Introduction to simulation-based inference (SBI)

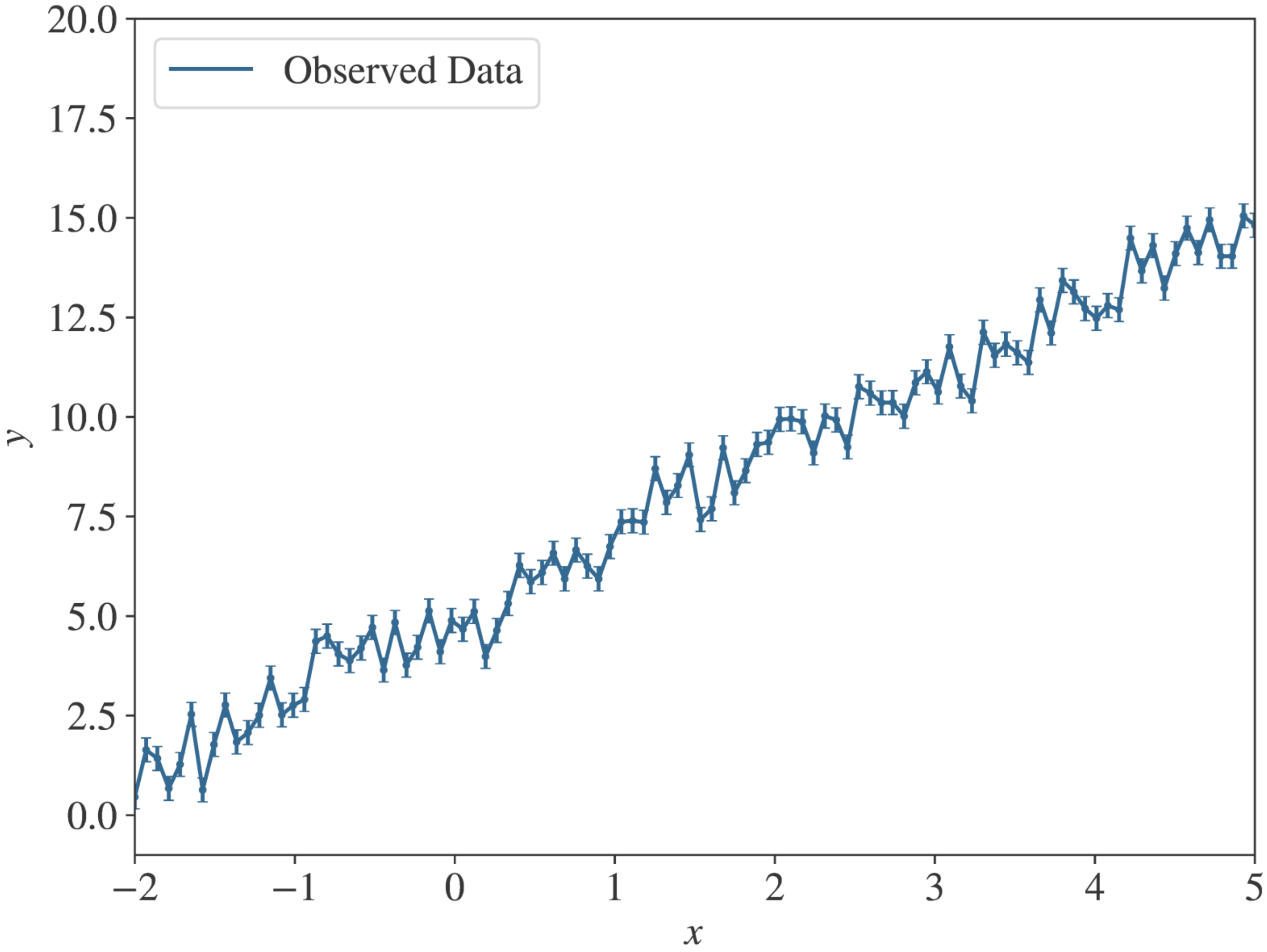
Kiyam Lin

Durham Astrodat, 10/09/2025

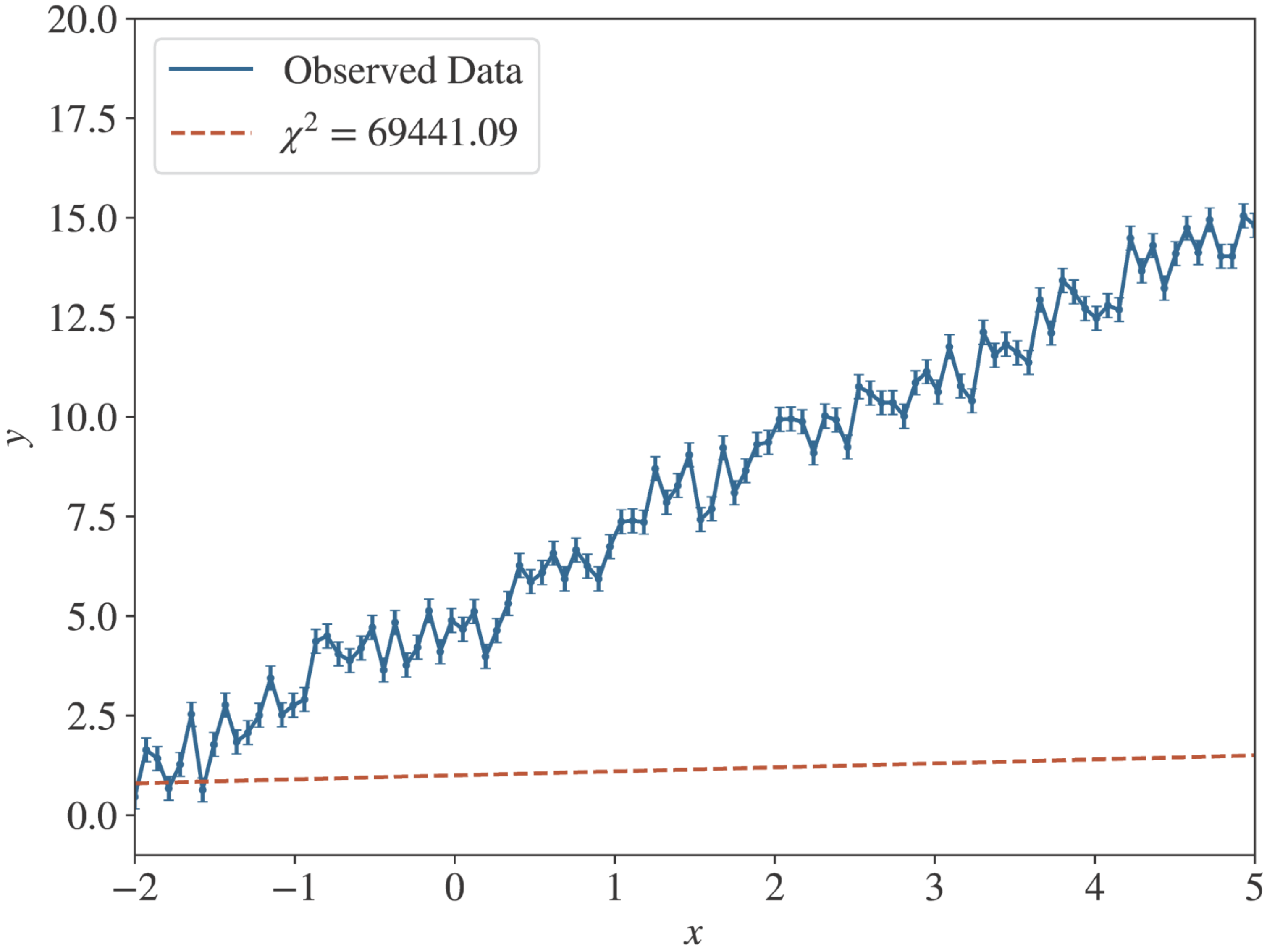
Accurate Likelihoods are hard

- Your likelihood probably is not Gaussian
- Observational systematics
  - Noise models (depth variability etc.)
  - Selection bias
- Complex physics is hard to cast into an analytic model

