BOSTON UNIVERSITY

A data-driven prediction for the primordial deuterium abundance

Tim Launders, Boston University

With Cara Giovanetti (UC Berkeley), Hongwan Liu (BU)

arXiv:2604.16600

Pheno, 11 May 2026

Punchline:

- We can do "precision cosmology" with BBN [1]!

[1] T. Yeh *et al.*, 2601.22239.

Punchline:

- We can do “precision cosmology” with BBN [1]!

Assuming Planck CMB baryon density, our prediction is 1.7σ discrepant with observation

[1] T. Yeh *et al.*, 2601.22239.

Punchline:

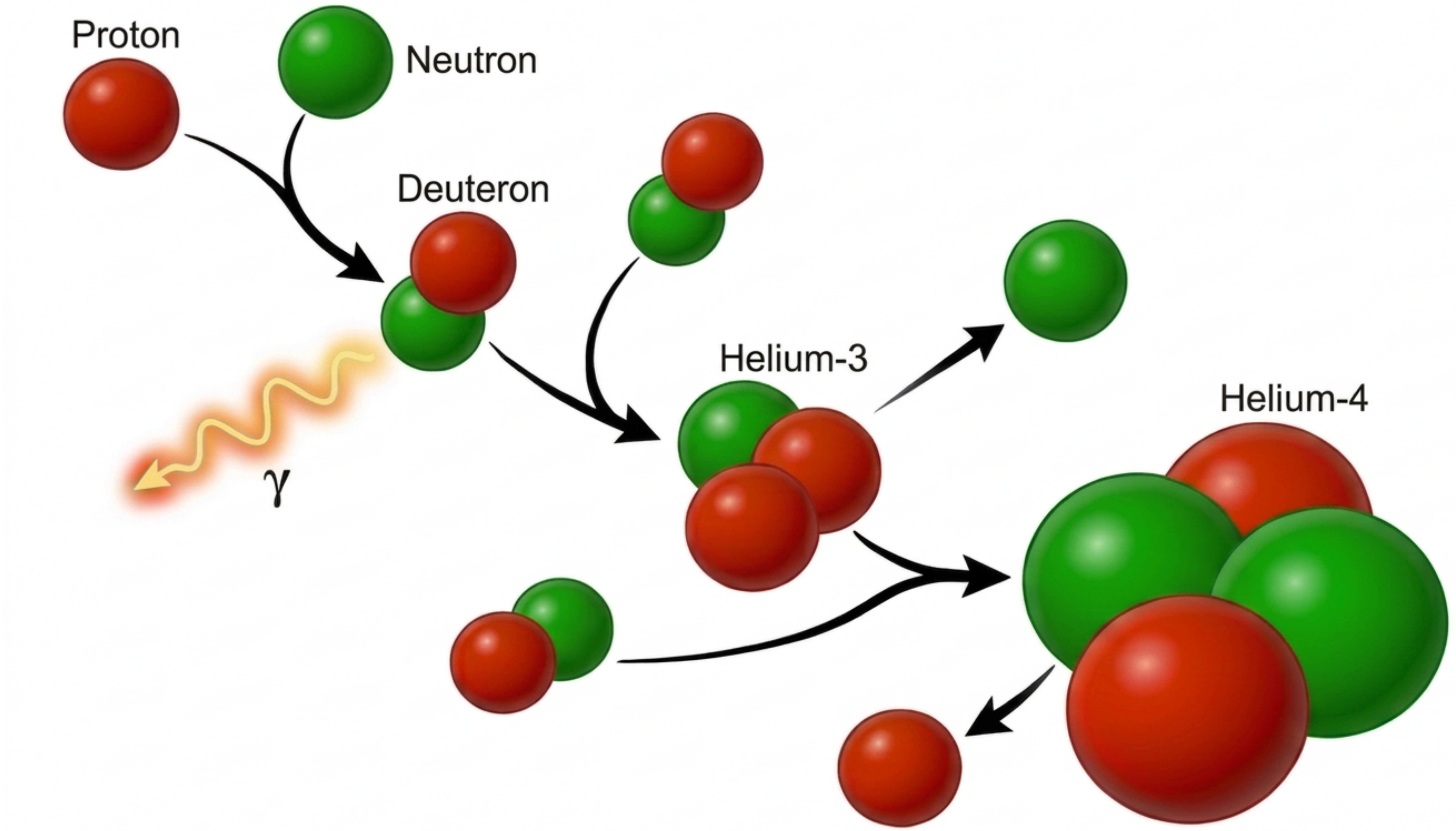
- We can do "precision cosmology" with BBN [1]!

Assuming Planck CMB baryon density, our prediction is 1.7σ discrepant with observation

Systematics? New physics?

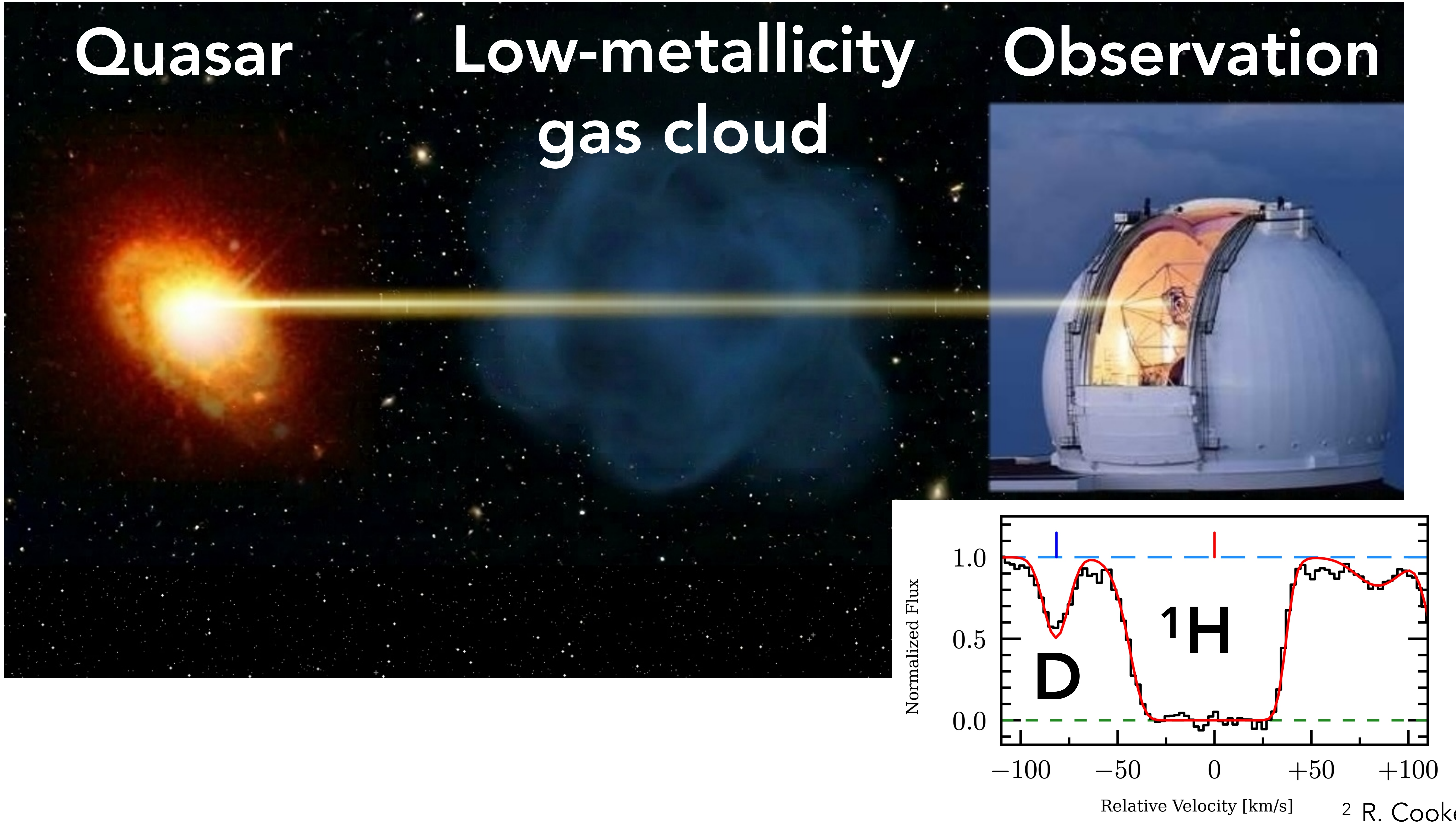
[1] T. Yeh *et al.*, 2601.22239.

Big Bang Nucleosynthesis (BBN)



Highly sensitive to primordial physics!

Percent-level measurement of primordial D



[2] R. Cooke, 2409.06015.

How do we make a theory prediction?

How do we make a theory prediction?

