Coverage is not enough: What SBI posteriors of f_{NL} are actually telling you

aka: “Frequentist tests of SBI for primordial non-Gaussianity”

Toka Alokda
PhD candidate - University of Bonn

(arXiv:2605.00980) with Cristiano Porciani and Alexander Eggemeier

10.09.2026 - AG Meeting 2026, Munich

UNIVERSITÄT BONN

Argelander-Institut für Astronomie

IMPRS for Astronomy & Astrophysics Bonn and Cologne
INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

DYNAVERSE

Andrew Pontzen and Fabio Governato

Primordial non-Gaussianity and the CMB wall

The CMB is a 2D map – gives access to limited no. of modes (cosmic variance)

Local

Equilateral

Orthogonal

- **Not** degenerate with gravitational non-Gaussianities.
- **Scale-dependent bias** in galaxy power spectrum.
- **Correlates large and small-scales** (e.g. squeezed bispectrum).

Direct LSS measurements from summaries achievable!

At late times: degenerate with gravitational non-Gaussianities

Require alternative approaches such as field-level inference, etc.

Current best

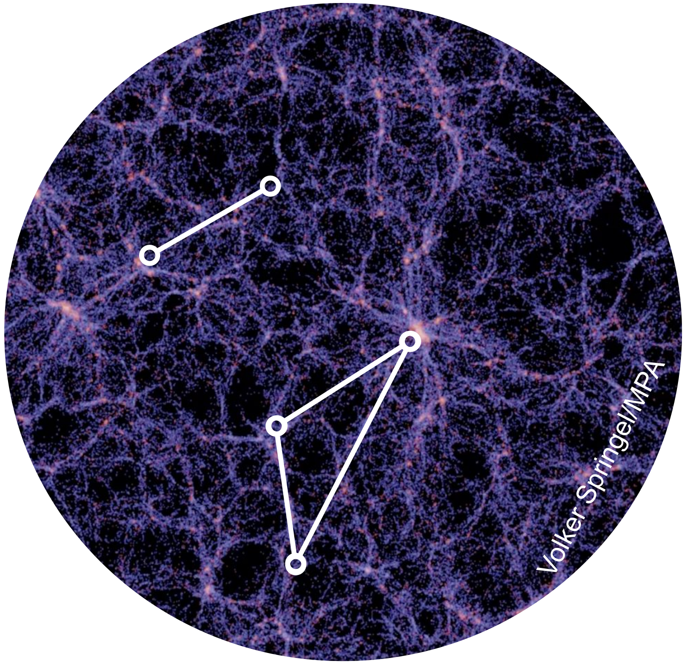
$$f_{\rm NL}^{\rm local} = -0.9 \pm 5.1$$
$$f_{\rm NL}^{\rm equil} = -26 \pm 47$$
$$f_{\rm NL}^{\rm ortho} = -38 \pm 24$$

Planck Collaboration (2020) arXiv:1905.05697

Primordial non-Gaussianity and the LSS promise

Traditional LSS analyses
(*n*-point statistics)

- The curse of **dimensionality**
- The better the model, the more the **(nuisance)** parameters
- **Gaussian likelihoods** may not be precise



Primordial non-Gaussianity and the LSS promise

Traditional LSS analyses (*n*-point statistics)

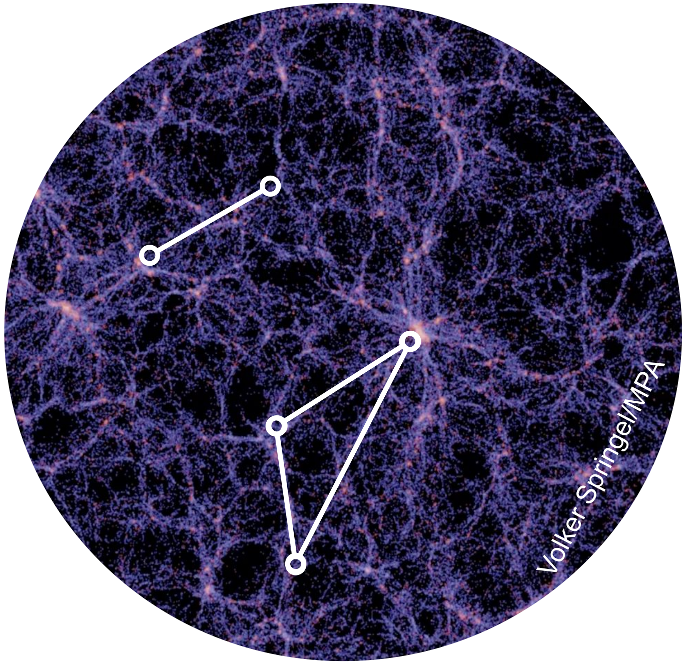
The curse of **dimensionality**

The better the model, the more the **(nuisance)** parameters

Gaussian likelihoods may not be precise

Alternative summaries + SBI

Can even be parameter-optimal!