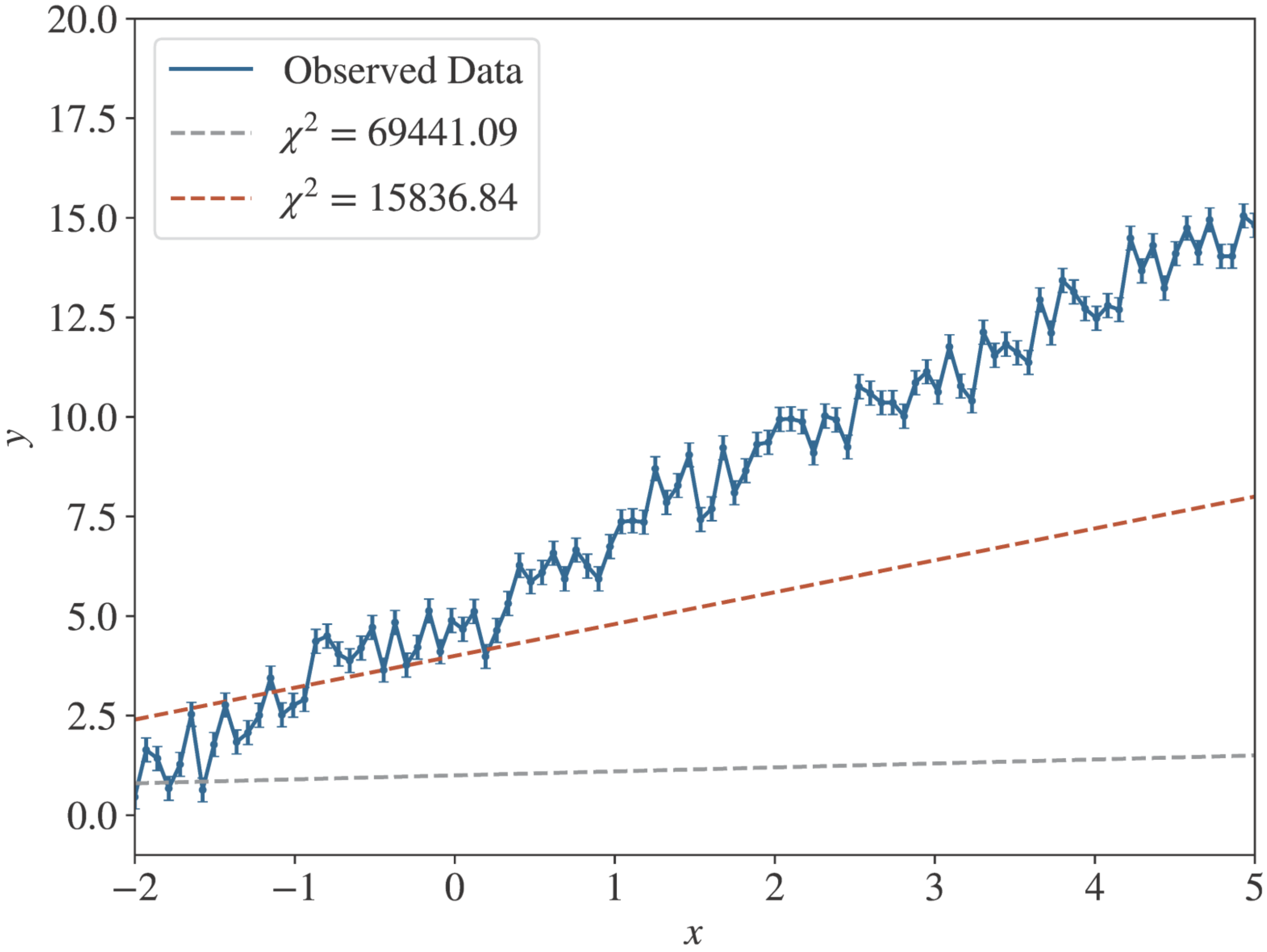
Learn from simulations



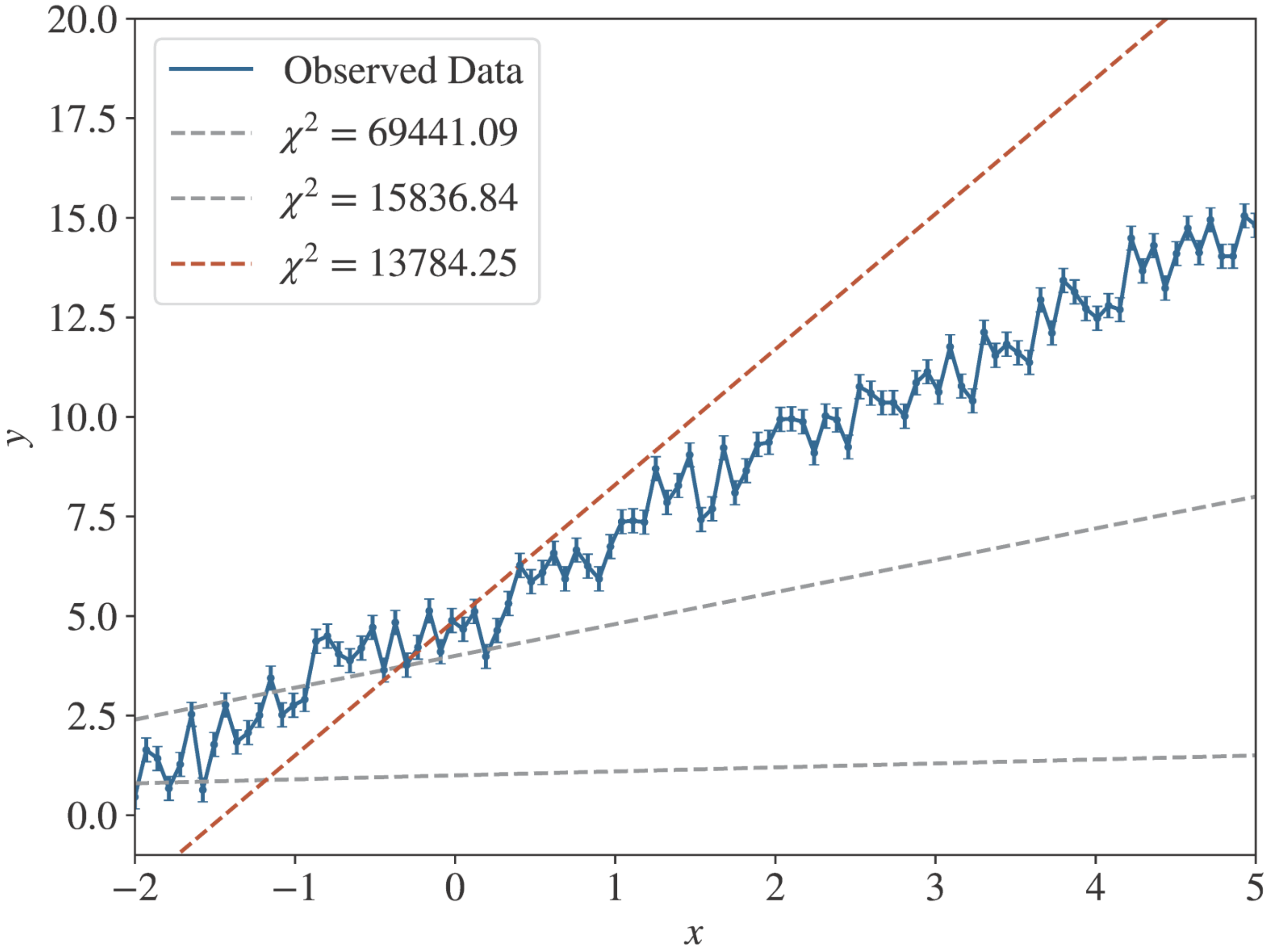
Learn from simulations



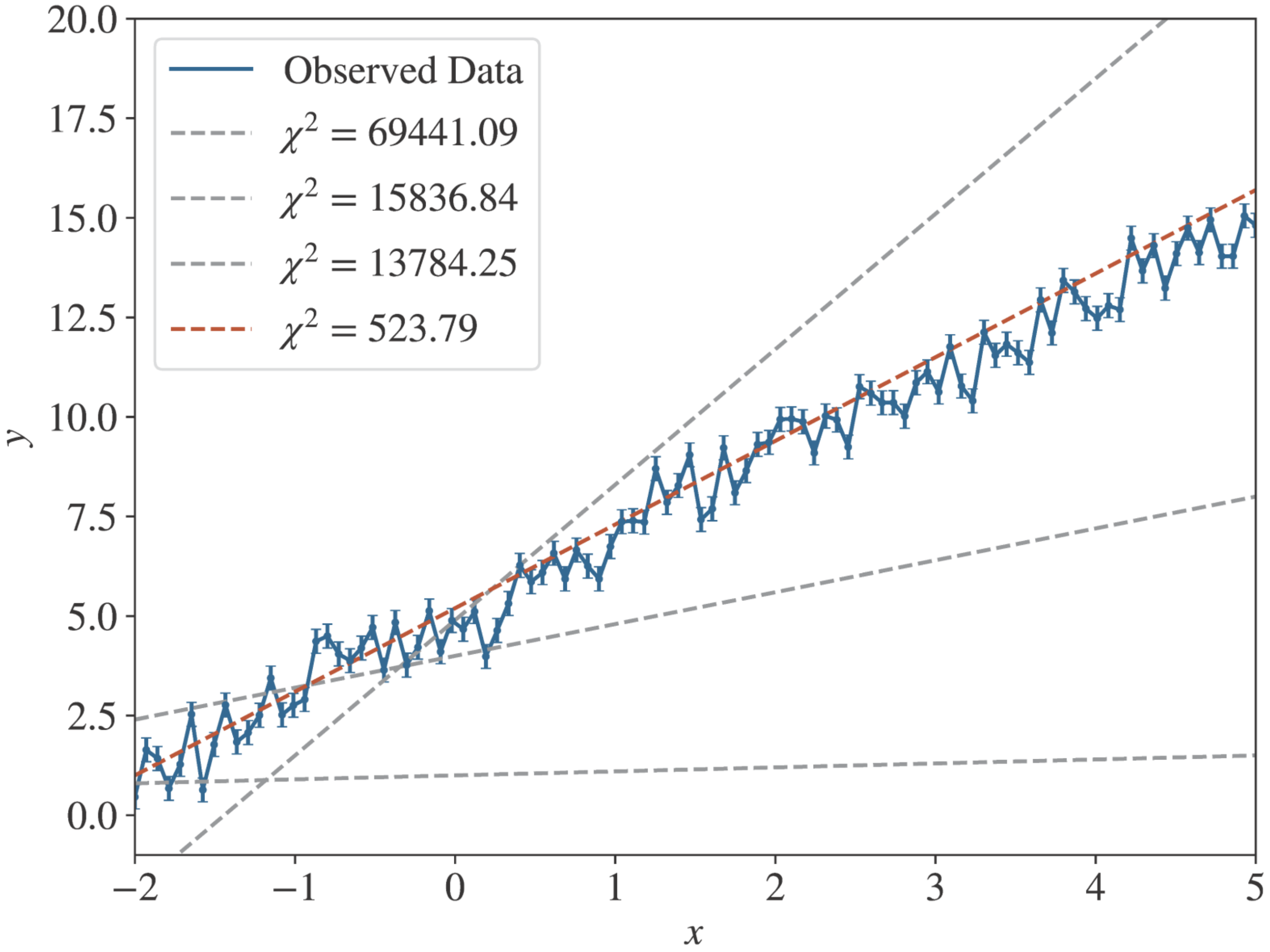
Learn from simulations



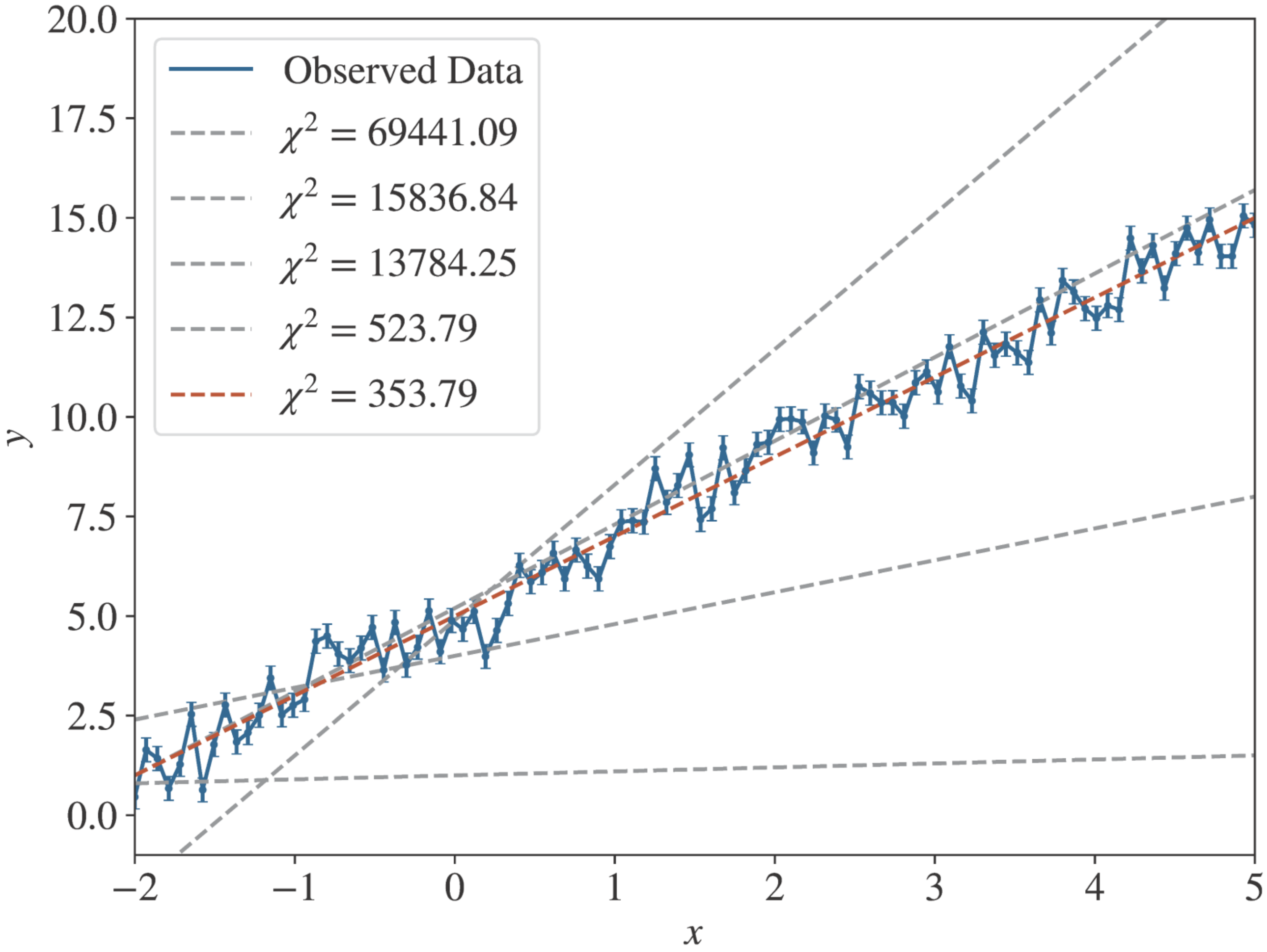
Learn from simulations



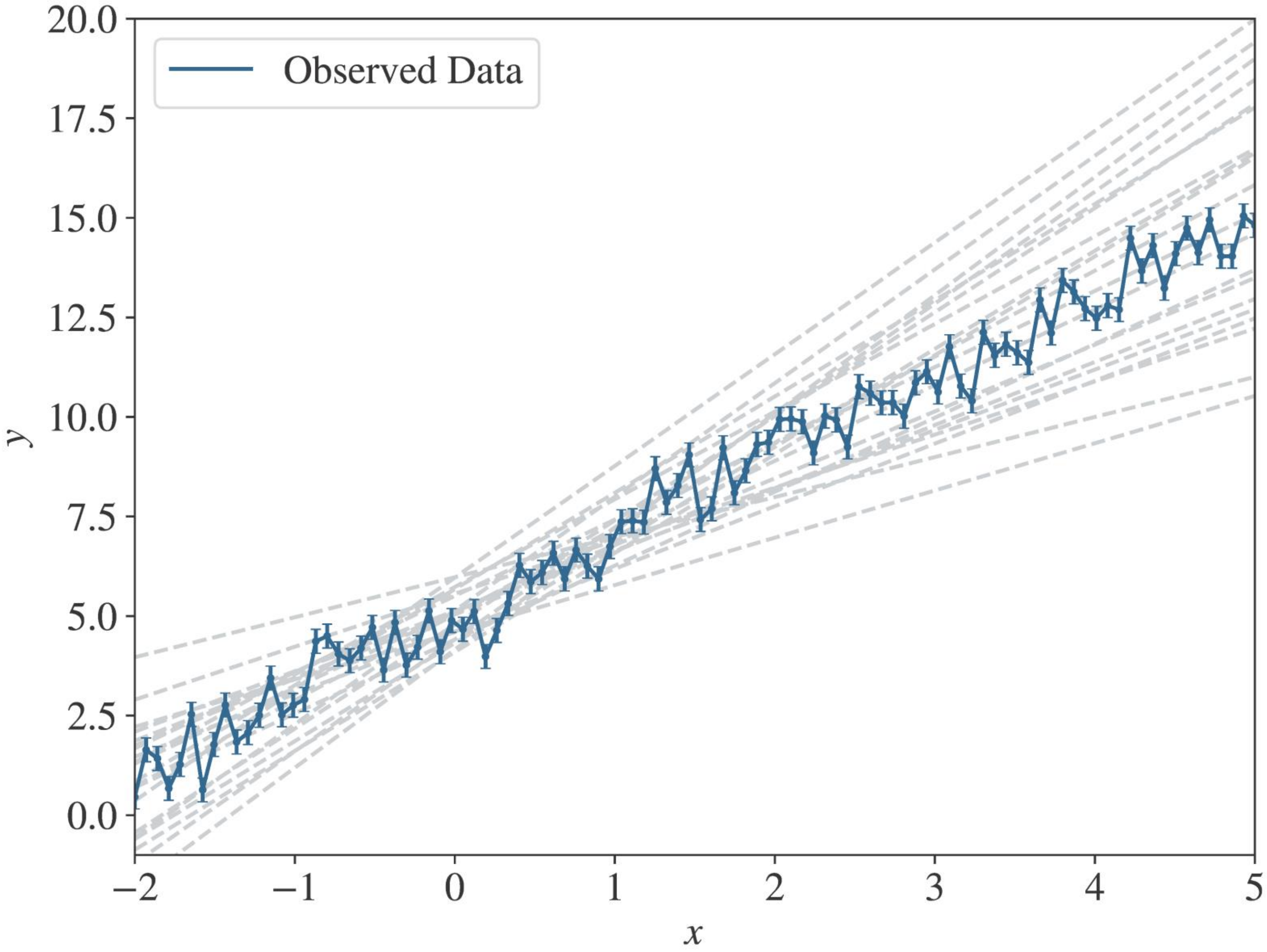
Learn from simulations



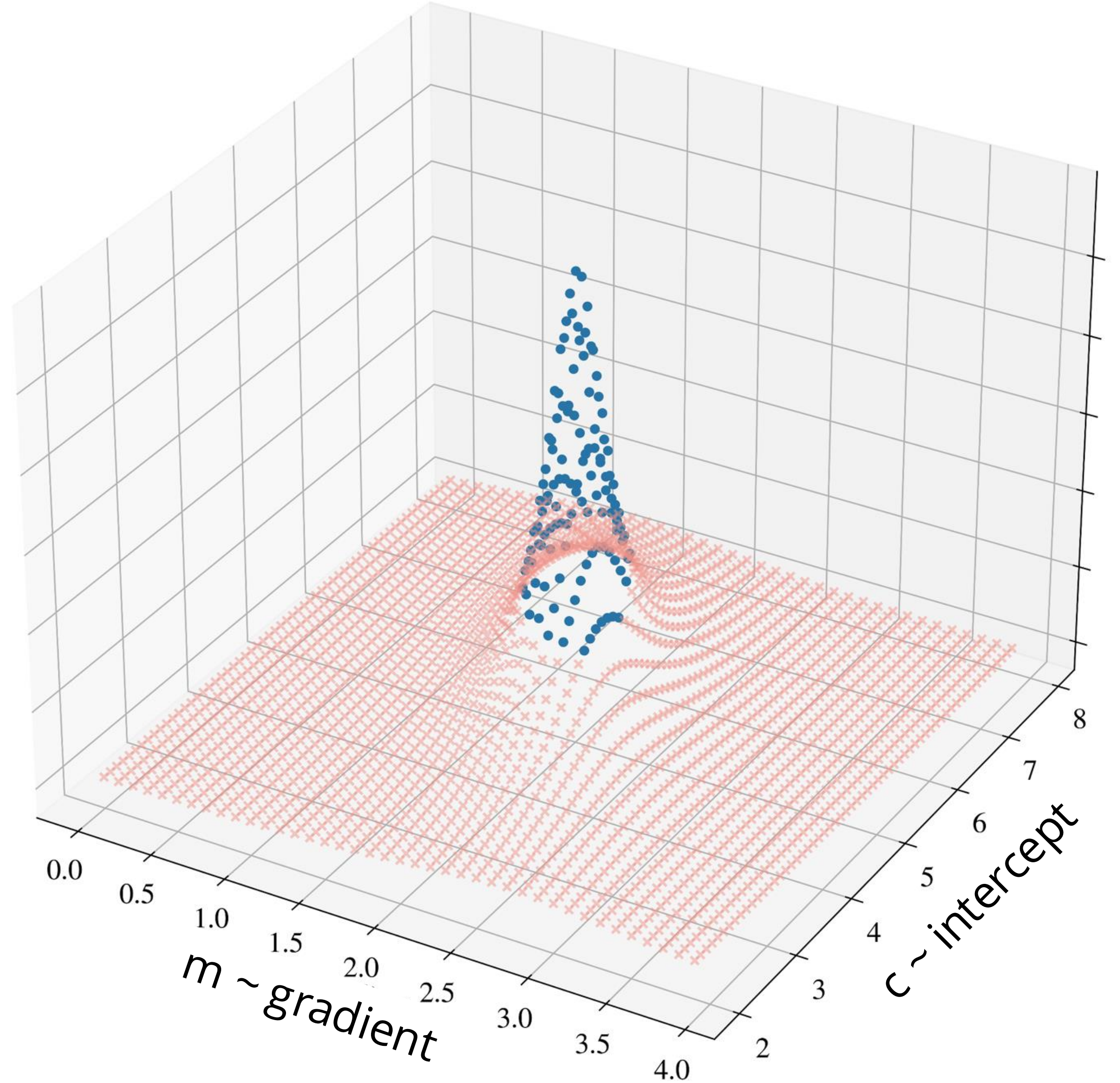
