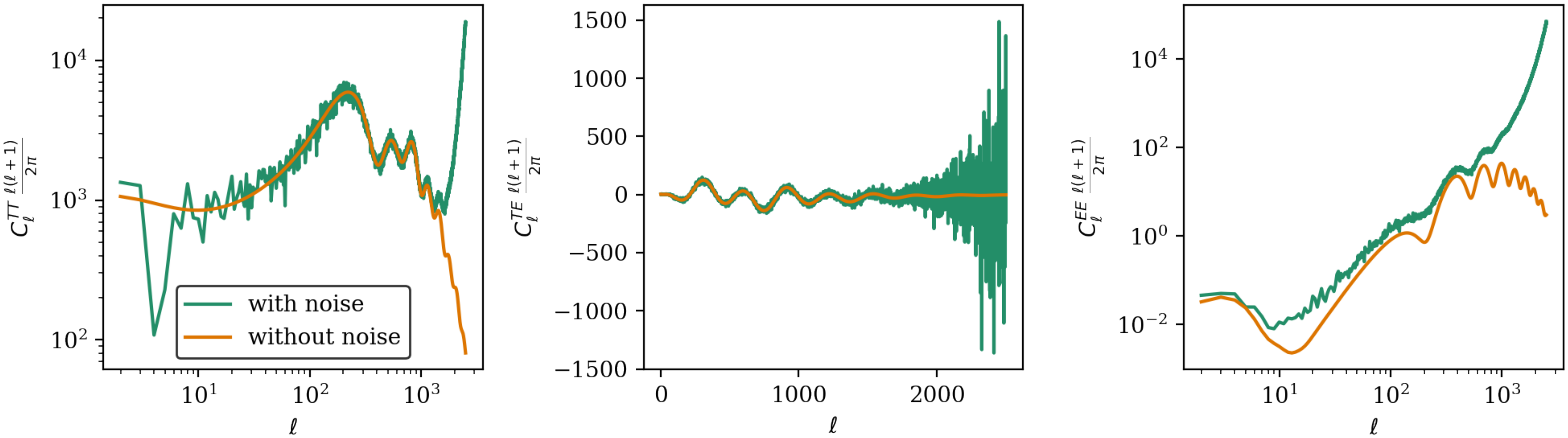


- A large fraction of experimental cosmology constraints come from the CMB power spectrum. A simulator (cf. **Likelihood_mock_cmb** in **monte_python**):

  1. Given cosmology, compute $C_\ell^{PP'}$ from Boltzmann code **vanilla ~ 1s** **“exotic” -> 30 min**
  2. Add instrumental noise $\overline{C}_\ell^{PP'} \equiv C_\ell^{PP'} + N_\ell^{PP'}$
  3. Sample the maximum likelihood $\hat{C}_\ell^{PP'}$, sampling full Wishart distribution at low ℓ and approximating with Gaussian at high ℓ. **[details in AC et al.]**

CMB forecasting

- A large fraction of cosmology constraints come from the CMB power spectrum. In **[AC et al. '21]** we defined a simulator for this.
- With Planck-like noise **[Di Valentino et al. '16]**, drawing from the simulator looks like:



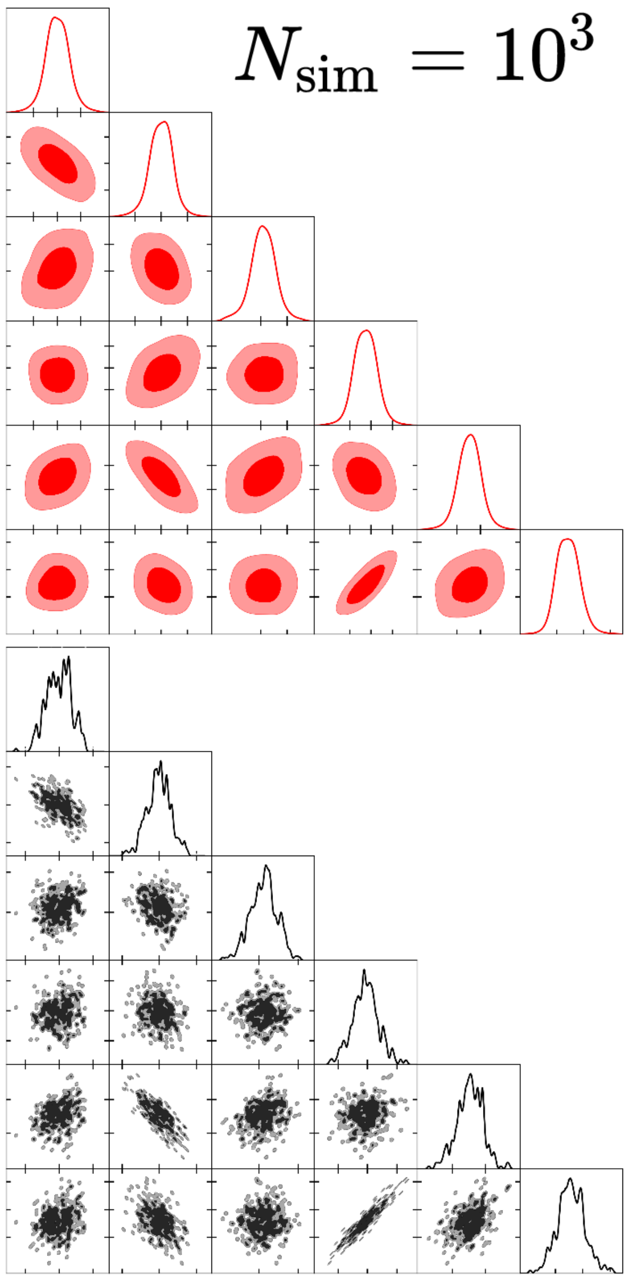
here noise means instrument contribution and cosmic variance