Lots of physics (and numerics)!

- Calculate background evolution of the universe
- Compute proton/neutron inventories at high temperatures
- Solve coupled differential equations for nuclear reactions!

How do we make a theory prediction?

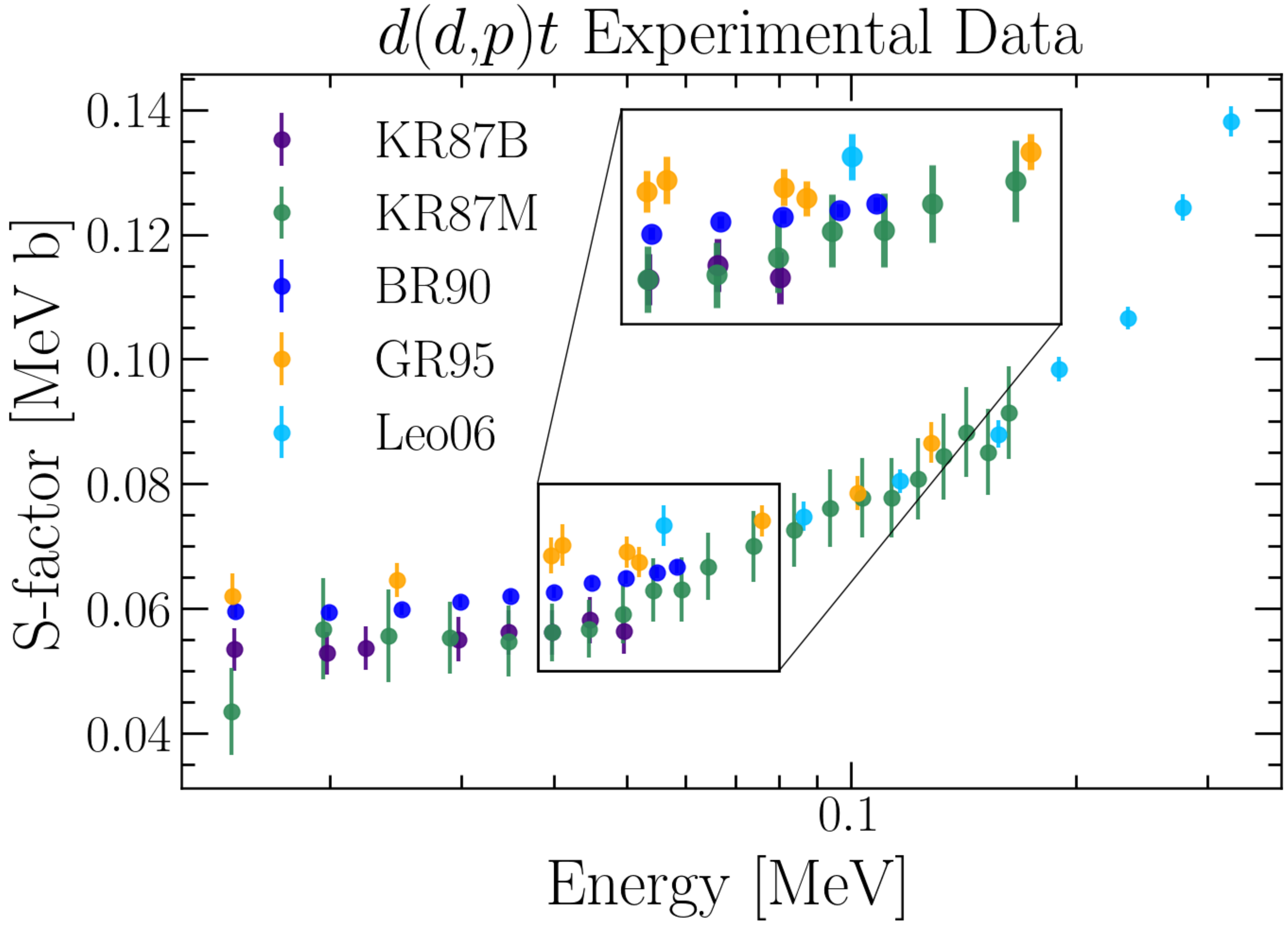
Lots of physics (and numerics)!

- Calculate background evolution of the universe
- Compute proton/neutron inventories at high temperatures
- Solve coupled differential equations for nuclear reactions!

↓

Primordial light nuclei abundances (D, ^{3}He, ^{4}He, ^{7}Li)

Nuclear Reactions

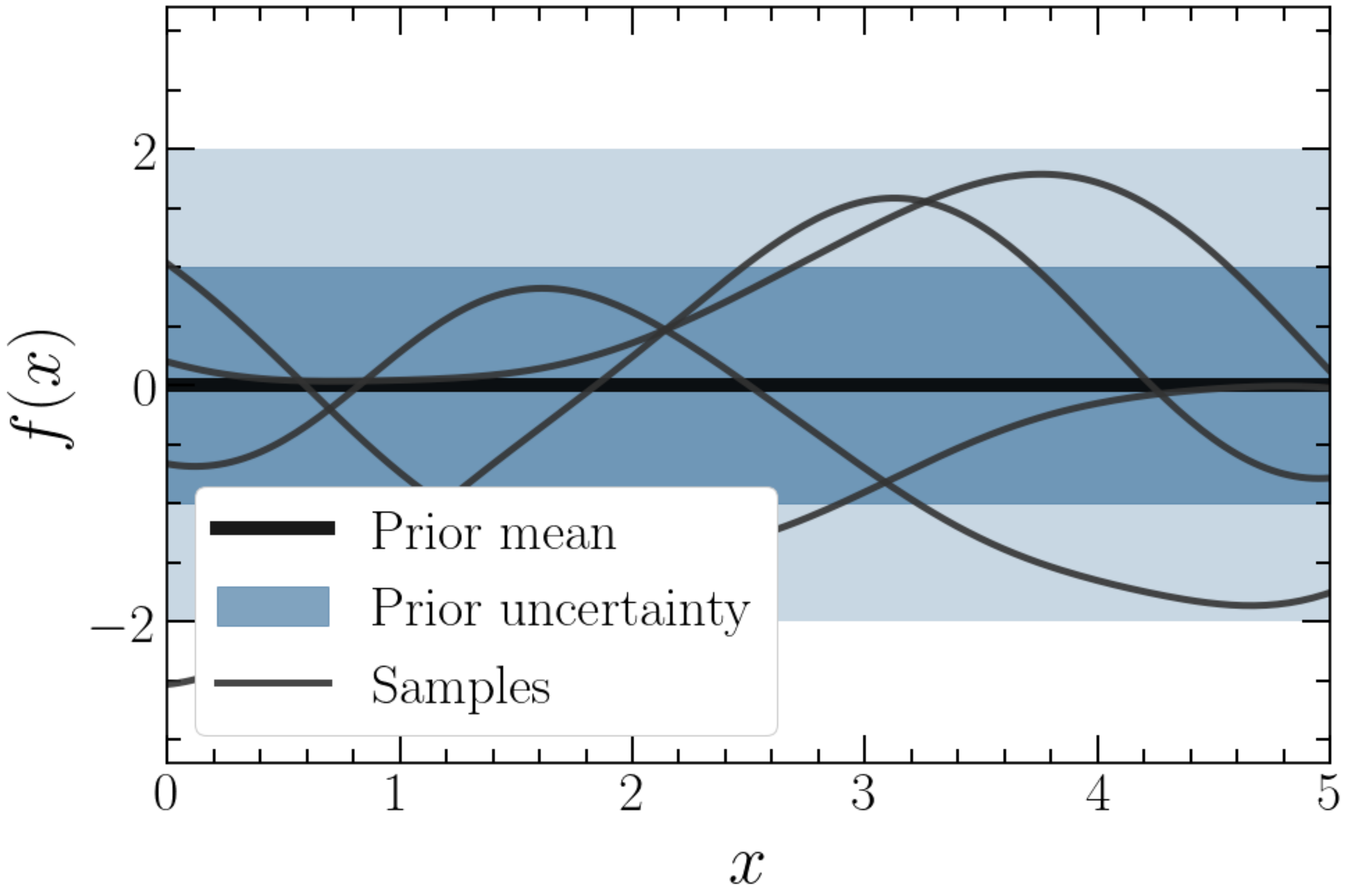


Must fit these data…somehow.

Polynomials? Nuclear theory?