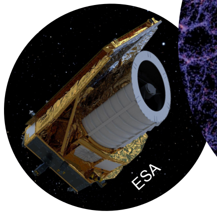


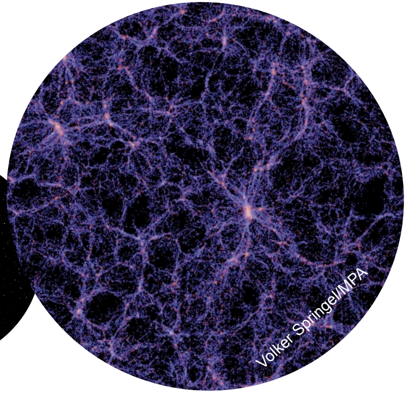
Motivation for frequentist testing of SBI

- We can simulate **infinite universes** with the **same cosmology**, but we only get **one** observation.

- **Frequentist tests** help us convince ourselves of whether our inference methods are **robust** to **sample noise/cosmic variance at a fixed truth**.

- SBI methods are validated through the distribution of coverage **over the prior**; meanwhile frequentist testing uncovers systemic biases at such **blind spots**.





Summary statistics

2-point

Power spectrum
"***P***"

D = 7

3-point

Bispectrum
"***B***"

D = 62

(mainly)
2- and 4-point info.

Wavelet Scattering Transform
"**WST**"
Coefficients (Eickenberg et al. 2018)

D = 41

$$\boxed{k_{\max} = 0.14\, h\,\mathrm{Mpc}^{-1}}$$

Summary statistics

2-point

Power spectrum
"*P*"

D = 7

3-point

Bispectrum
"*B*"

D = 62

(mainly)
2- and 4-point info.

Wavelet Scattering Transform
"**WST**"
Coefficients (Eickenberg et al. 2018)

D = 41

$$\boxed{k_{\max} = 0.14\,h\,\mathrm{Mpc}^{-1}}$$

Multivariate normality tests
mnt-R library
(Ebner and Henze 2020)

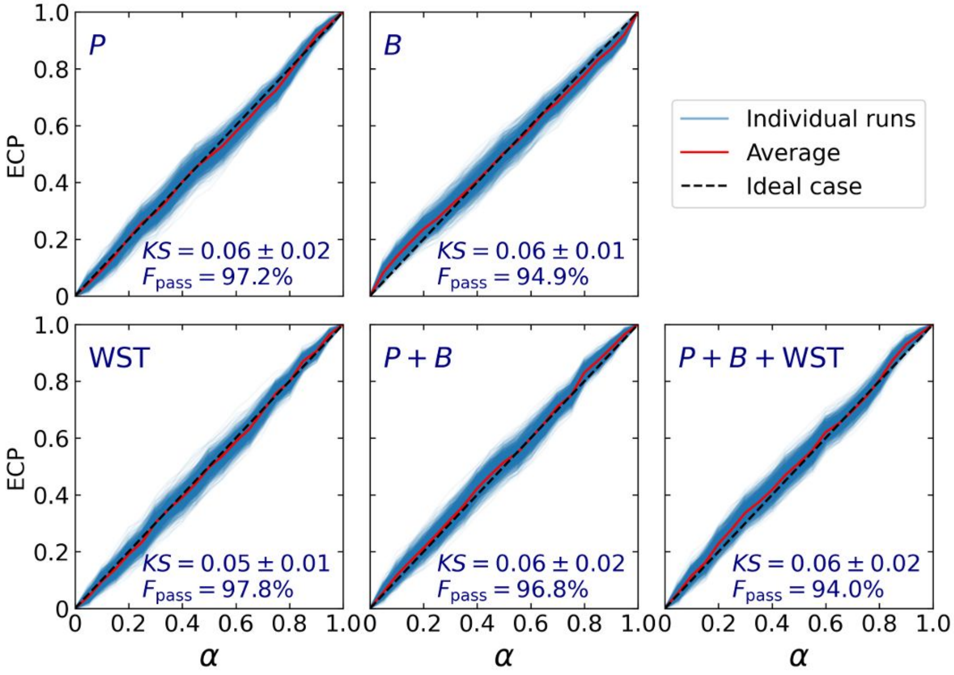
Test	P	B	$P+B$	WST	$P+B$+WST
Mardia γ	✓	✗	✗	✗	✗
Mardia κ	✓	✓	✓	✗	✓
HZ	✓	✗	✗	✓	✓
SR Energy	✓	✗	✗	✓	✓
DEHT	✓	✗	✗	✗	✗
DEHU	✓	✗	✗	✗	✗