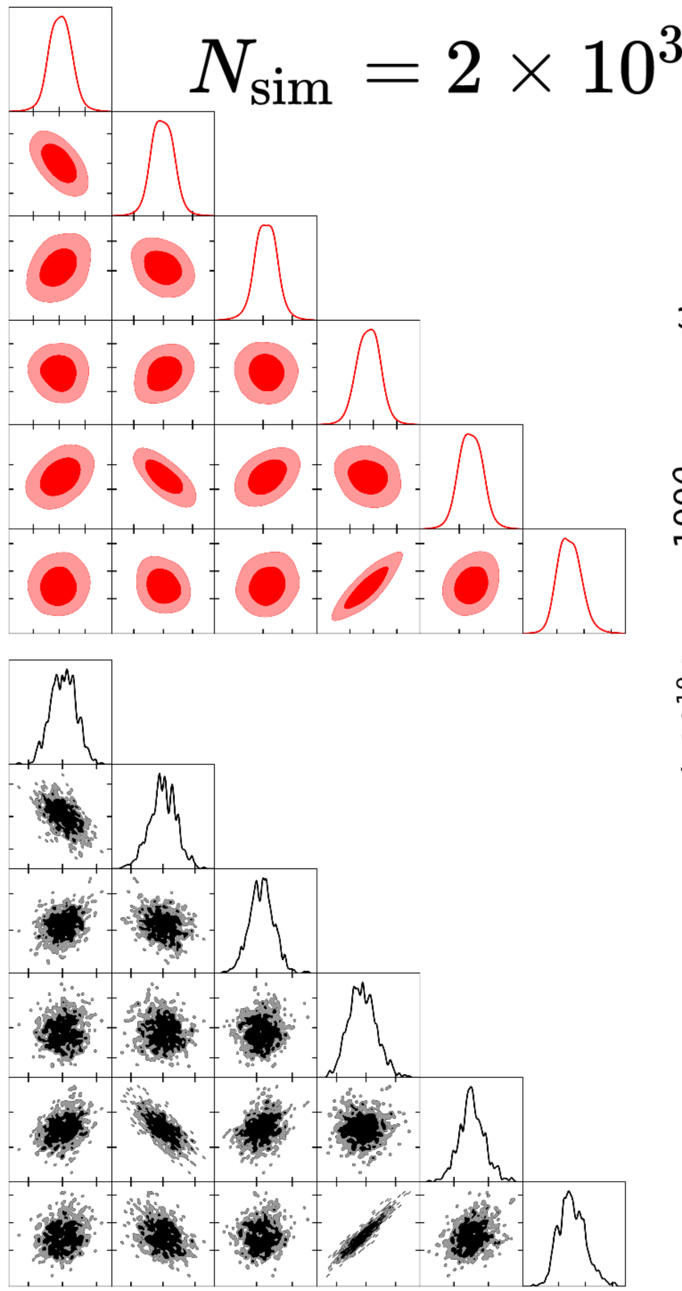
CMB forecasting

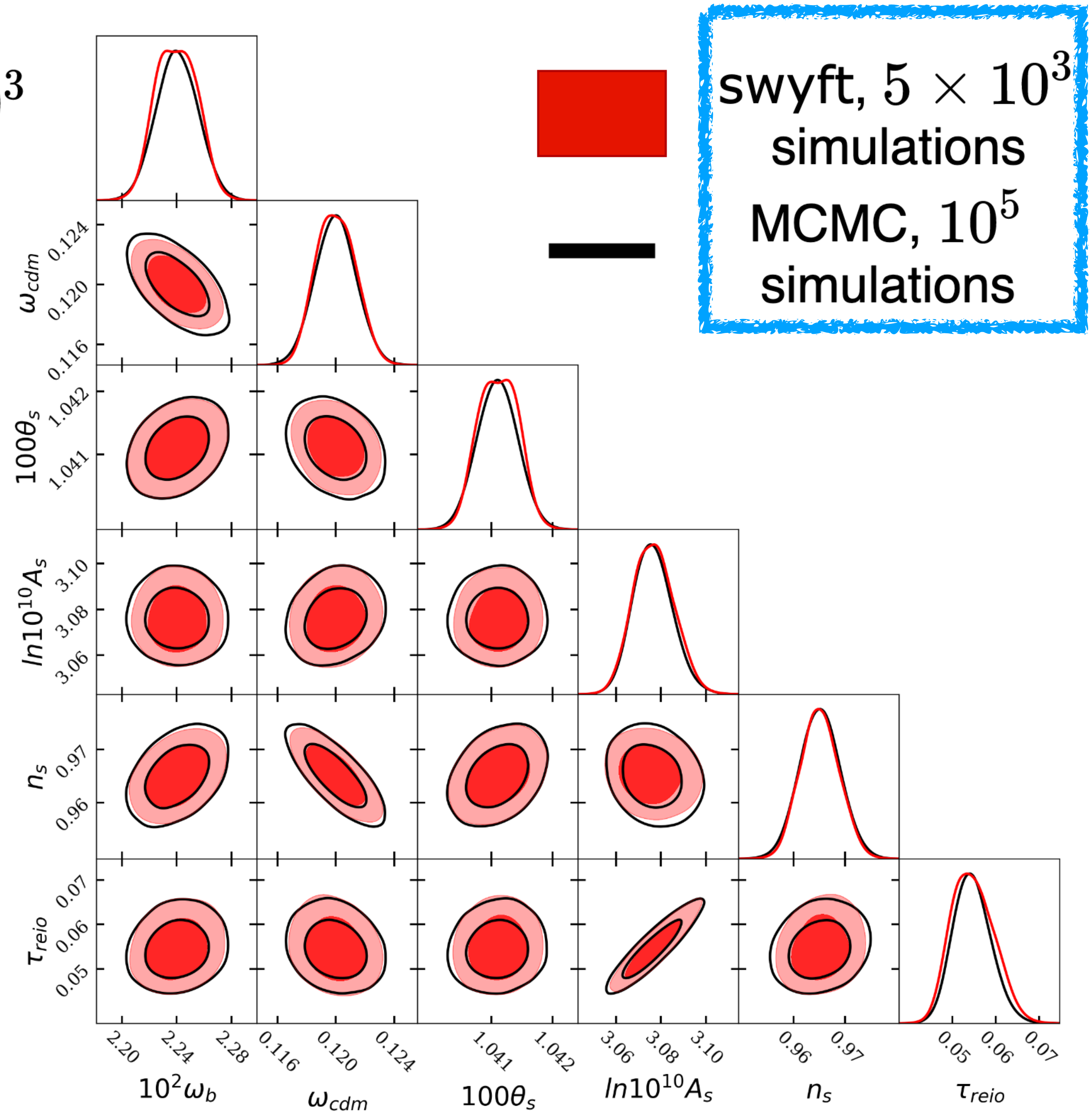
- Let's apply TMNRE to this simulator.
- There are 6 ΛCDM parameters to infer. For a prior, we use $\pm 5\sigma$ from a Fisher estimate. (i.e. truncation not necessary)
- The likelihood in this case is known, so we can compare convergence against MCMC.
- To compress the data, we use a **linear embedding network,** which compresses from 7500 to 10 features. **[cf. Tegmark,Taylor,Heavens '97; Heavens,Jimenez,Lahav '00]**