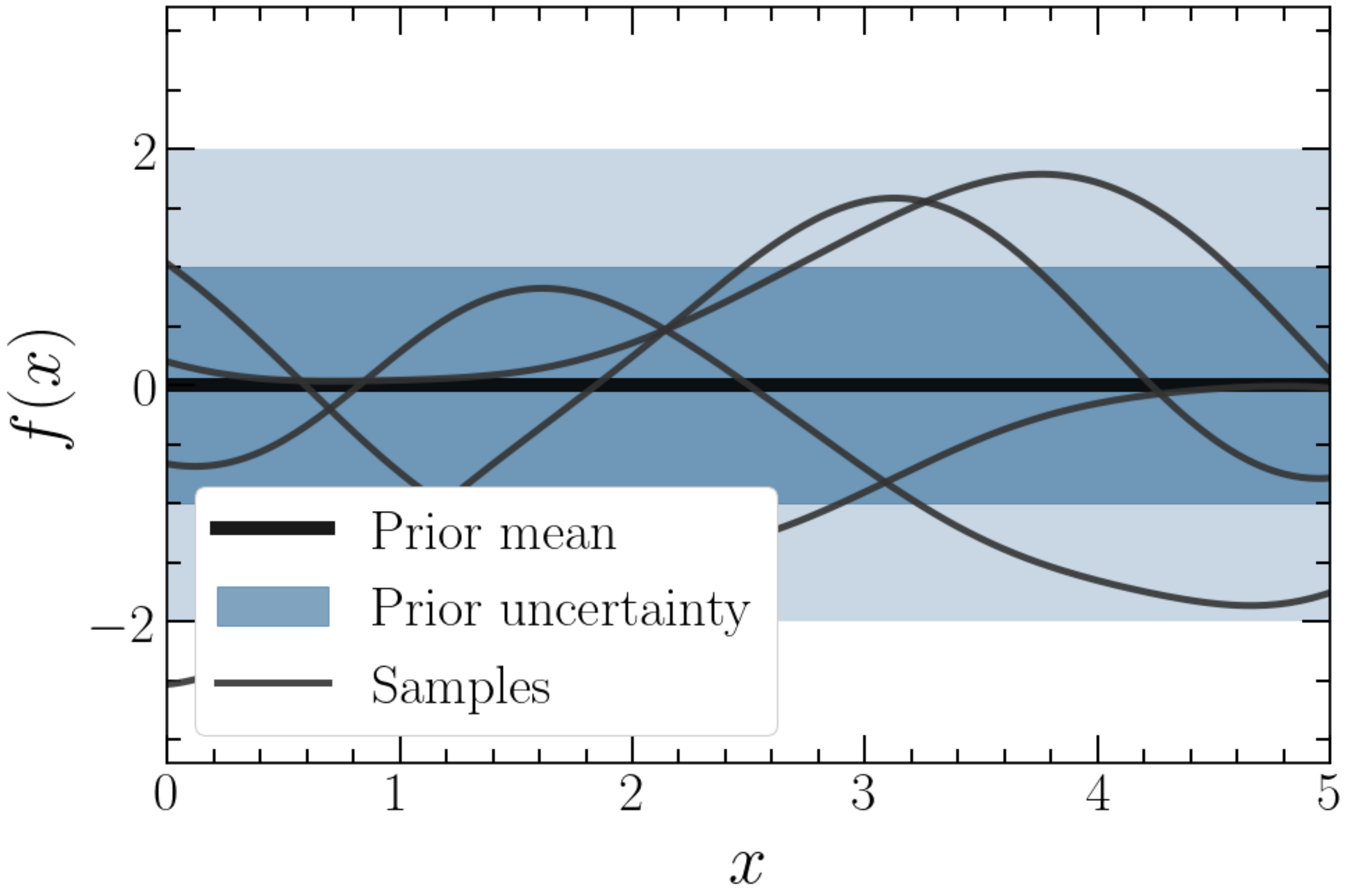
Gaussian Processes

- **Probability distribution over functions that can explain the data**



Gaussian Processes

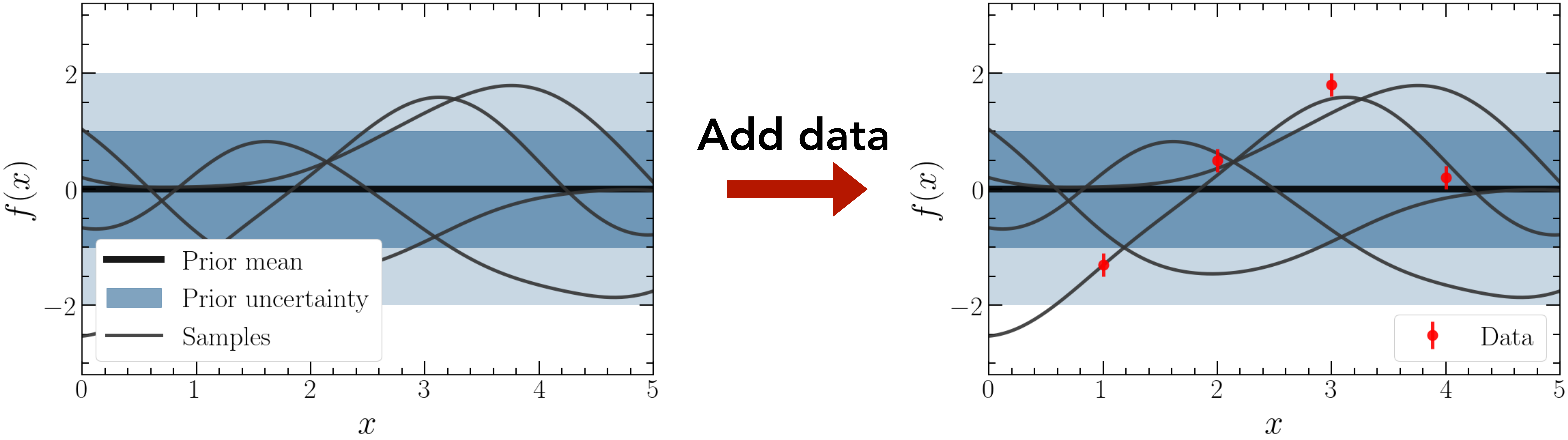
- **Probability distribution over functions that can explain the data**