Learn from simulations



Learn from simulations

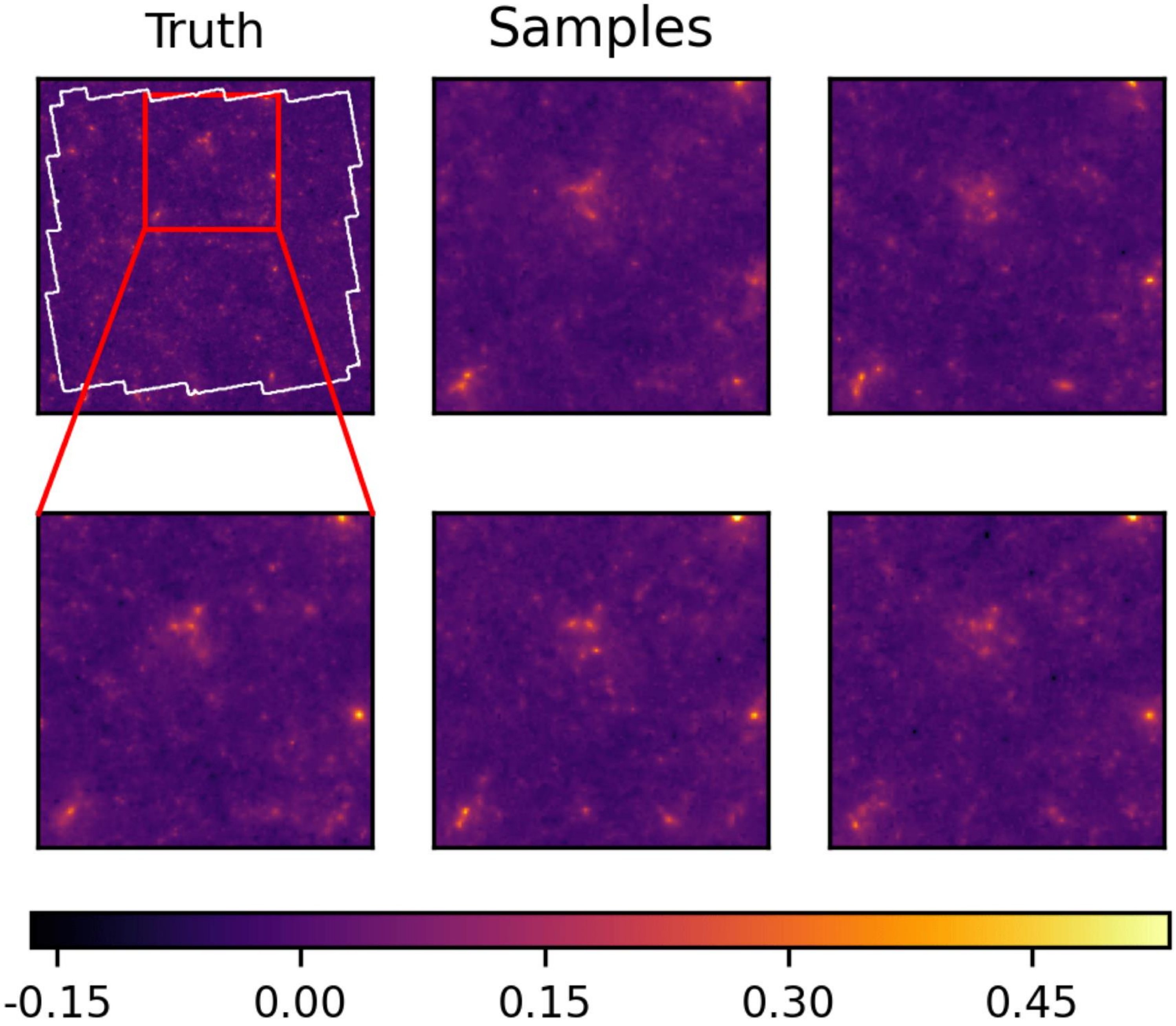


Learn from simulations



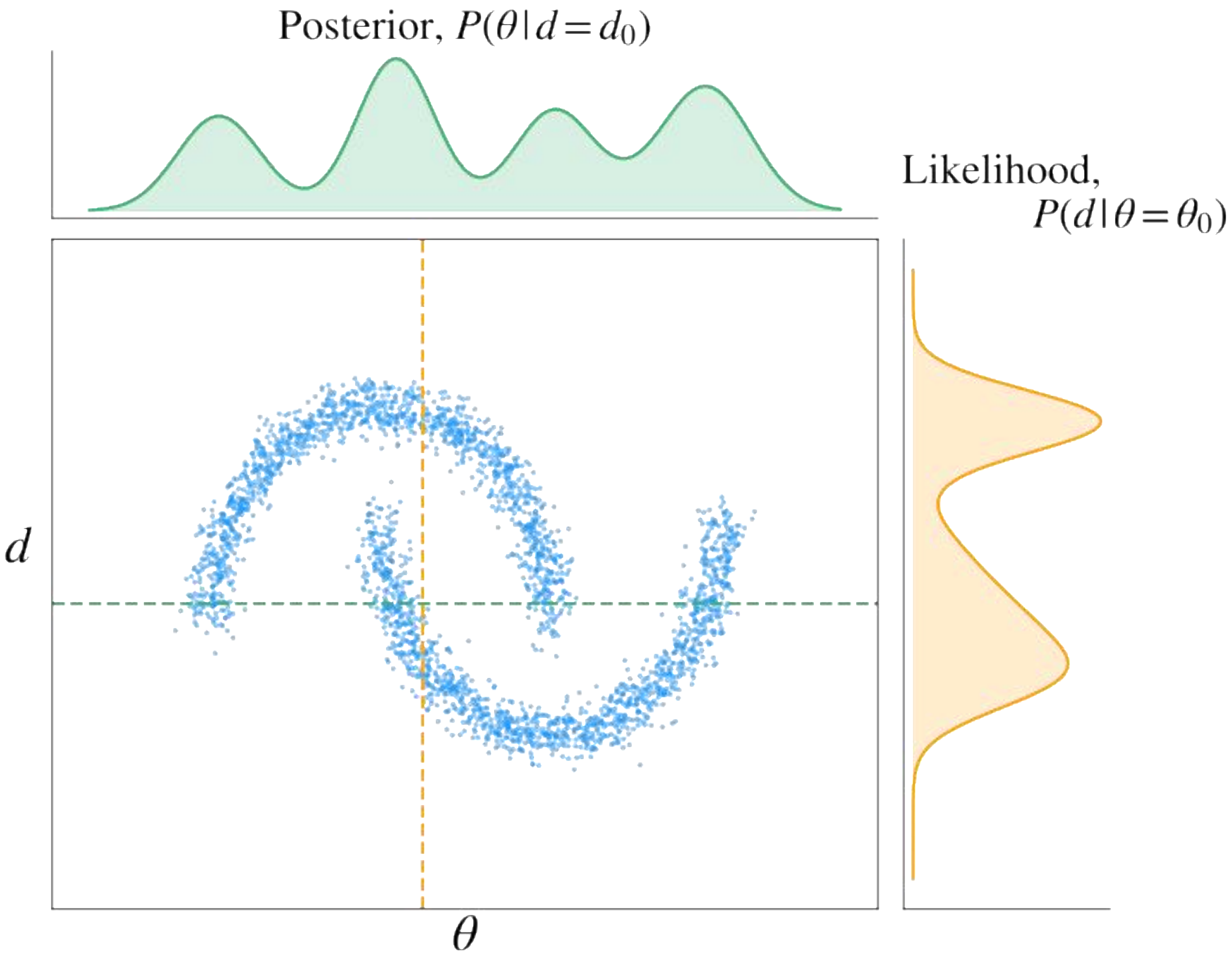
Learn from simulations

- Approximate Bayesian computation (ABC)
  - How do you define closeness
  - Incredibly inefficient



Whitney et al. (2024)

Joint probability is everything $p(\theta, d)$