[AC et al. '21]

Convergence





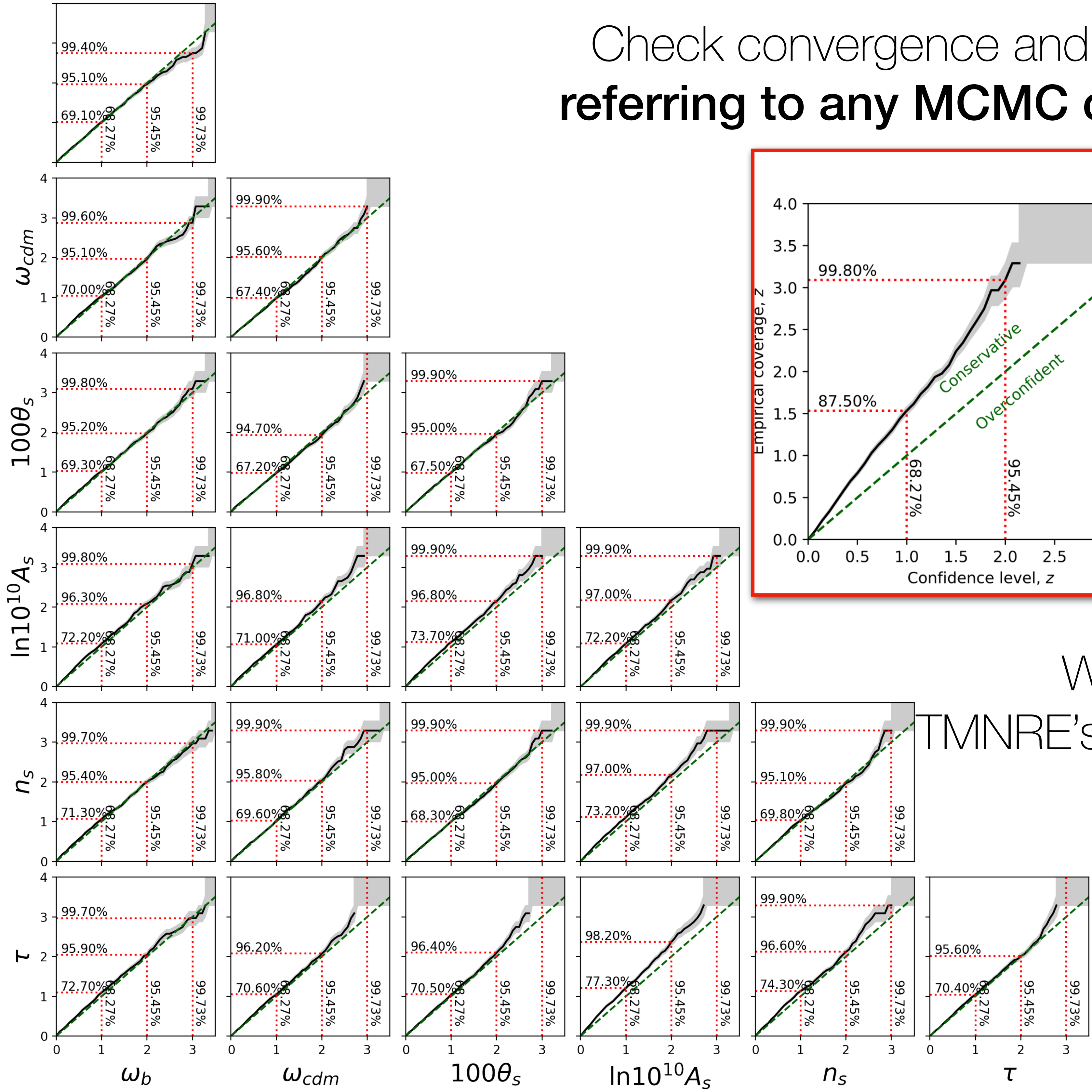


- General question: when should we trust results generated by SBI? What techniques do we have when ground-truth MCMC is not available?