Primordial non-Gaussianity as a test case

- We focus on inferring a **single parameter:** $f_{\rm NL}^{\rm local}$ from **1000 realizations** from the **Quijote** suite (Villaescusa-Navarro et al., 2020)–These have $f_{\rm NL}^{\rm local}$ = 0.

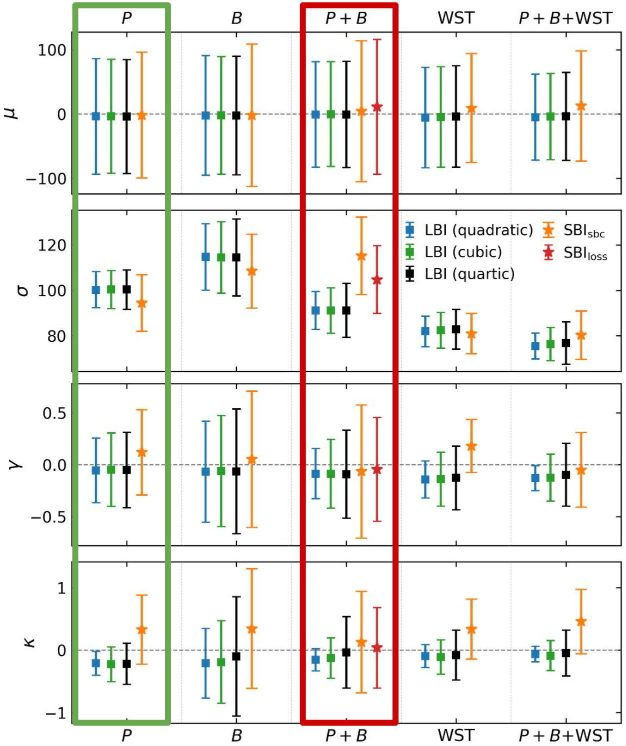
- **SBI** with Contrastive NRE (Miller et al. 2022) is trained with the **Quijote-PNG** (Coulton et al. 2022) simulations: $p(f_{\rm NL}) = \mathcal{U}(-300, 300)$ – optimized with **SBC** (Talts et al. 2018), and validated with **TARP coverage** (Lemos et al. 2023).

- **LBI** posteriors obtained straightforwardly via the Bayes' formula (with simulation-based covariance estimates).