Add data

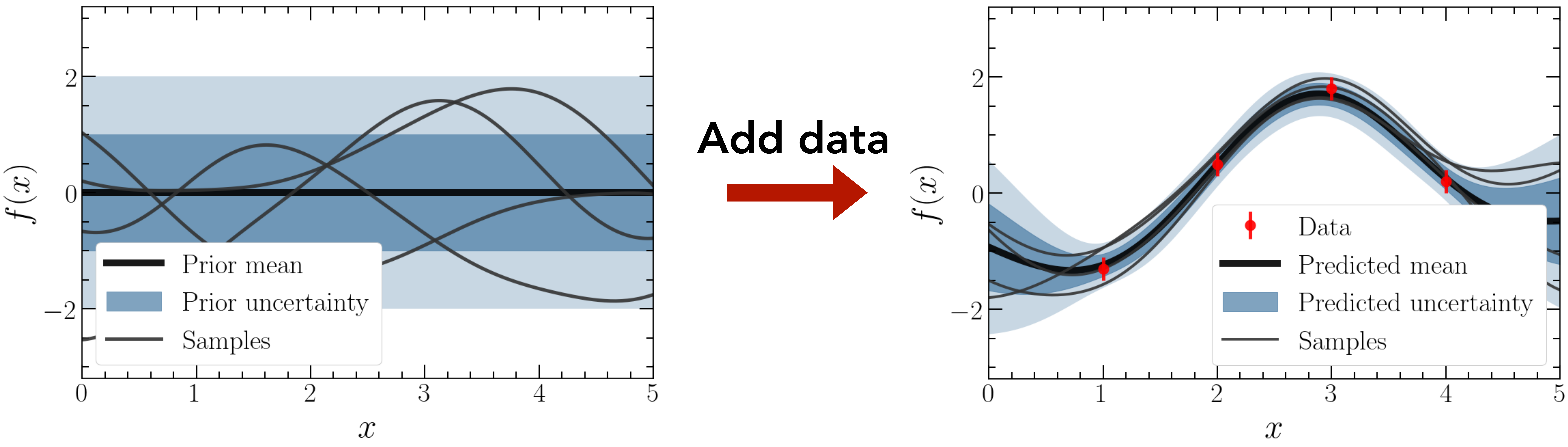
Gaussian Processes

- **Probability distribution over functions that can explain the data**

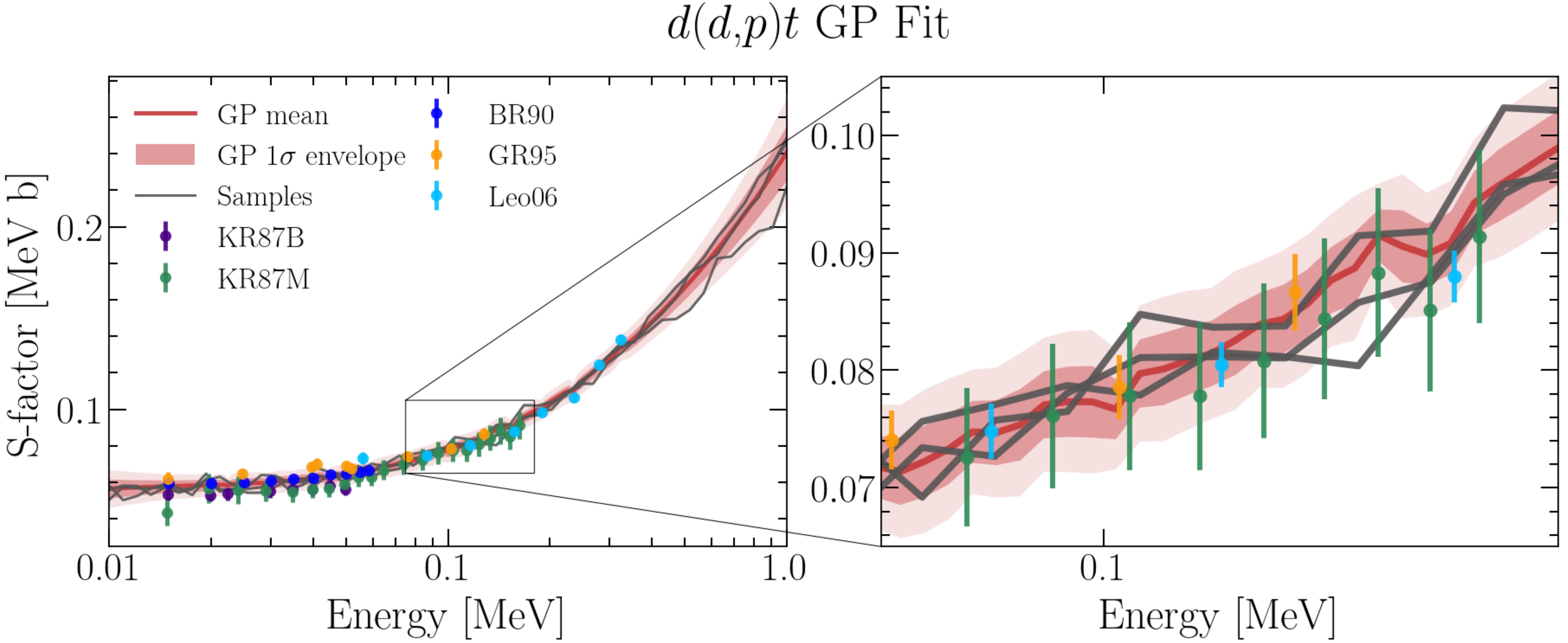


Gaussian Processes

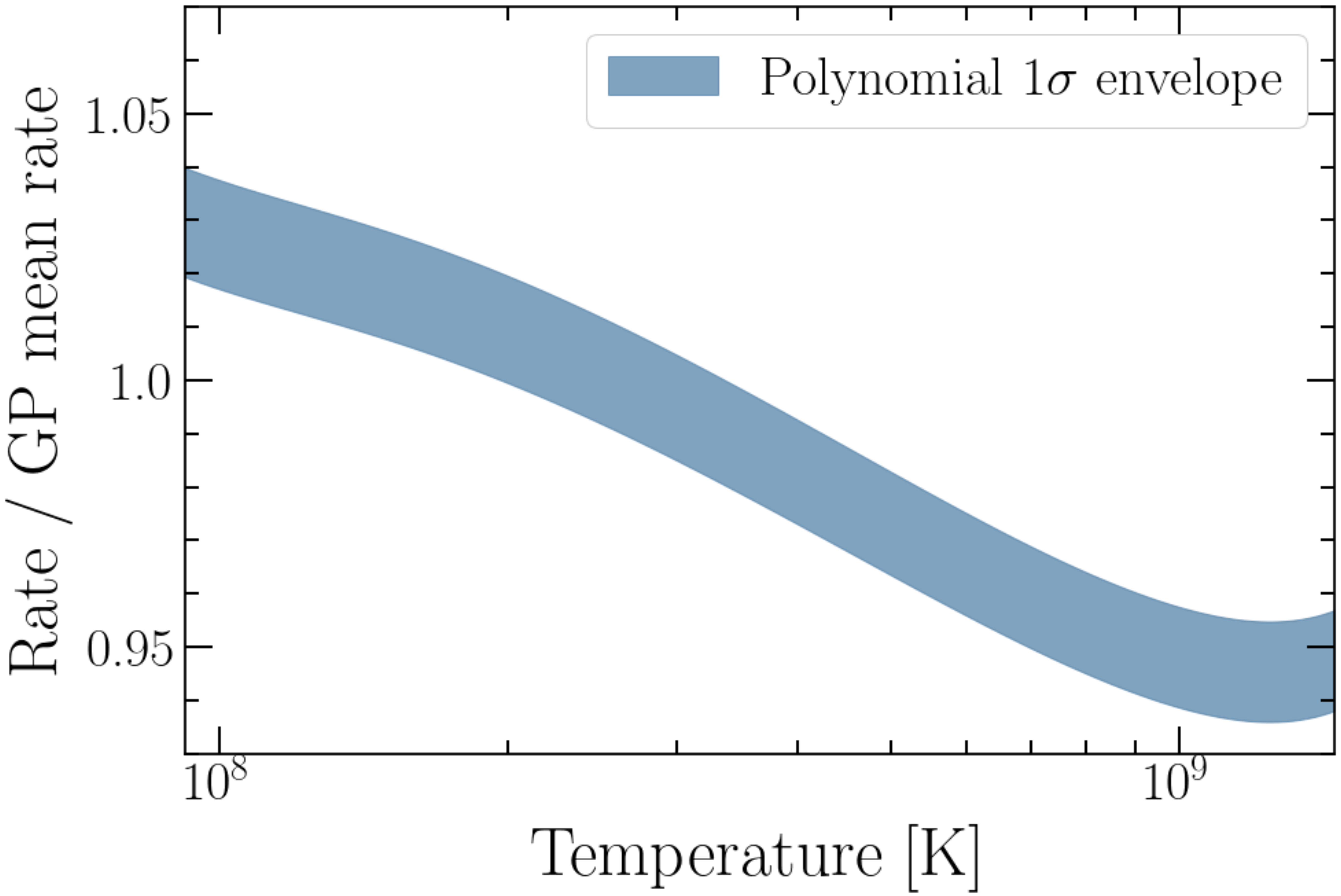
- **Probability distribution over functions that can explain the data**