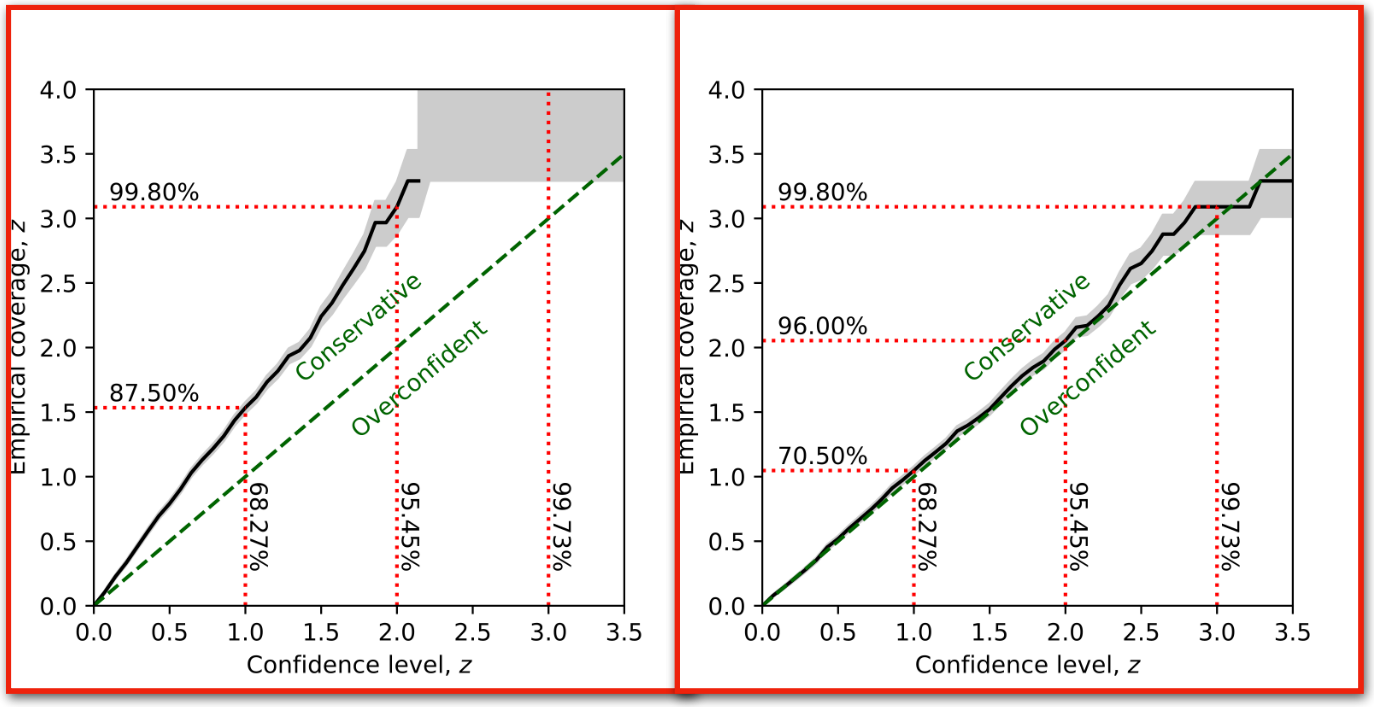
- **Really important! (S)NPE, (S)NRE, SNLE, ABC are *all* capable of overconfidence** [cf. Hermans et. al "Averting a Crisis in Simulation-Based Inference" 2110.06581]

Consistency check

[AC et al. '21]



Check convergence and consistency **without referring to any MCMC chains or likelihoods!**



Well-defined thanks to TMNRE's local amortization property.

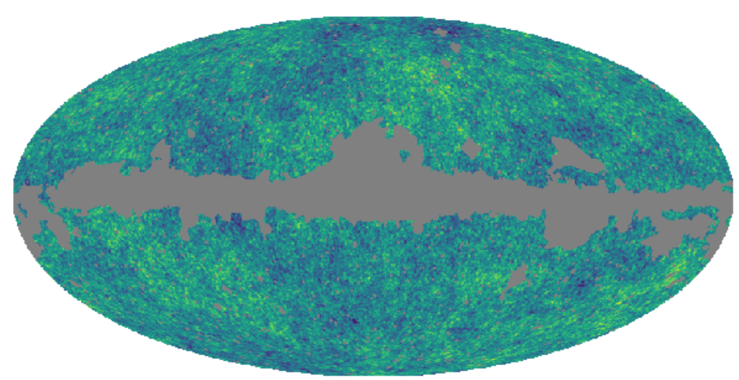
Sky Mask

[AC et al. '21]

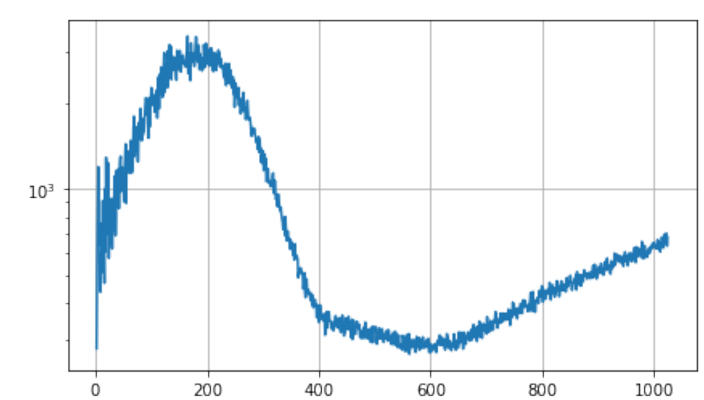
- When including a sky mask, the likelihood becomes pretty nasty. On the other hand, as a simulator the process is simple.

- Inference with MNRE is straightforward (5000 sims)