SBI – Neural methods

- Neural Posterior
- Neural Likelihood
- Neural Ratio

Neural SBI, the premise

- Have a simulator such that one can draw samples from the joint distribution

$$p(\theta, d)$$

SBI – Training

- Train a neural density estimator (NDE) such that $q_\phi(\theta|\mathrm{d}) \rightarrow p(\theta|d)$
- Minimise Kullback-Leibler (KL) Divergence

$$KL\left(p(\theta|d)||q_\phi(\theta|d)\right) = \mathrm{E}_{p(\theta,d)}\left[\log\frac{p(\theta|d)}{q_\phi(\theta|d)}\right]$$

$$KL = \mathrm{E}_{p(\theta,d)}[\log p(\theta|d)] - \mathrm{E}_{p(\theta,d)}\left[\log q_\phi(\theta|d)\right]$$

SBI – Training

$$KL = \mathrm{E}_{p(\theta,d)}[\log p(\theta|d)] - \mathrm{E}_{p(\theta,d)}\left[\log q_\phi(\theta|d)\right]$$

- Minimise the log likelihood of θ

$$\mathcal{L}(\phi) = \mathrm{E}_{p(\theta,d)}\left[-\log q_\phi(\theta|d)\right]$$

SBI – Neural Posterior

- Curse of dimensionality
- $\tilde{p}(\theta|d) \propto \frac{\tilde{p}(\theta)}{p(\theta)} p(\theta|d)$

SBI – Neural Posterior

- $\tilde{p}(\theta|d) \propto \frac{\tilde{p}(\theta)}{p(\theta)} p(\theta|d)$
- NPE-A – Papamakarios & Murray 2016