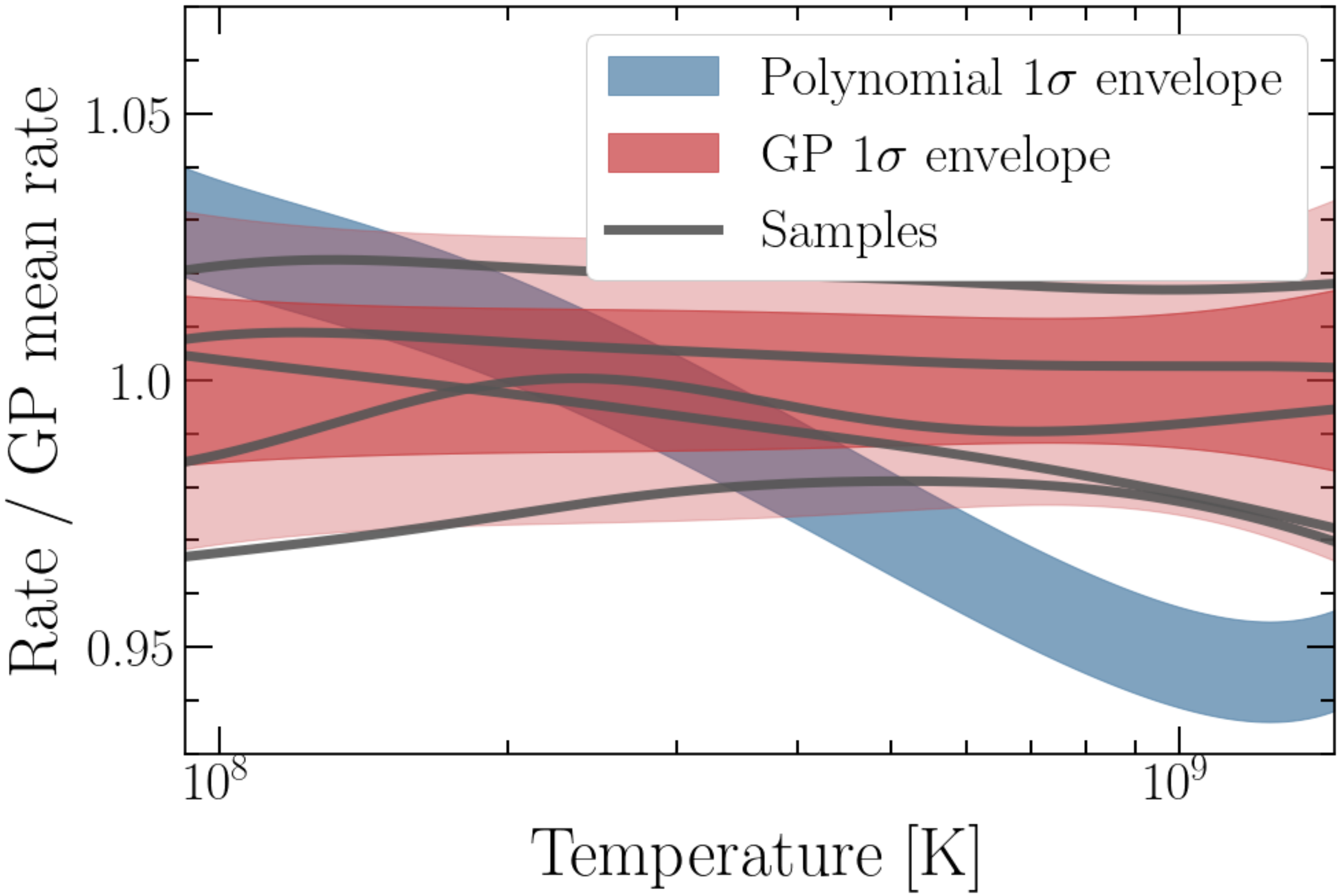
Nuclear Reactions with GPs



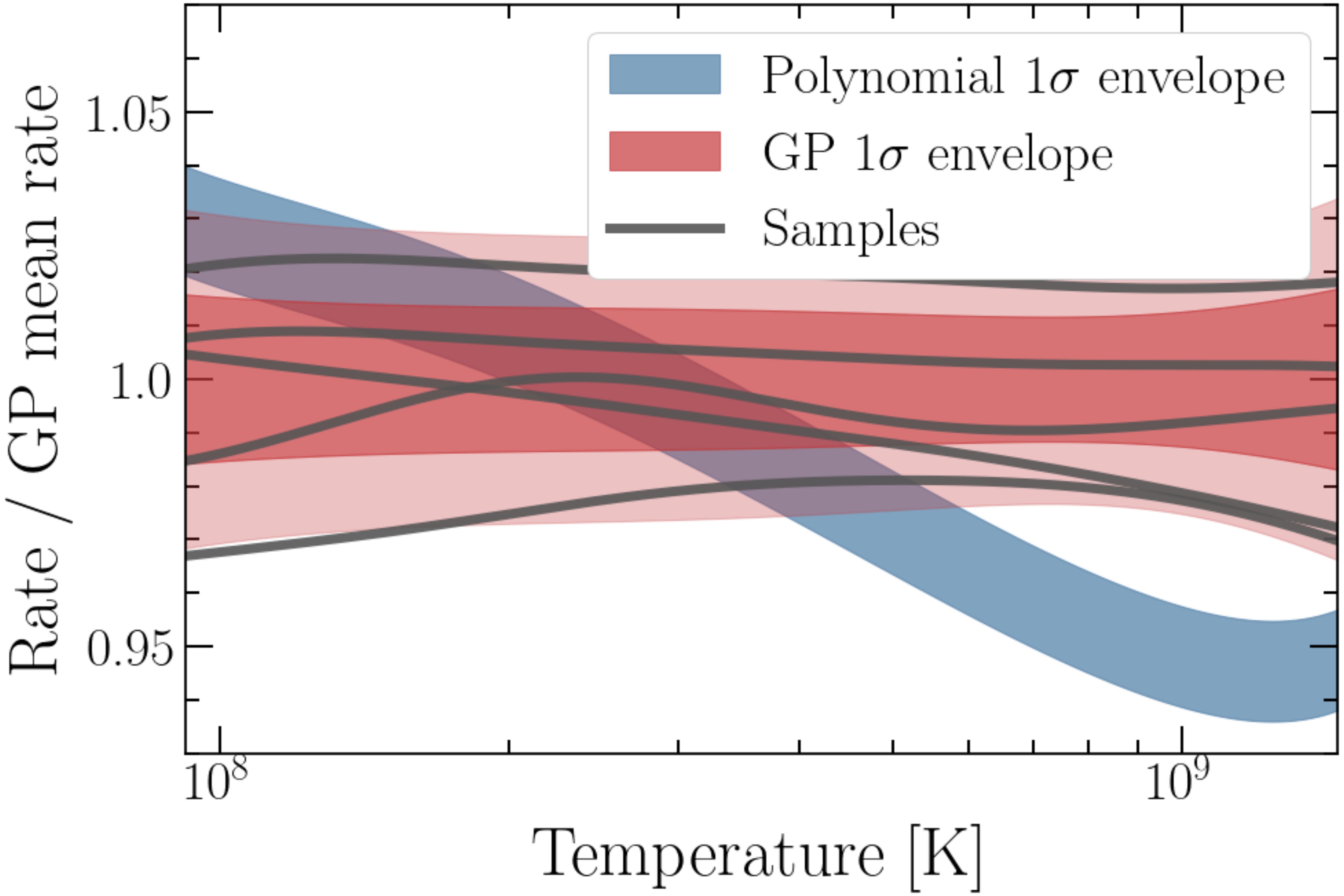
After thermal averaging,



After thermal averaging,



After thermal averaging,

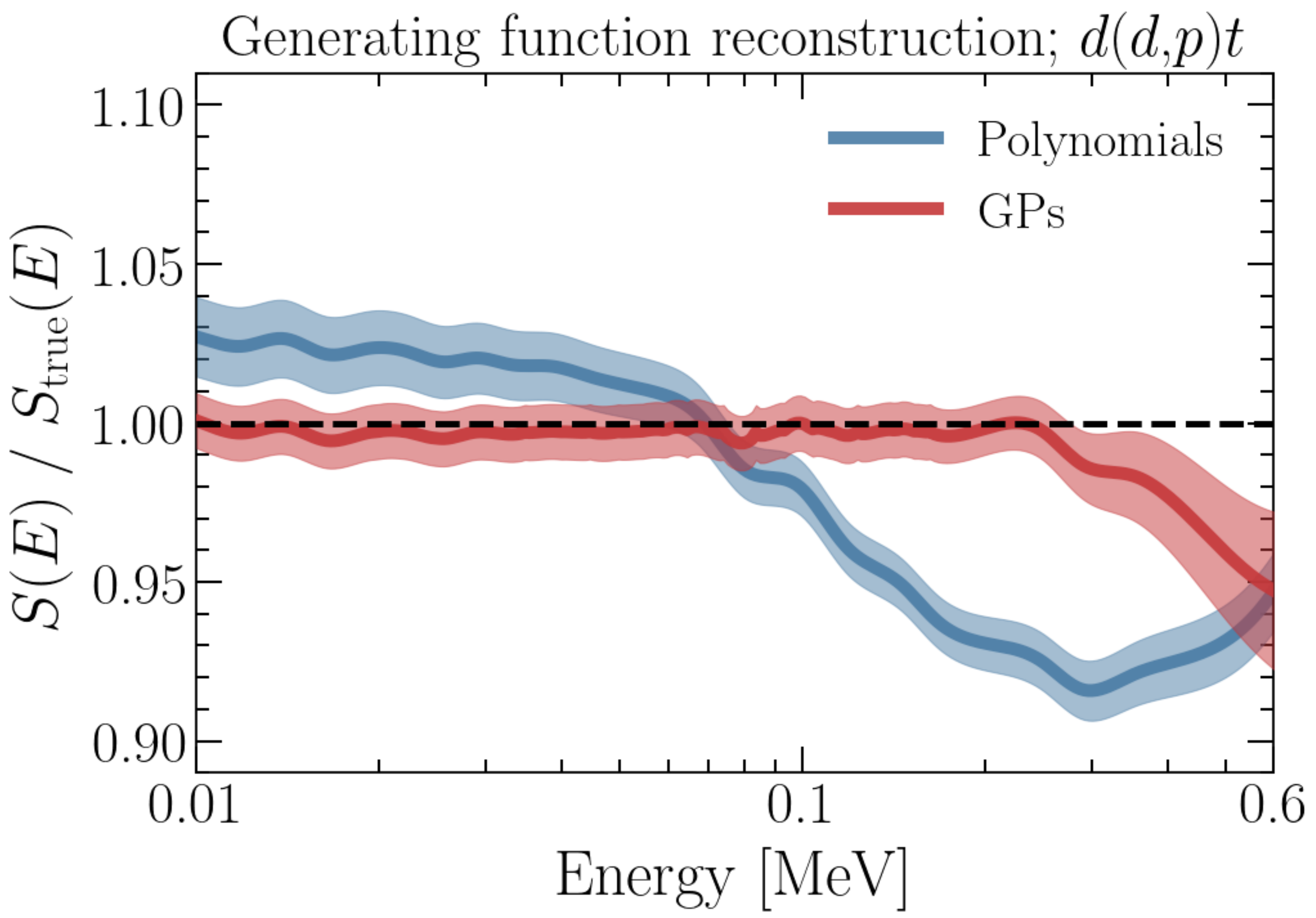


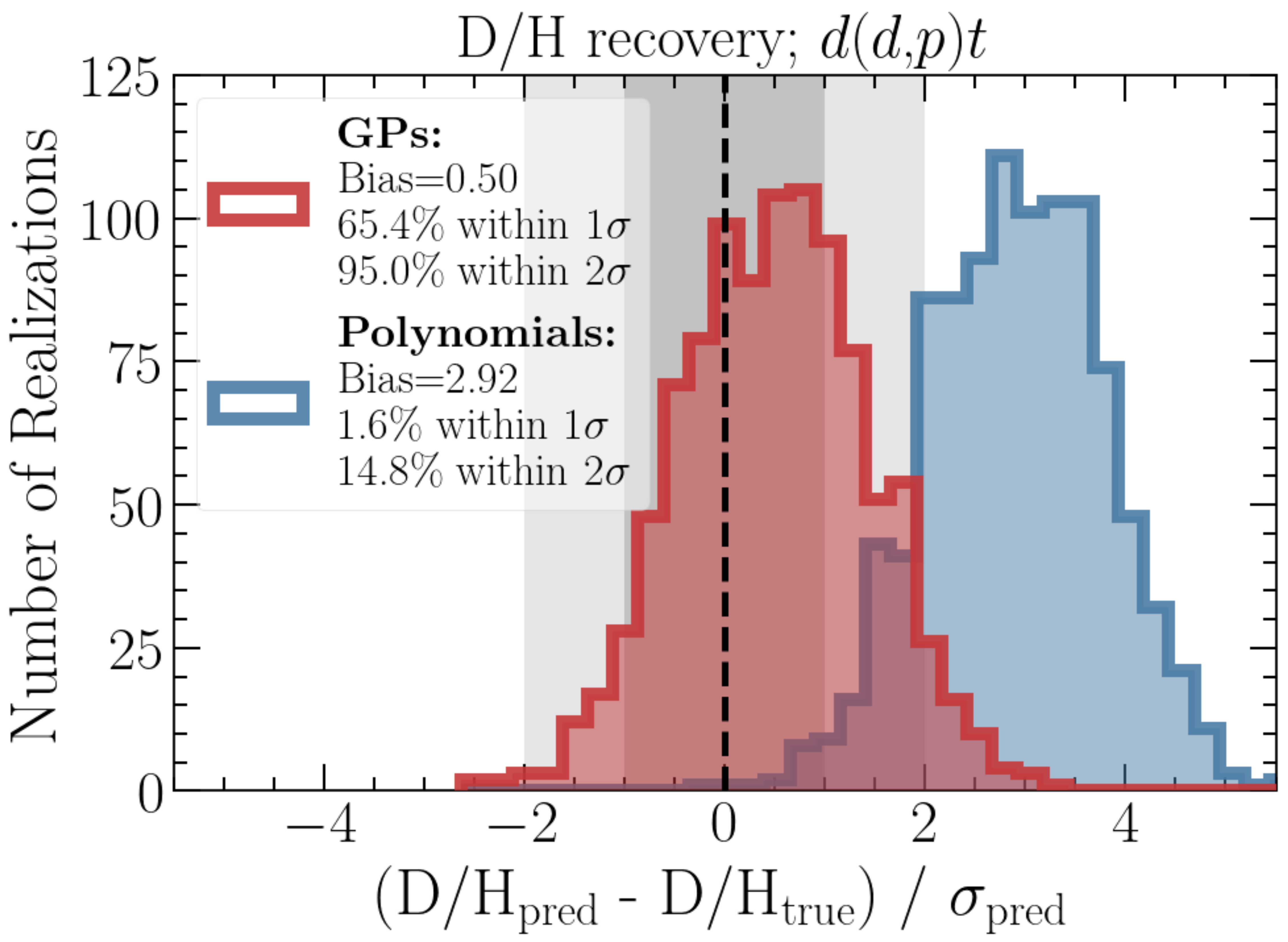
Each reaction rate sample goes directly into LINX[3] (BBN code)

[3] C. Giovanetti *et al.*, *Phys. Rev. D.* **112**, 063531 (2025), 2408.14538.

Monte Carlo validation tests

Can GPs learn S-factors and D/H from many realizations of mock data?