$$\Theta \rightarrow C_\ell \rightarrow$$

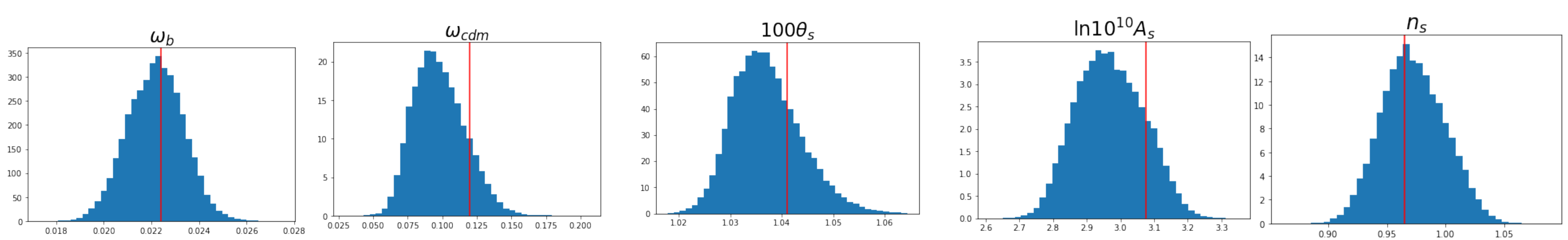


$\rightarrow$



NB: systematics
from WMAP

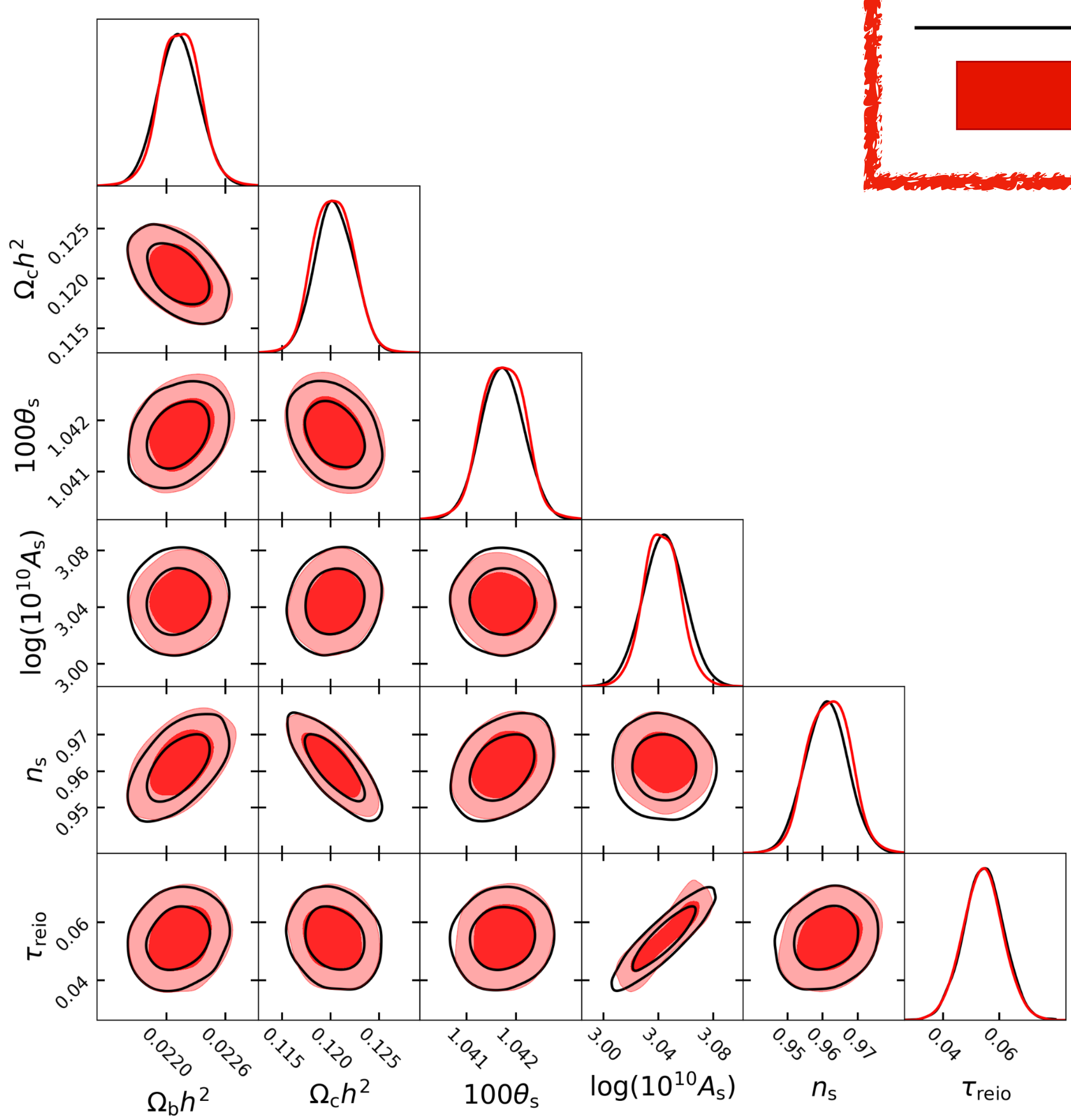
$\rightarrow$

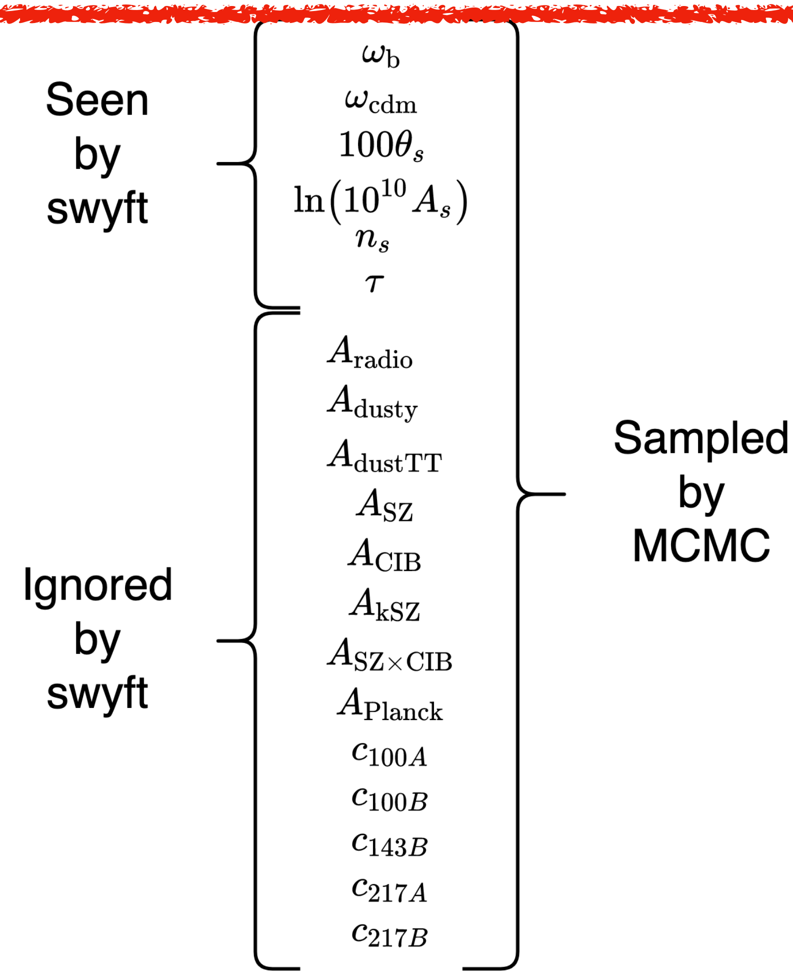


Realistic CMB

- Ramping up in realism! **HiLLiPoP** likelihood: Planck likelihood, 13 varying nuisance parameters [Couchot et al. '16]