  - Post-hoc analytical correction needed
- NPE-B – Lueckmann et al. 2017
  - Proposal embedded as an importance weight
- NPE-APT – Greenberg et al. 2019
  - Automatic Posterior Transformations

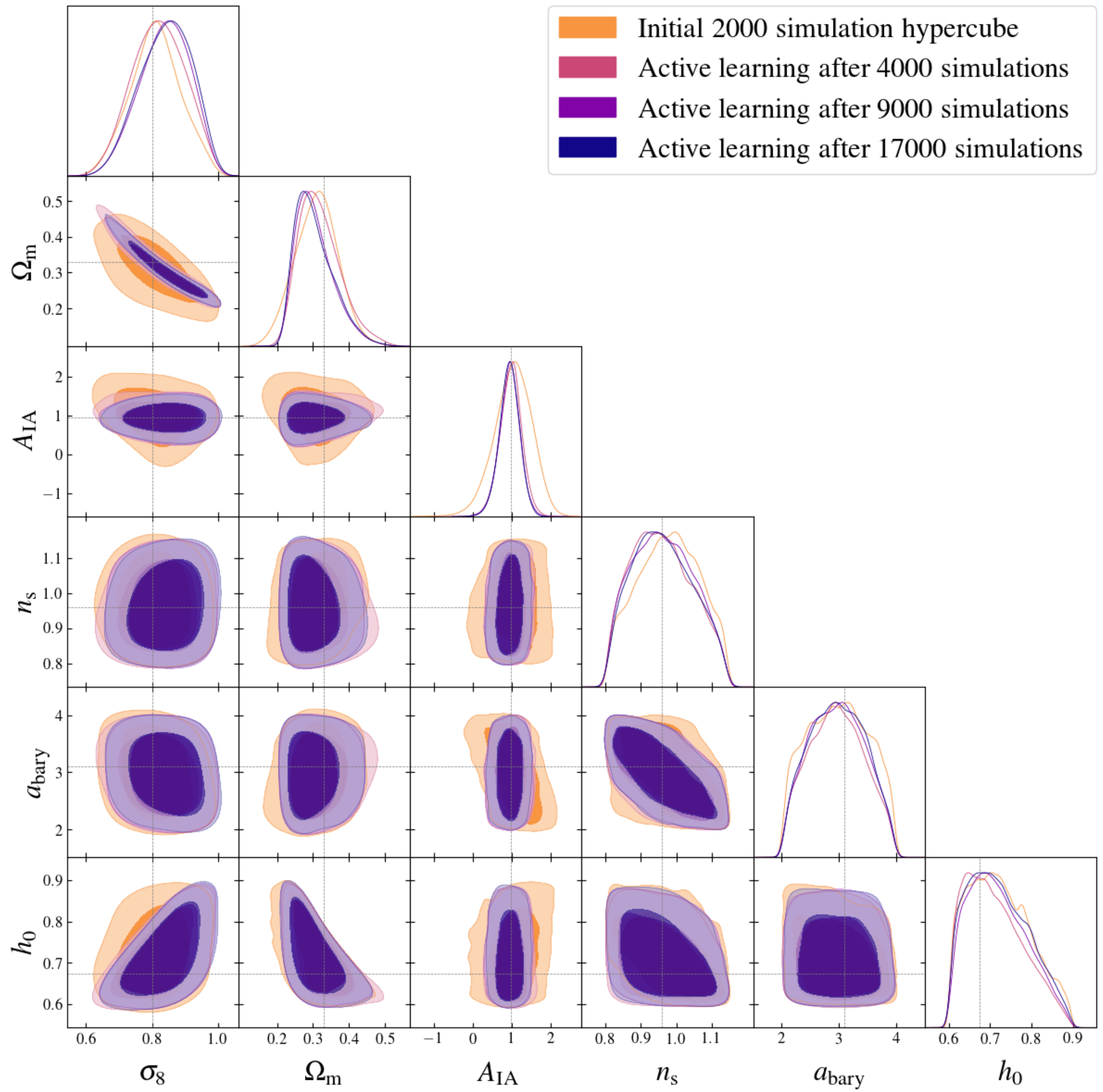
SBI – Neural Posterior

- NPE-APT is the standard approach now
- Amortized for data, but is prior dependent
- Can be sequential