Deuterium Abundance Comparison

Reaction network	D/H x 10^{-5}
Gaussian Processes (This work)	**2.442 ± 0.040**
Scaled theory curve (PRIMAT)	2.444 ± 0.037
Polynomials (PArthENoPE)	2.512 ± 0.042
Measurement [4]	2.527 ± 0.030

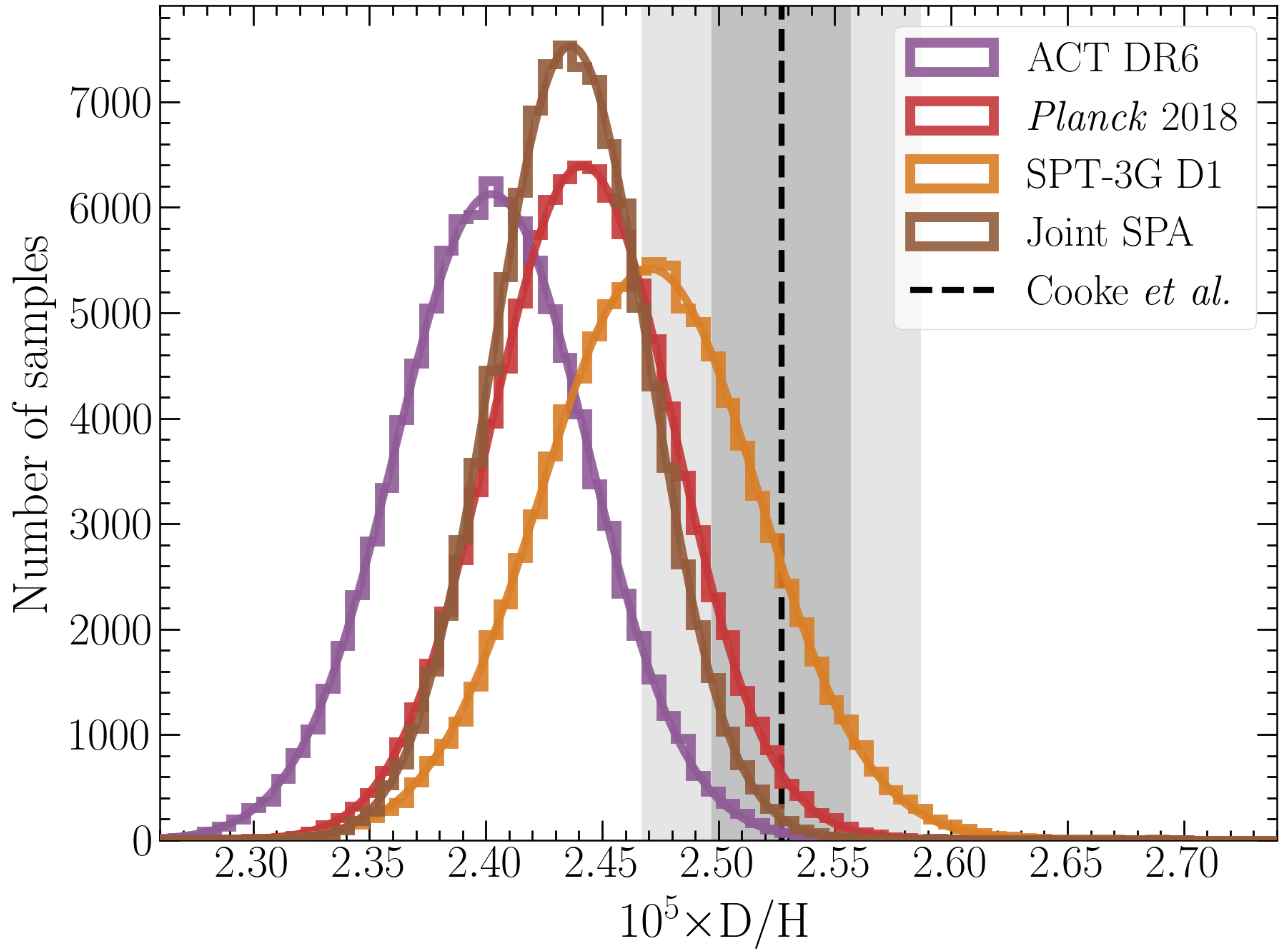
Assuming Planck CMB ω_b, agrees with measurement to 1.7σ — *view as discrepancy between CMB and BBN*

[4] R. Cooke, M. Pettini, and S. Steidel, *ApJ*, **855**, 102 (2018), 1710.11129

Thanks for listening!