swyft	parameters	MCMC
3×10^3 simulations		5×10^6 simulations

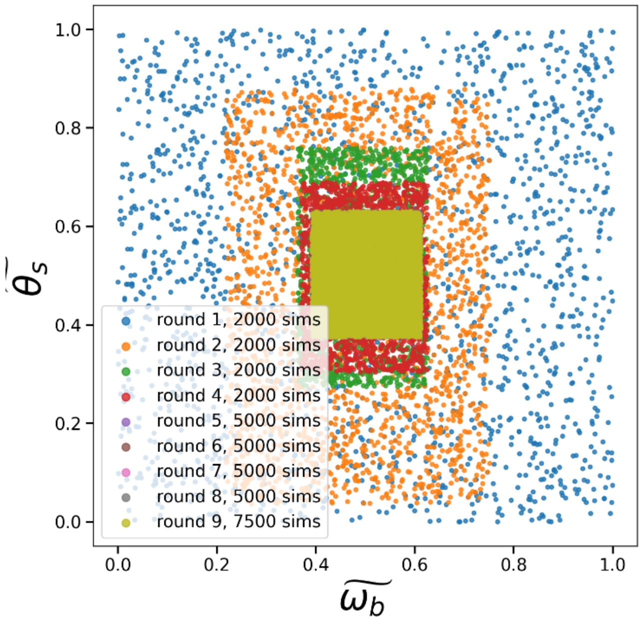


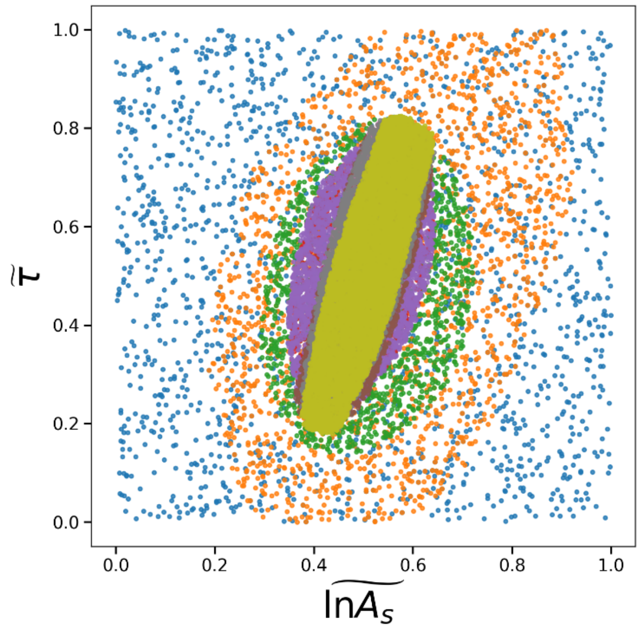


Zooming in

- Demonstration on prior that is “too big” by factor of 5 in each parameter ($8/5$ for τ) — prior volume “too big” by factor 5000

- Truncation efficiently identifies relevant region with $20{,}000$ sims over several rounds.



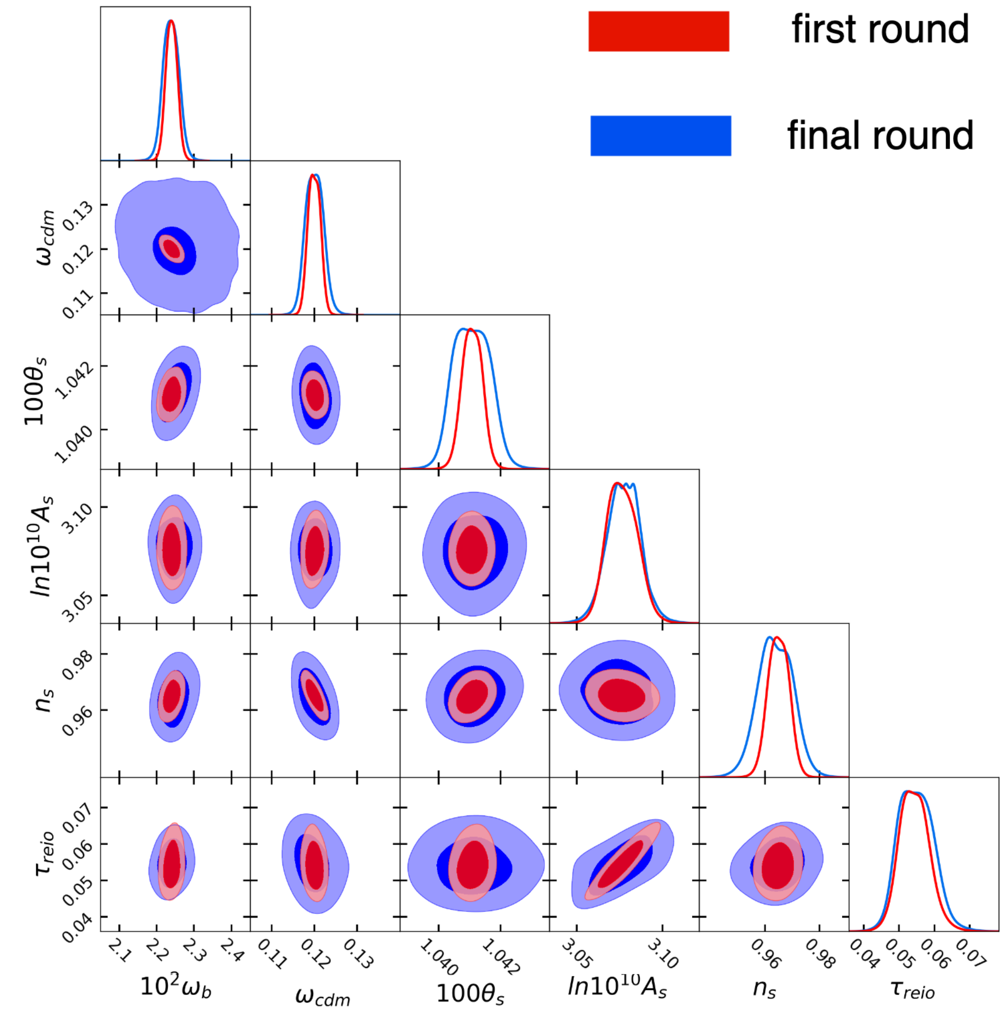


[AC et al. ’21]