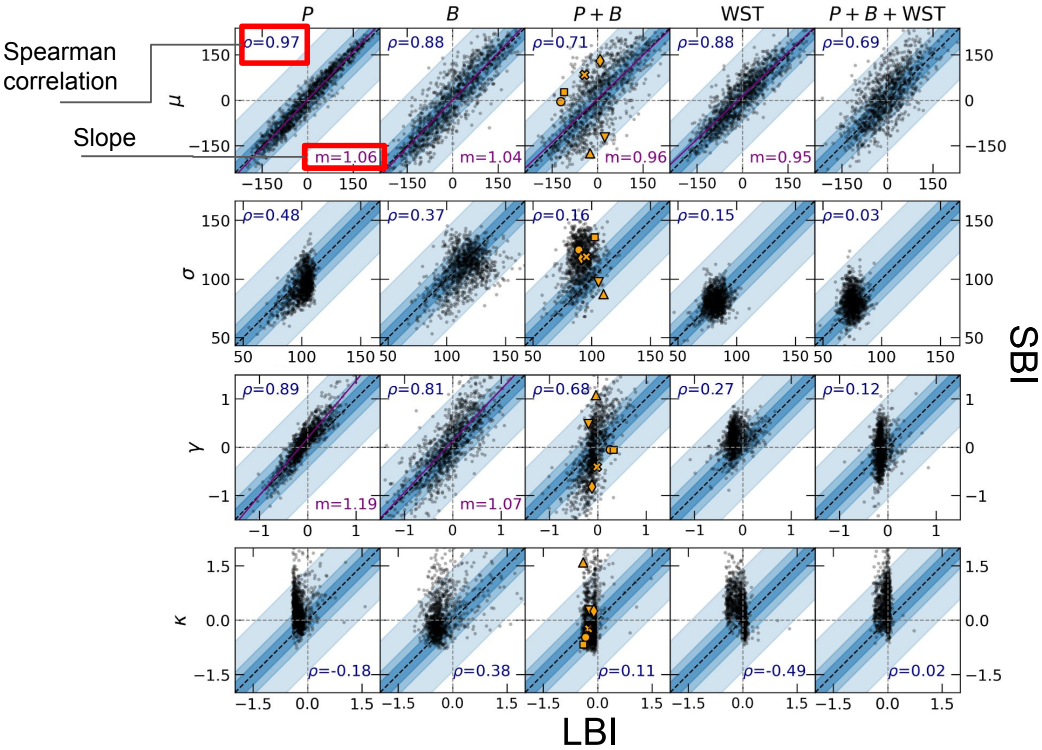


Direct comparison of posterior moments

On average

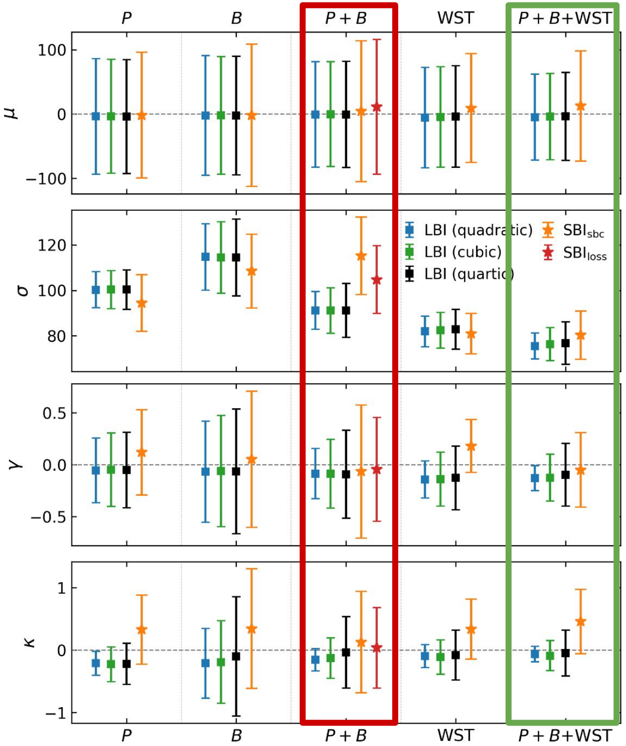


realization-wise

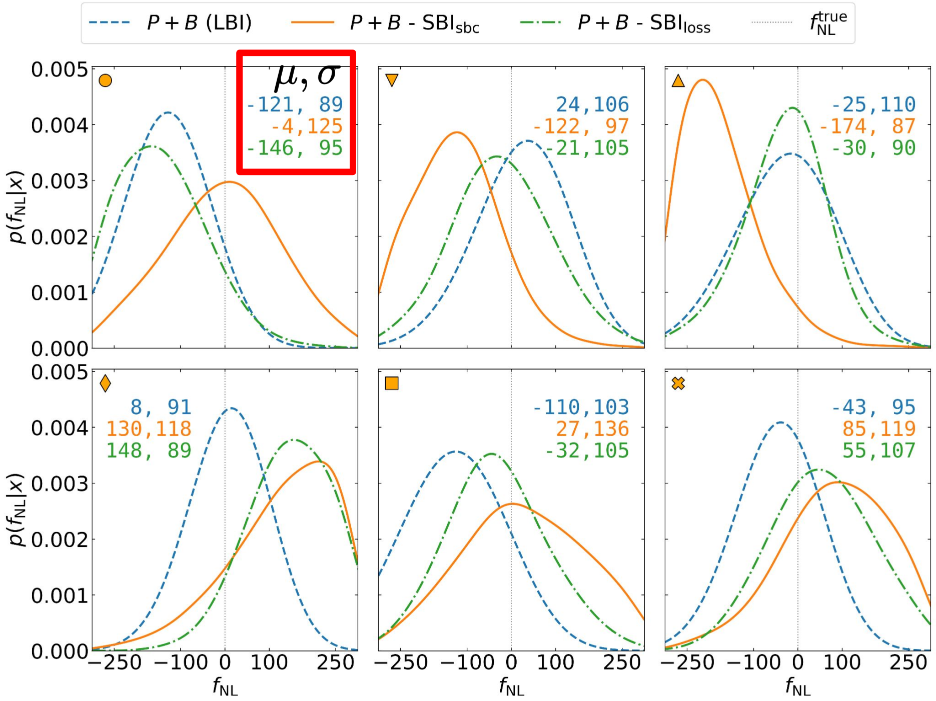


The case of *P+B*

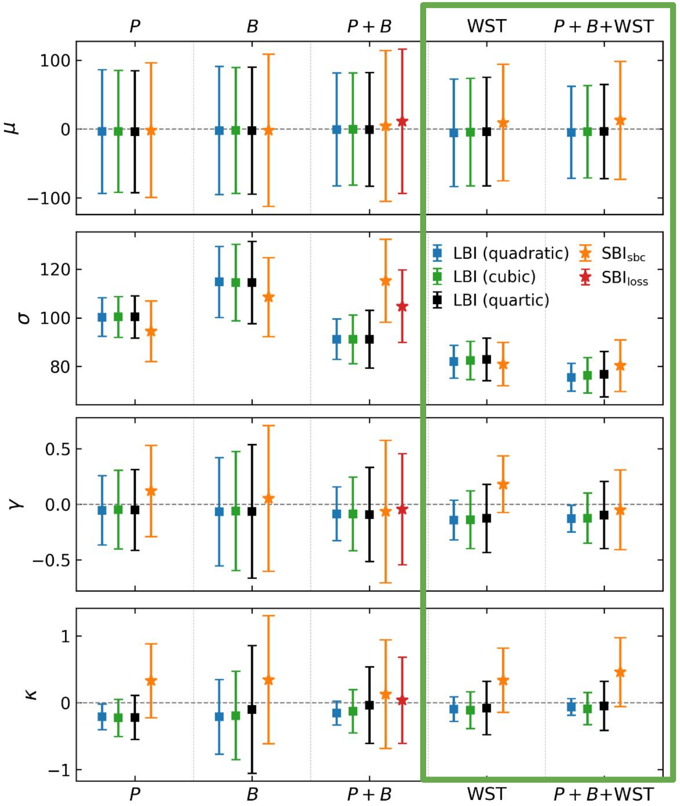
On average



Some individual cases

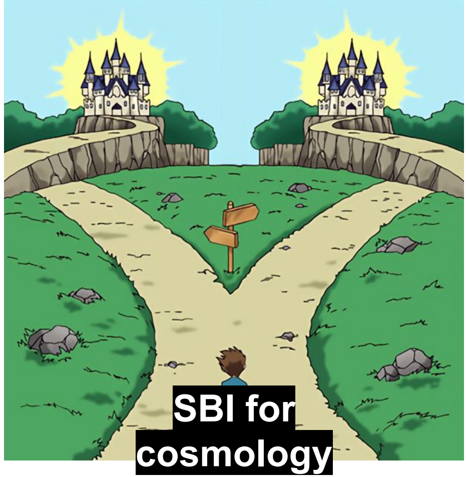


Implications on $f_{\mathrm{NL}}^{\mathrm{local}}$ constraints



- Although we **restrict ourselves** to $k_{\max} = 0.14\, h\, \mathrm{Mpc}^{-1}$, WST coefficients contain **additional information** on $f_{\mathrm{NL}}^{\mathrm{local}}$.
- This motivates serious pursuit of **robust SBI** methods and **additional diagnostics** to **safely pursue** novel higher-order summaries, beyond *P*+*B*.

Some ways forward



- **Diagnostically:**
  In theory, **coverage is indeed enough**. However, frequentist discrepancies would only show in practice if each point in the prior is sampled **many times**–may require even more simulations to validate than to train!

- **Methodically:**
  Perhaps new SBI methods (architectures or training strategies to control these implicit biases? – happy to discuss!).

Read the paper!