Questions?

Dependence on ω_b



Combined CMB inference (*Planck*, ACT, SPT) — tighter constraint

More discrepant with measurement (2σ)

Gaussian Processes

Add statistical noise **AND** systematic uncertainties

- **Collection of random variables, any finite number of which are normally distributed**

$$\left(\begin{array}{c} \text{Known data} \\ \hdashline \text{Prediction points} \end{array}\right) \sim \mathcal{N}\left(\text{Prior mean}, \ \text{Cov} = \left[\begin{array}{c:c} \text{Upper left} & \text{Upper right} \\ \hdashline \text{Lower left} & \text{Lower right} \end{array}\right]\right)$$

Determined by the kernel

Gaussian Processes

- Make predictions with conditional on known data

$$\left(\begin{array}{c} \boldsymbol{S} \\ \hdashline \boldsymbol{S}_* \end{array}\right) \sim \mathcal{N}\left(\left[\begin{array}{c} \boldsymbol{\mu} \\ \hdashline \boldsymbol{\mu}_* \end{array}\right], \left[\begin{array}{c:c} K_{11} & K_{12} \\ \hdashline K_{21} & K_{22} \end{array}\right]\right)$$

Simply draw from this distribution!

$$\boldsymbol{\mu}_{pred} = \boldsymbol{\mu}_* + K_{21} K_{11}^{-1} \left(\boldsymbol{S} - \boldsymbol{\mu}\right),$$

$$\boldsymbol{\Sigma}_{pred} = K_{22} - K_{21} K_{11}^{-1} K_{12}$$

Kernel Choice

Kernel: function setting correlation between points in input space

- Choice of function: determines properties of sample draws

$$k(d;\boldsymbol{\theta}) = k_{SE}(d;\sigma_0,\ell_0) + k_{Matern,\nu=1/4}(d;\sigma_1,\ell_1),$$

$$d = \left|\log E_i/\mathrm{MeV} - \log E_j/\mathrm{MeV}\right|$$

- "Hyperparameter" optimization: fixes correlation length, amplitude
    - "Leave-Dataset-Out" Cross-Validation: leave out one dataset, fit the GP to the others, evaluate performance on removed dataset

BBN Pipeline

Nuclear reaction data → Gaussian process regression with experimental uncertainties → Draw reaction rate samples

↓

Primordial abundance predictions with uncertainties ← Input samples into BBN code

Are GPs including correlations a biased estimator?

- Including systematics in the covariance matrix biases estimators[5]
  - Is this also true for GPs?
- Are GP 1σ error bars accurate?