Zooming in

- Demonstration on prior that is “too big” by factor of 5 in each parameter ($8/5$ for τ) — prior volume “too big” by factor 5000

- Truncation efficiently identifies relevant region with $20{,}000$ sims over several rounds.

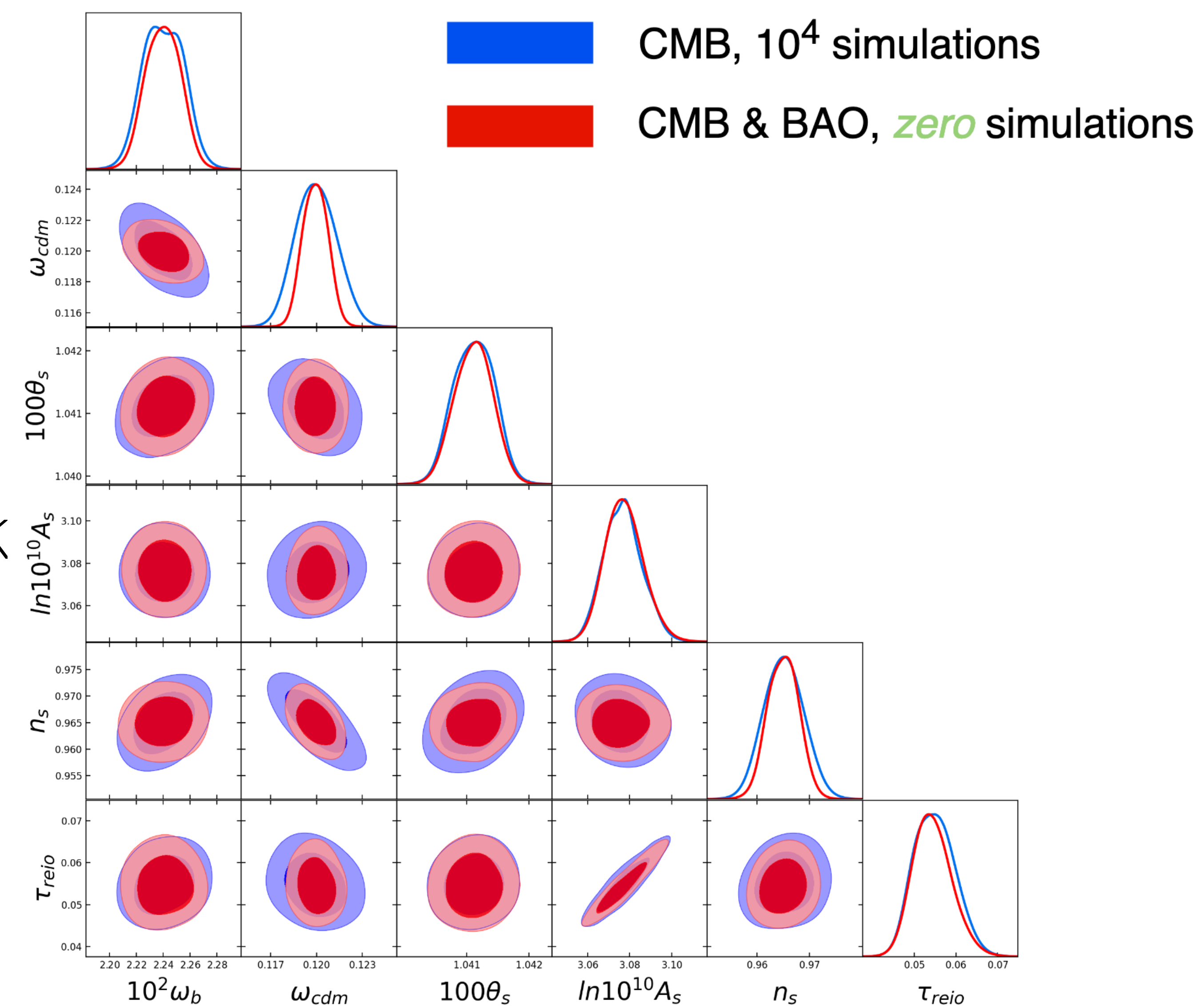


[AC et al. ’21]

Simulation reuse

- In TMNRE, **saved simulations can be reused** by subsequent inferences with different experimental configurations, priors, network structures, etc.

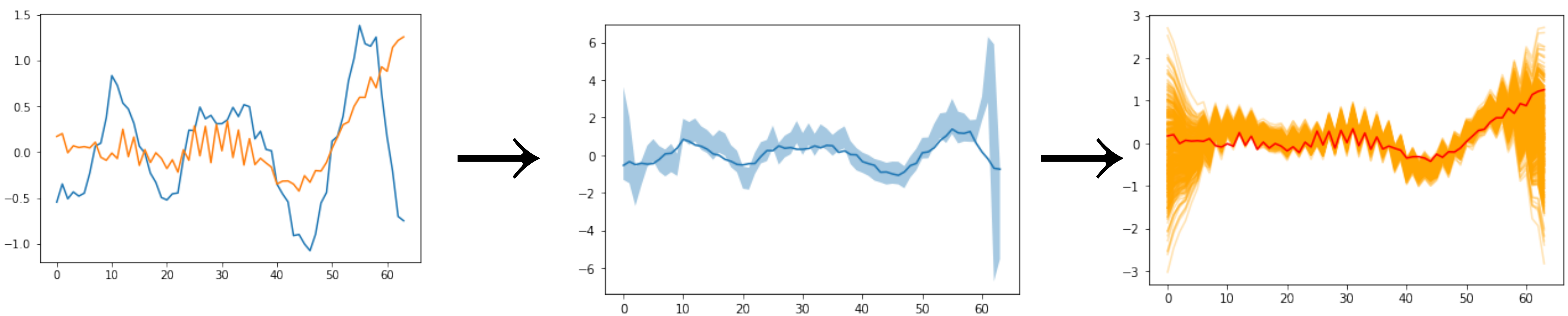
- Promising for speeding up massive forecasting efforts.



Ratio Estimation is Flexible

[AC, Weniger in progress]

- There are plenty of odds ratios beyond the likelihood-to-evidence ratio that are relevant to SBI. Conditionals, …
- Can use ratio estimation to **constrain latent variables** and generate realistic data via **constrained simulators**.



$\Theta \to_{\text{stochastic}}$ latent field $\to_{\text{stochastic PDE}}$ observed data

4. Discussion

Summary

- By directly targeting marginal posteriors, we can unlock flat scaling of simulation cost w.r.t. parameter space dimension.
- TMNRE agrees with long-run MCMC and requires order of magnitude fewer simulations.
- Rapid evaluation of many posteriors with a trained network allows for consistency tests beyond MCMC.

Discussion

- How best to combine aspects of NLE, NPE, NRE, various proposals for zooming in, marginalizing, for cosmology? For field X?
- More consistency tests?
- Pretraining (cf. LLMs) for scientific data?

Thank you!

Extra Slides

CMB forecasting

- The CMB spherical harmonic coefficients obey

$$\langle a_{\ell m}^{P*} a_{\ell' m'}^{P'} \rangle = \left(C_\ell^{PP'} + N_\ell^{PP'} \right) \delta_{\ell\ell'} \delta_{mm'} \equiv \overline{C}_\ell^{PP'} \delta_{\ell\ell'} \delta_{mm'}$$

- $C_\ell^{PP'}$ computed from e.g. $\texttt{CLASS}$, $N_\ell^{PP'}$ is instrument noise

$$N_\ell^{PP'} \equiv \langle n_{\ell m}^{P*} n_{\ell m}^{P'} \rangle = \delta_{PP'} \theta_{\rm fwhm}^2 \sigma_P^2 \exp\left(\ell(\ell+1) + \frac{\theta_{\rm fwhm}^2}{8 \ln 2} \right)$$

- Then the likelihood for $a_{\ell m}^{PP'}$ is given by

$$p(\mathbf{a} \mid \boldsymbol{\theta}) \propto \frac{1}{|\overline{C}(\boldsymbol{\theta})|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{a}^\dagger [\overline{C}(\boldsymbol{\theta})^{-1}] \mathbf{a} \right) \qquad \mathbf{a} = \{a_{\ell m}^T, a_{\ell m}^E\}$$

also B-modes, weak lensing, …
here we restrict for simplicity

CMB forecasting

- Given $\overline{C}_\ell^{PP'}$, we can sample $a_{\ell m}^P$ according to

$$p(\mathbf{a} \mid \boldsymbol{\theta}) \propto \frac{1}{|\overline{C}(\boldsymbol{\theta})|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{a}^\dagger[\overline{C}(\boldsymbol{\theta})^{-1}]\mathbf{a}\right) \qquad \mathbf{a} = \{a_{\ell m}^T, a_{\ell m}^E\}$$

$$\begin{pmatrix} a_{\ell m}^T \\ a_{\ell m}^E \end{pmatrix} = L \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \qquad LL^T = \begin{pmatrix} \overline{C}_\ell^{TT} & \overline{C}_\ell^{TE} \\ \overline{C}_\ell^{TE} & \overline{C}_\ell^{EE} \end{pmatrix} \qquad n_i \sim \mathcal{N}(\mu = 0,\, \sigma = 1)$$

- For a single realization of the universe, we can only determine the maximum likelihood values for $\overline{C}_\ell^{PP'}$, denoted $\hat{C}_\ell^{PP'}$

$$\hat{C}_\ell^{PP'} = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} a_{\ell m}^{P*} a_{\ell m}^{P'} = \frac{1}{2\ell+1}\left(a_{\ell 0}^{P} a_{\ell 0}^{P'} + 2\sum_{m=1}^{\ell} a_{\ell m}^{P*} a_{\ell m}^{P'}\right)$$

CMB forecasting

- The likelihood for $\hat{C}_\ell^{PP'}$ is a Wishart distribution

$$-2\ln p\left(\hat{C}(\boldsymbol{\theta}) \mid \overline{C}\right) = \chi^2_{\rm eff} = \sum_\ell (2\ell+1)\left[\frac{D}{|\overline{C}|} + \ln\frac{|\overline{C}|}{|\hat{C}|} - 2\right]$$

$$D = \overline{C}_\ell^{TT}\hat{C}_\ell^{EE} + \hat{C}_\ell^{TT}\overline{C}_\ell^{EE} - 2\overline{C}_\ell^{TE}\hat{C}_\ell^{TE}$$

- At high ℓ, this is approximately Gaussian with covariance

$$\mathrm{Cov}_{C_\ell} = \frac{2}{2\ell+1}\begin{pmatrix} \left(\overline{C}_\ell^{TT}\right)^2 & \overline{C}_\ell^{TT}\overline{C}_\ell^{TE} & \left(\overline{C}_\ell^{TE}\right)^2 \\ \overline{C}_\ell^{TT}\overline{C}_\ell^{TE} & \frac{1}{2}\left(\overline{C}_\ell^{TT}\overline{C}_\ell^{EE} + \left(C_\ell^{TE}\right)^2\right) & \overline{C}_\ell^{TE}\overline{C}_\ell^{EE} \\ \left(\overline{C}_\ell^{TE}\right)^2 & \overline{C}_\ell^{TE}\overline{C}_\ell^{EE} & \left(\overline{C}_\ell^{EE}\right)^2 \end{pmatrix}$$