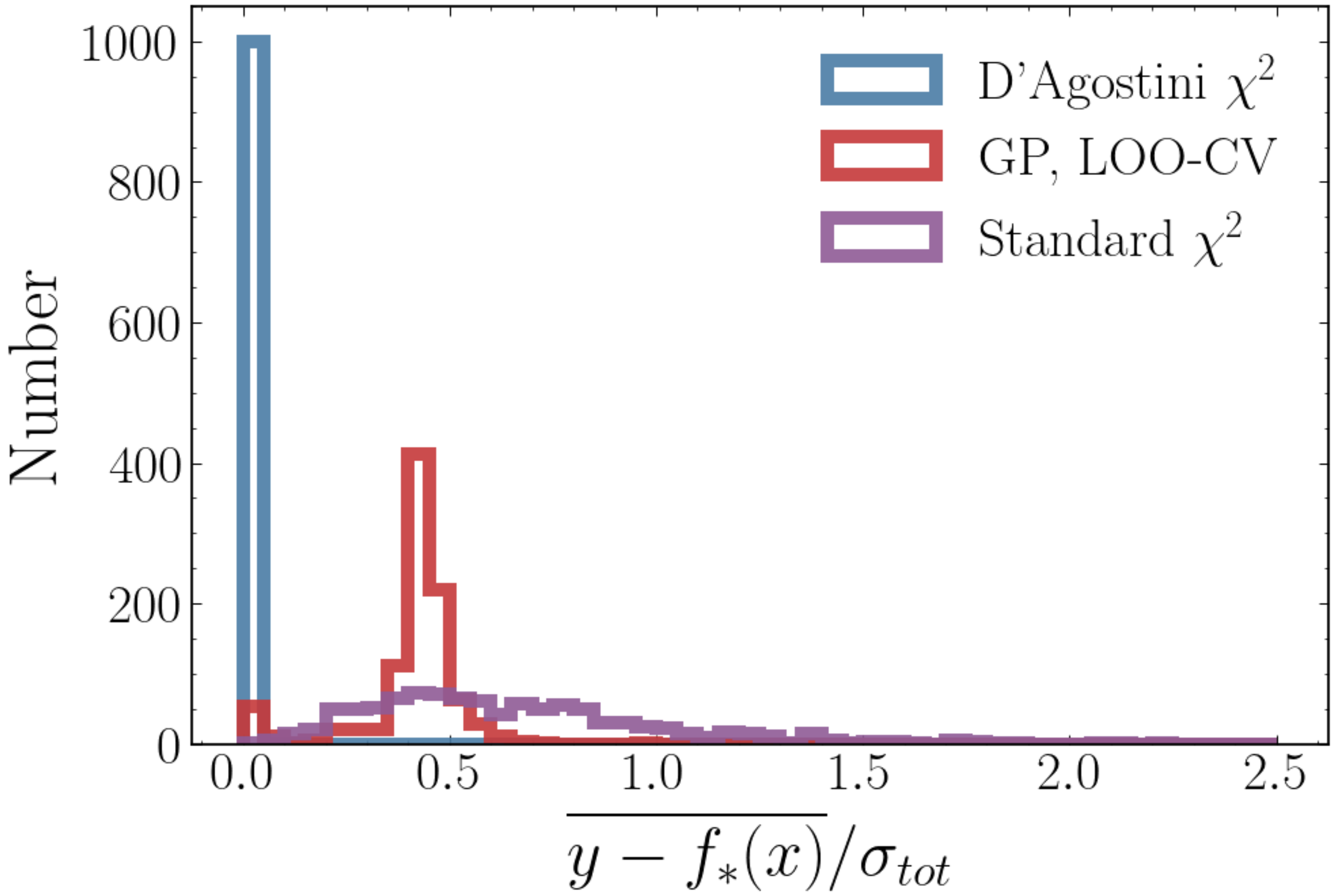
↓

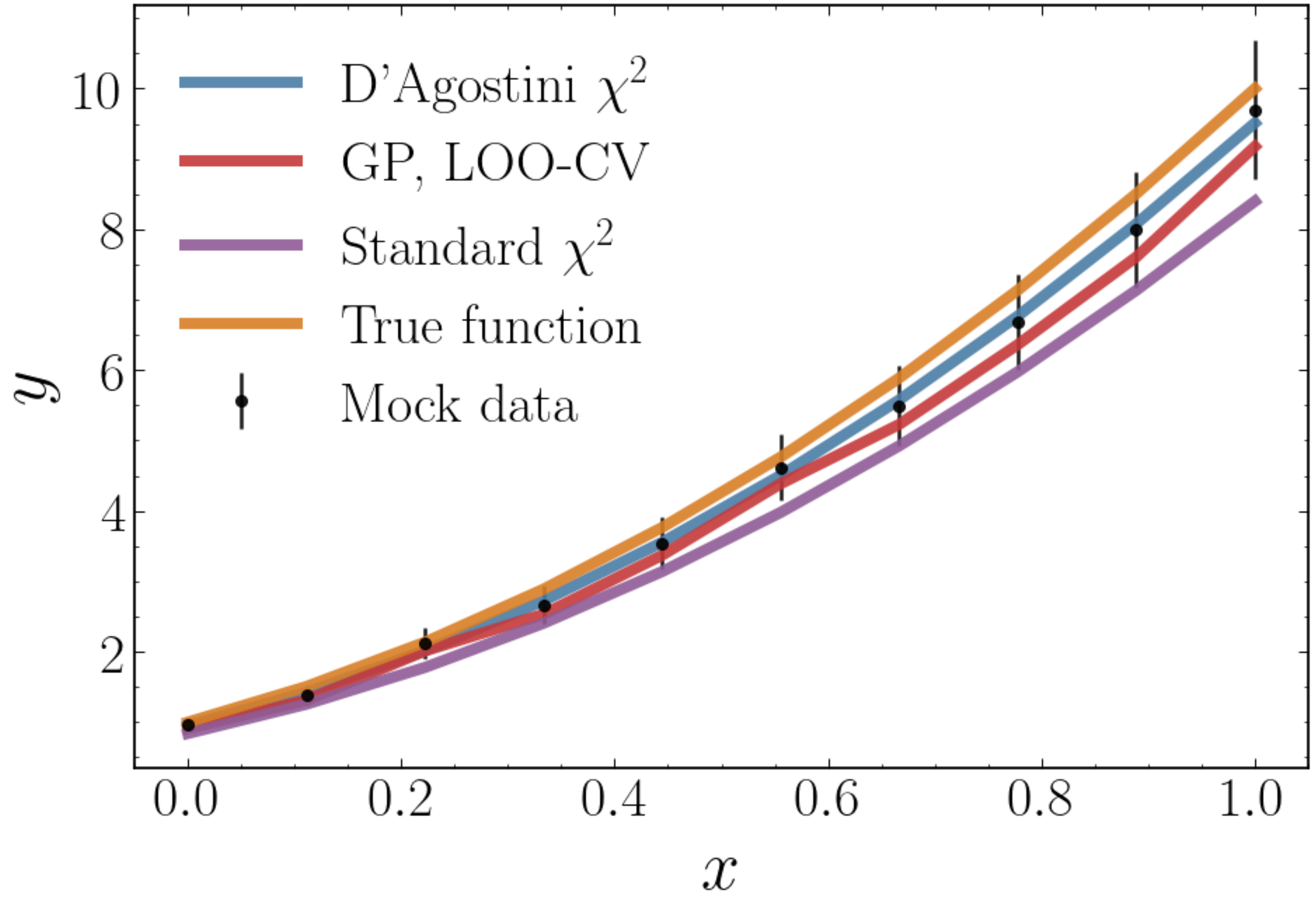
Test both by fitting GPs to mock data

[5] G. D'Agostini, Nucl. Instr. and Meth. in Phys. Res. A 346 (1994)

Are GPs including correlations a biased estimator?

Fits to data generated by a quadratic:





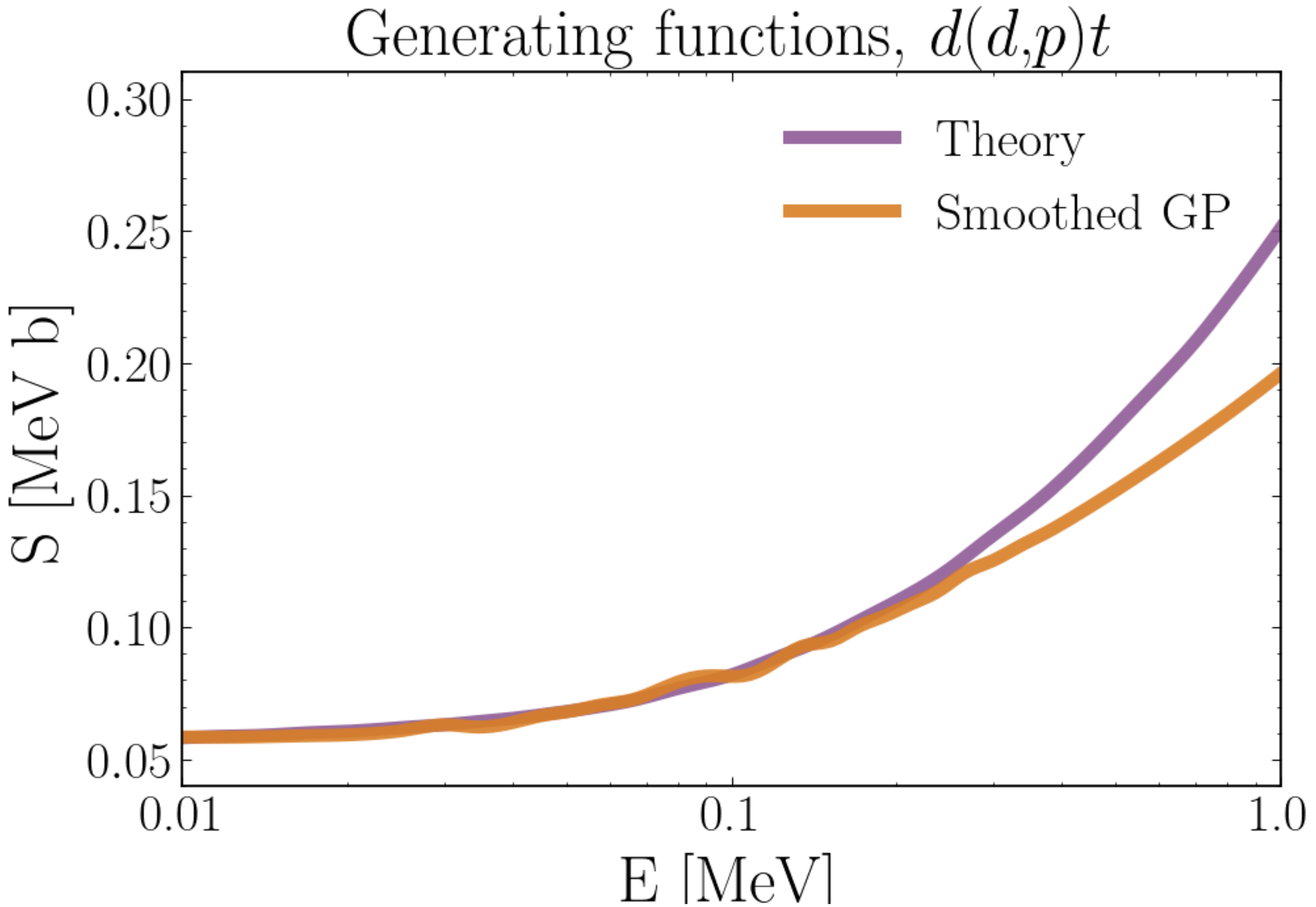
Monte Carlo validation tests: mock data generation

Mock data for point i of experiment k:

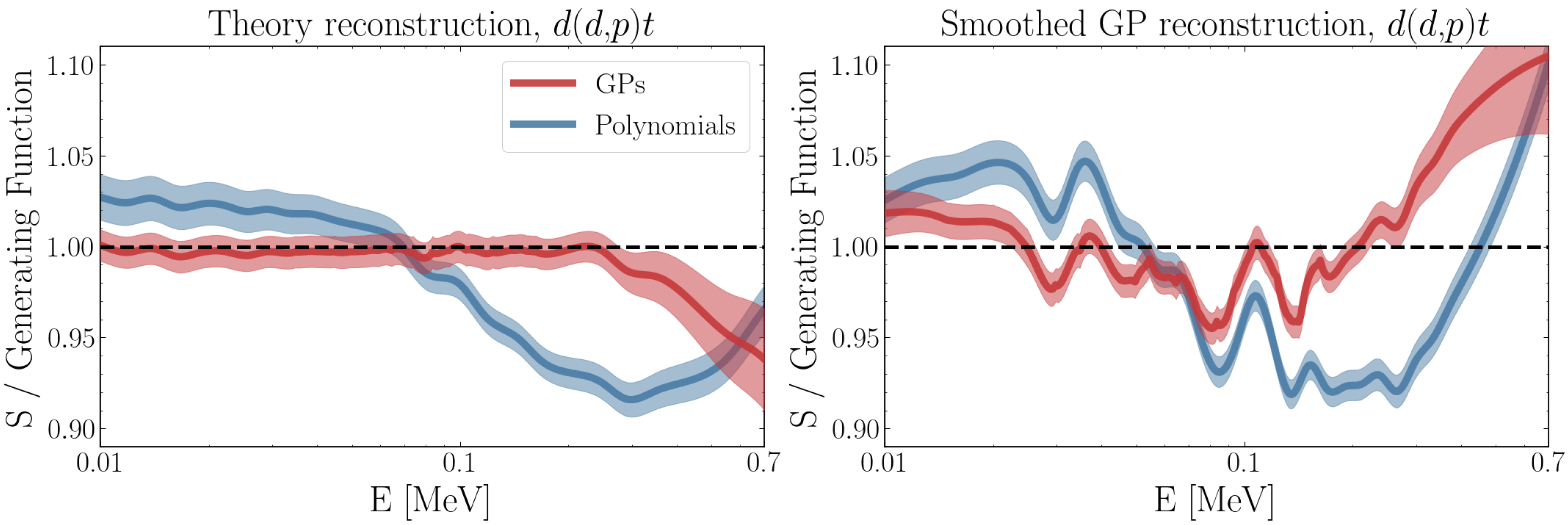
$$S_{ik}(E) = \alpha_k S^*(E)(1 + \beta_{ik}),$$

$$\alpha_k \sim \mathcal{N}(1, \epsilon_k), \quad \beta_{ik} \sim \mathcal{N}(0, \sigma_{ik})$$

Repeat 1,000 times for many realizations!



Monte Carlo validation tests: generating function reconstruction



Monte Carlo validation tests: generating function reconstruction

Monte Carlo validation tests: D/H recovery