SBI jargon

- Amortization
  - Learned model *not* specific to one data vector
- Sequential
  - Adaptively acquire simulations

Active learning posteriors



SBI – Neural Likelihood

- Learn conditional likelihood as a function of θ
- Minimise the conditional log likelihood

$$\mathcal{L}(\phi) = \mathrm{E}_{p(\theta,d)}\left[-\log q_{\phi}(d|\theta)\right]$$

SBI – Neural Likelihood

- Can draw simulations any way we want
- Gives access to the likelihood!
- Need sampling for final posterior
- Amortized
- Can be sequential

SBI - Neural Ratio

- Train a classifier to distinguish samples drawn from the joint distribution vs. marginal distribution
- Likelihood ratio trick:

$$r_\phi(d|\theta) = \frac{\boldsymbol{d}^*(d,\theta)}{1-\boldsymbol{d}^*(d,\theta)} \approx \frac{p(\theta|d)}{p(\theta)}$$

- Need sampling for final posterior

Hermans et al. (2020)

SBI – Neural Ratio

- In practice, little overlap between joint and marginal distributions to use as training data set.
- Truncated marginal neural ratio estimation exists to only learn the marginals
- Can be amortized
- Can be sequential

Hermans et al. (2020)

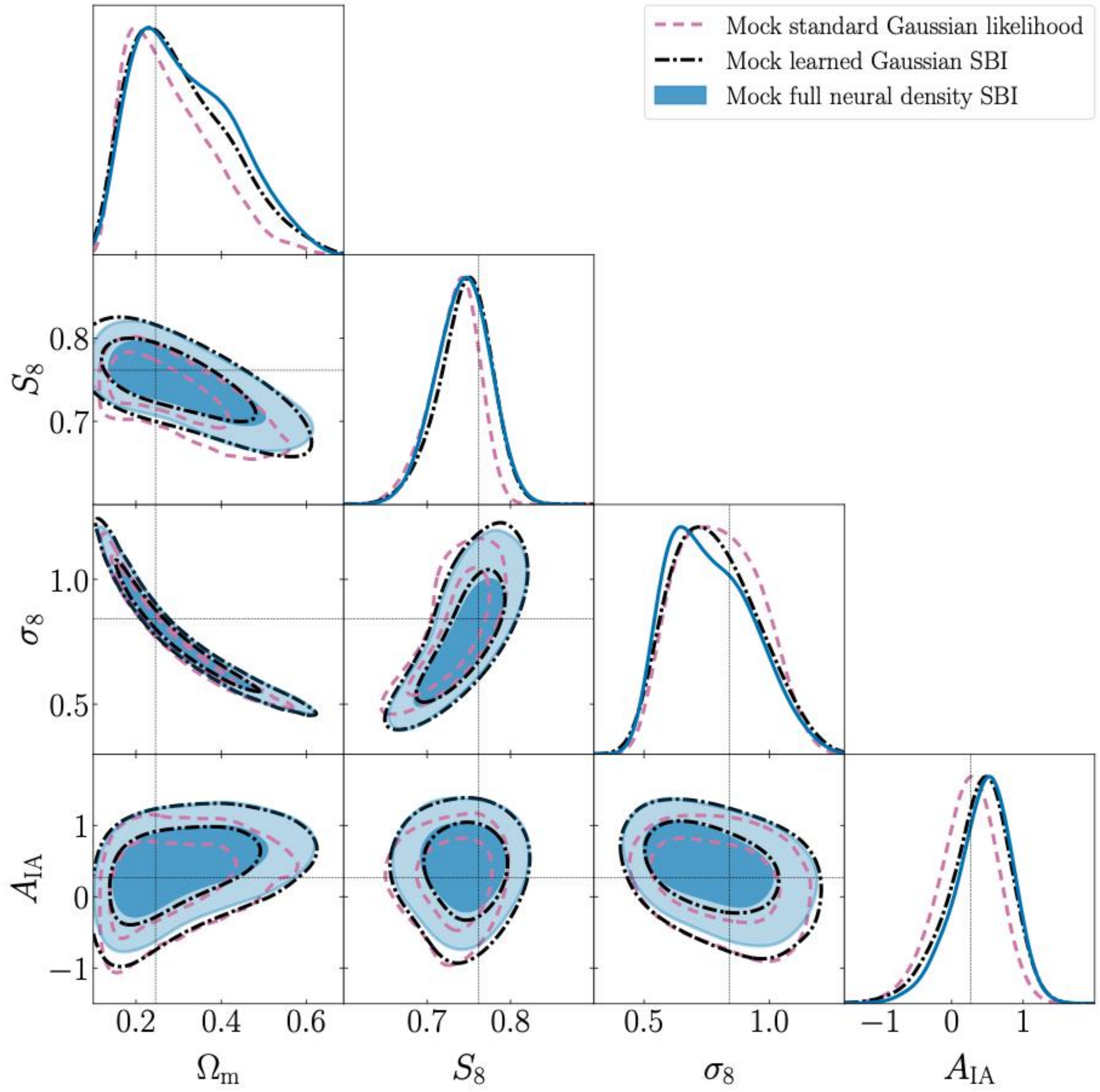
SBI – Software

Github links	NPE	NLE	NRE
https://github.com/sbi-dev/sbi	✅	✅	✅
https://github.com/probabilists/lampe	✅	❌	❌
https://github.com/maho3/ltu-ili	✅	✅	✅
https://github.com/undark-lab/swyft	❌	❌	✅ TMNRE
https://github.com/justinalsing/pydelfi	❌	✅	❌

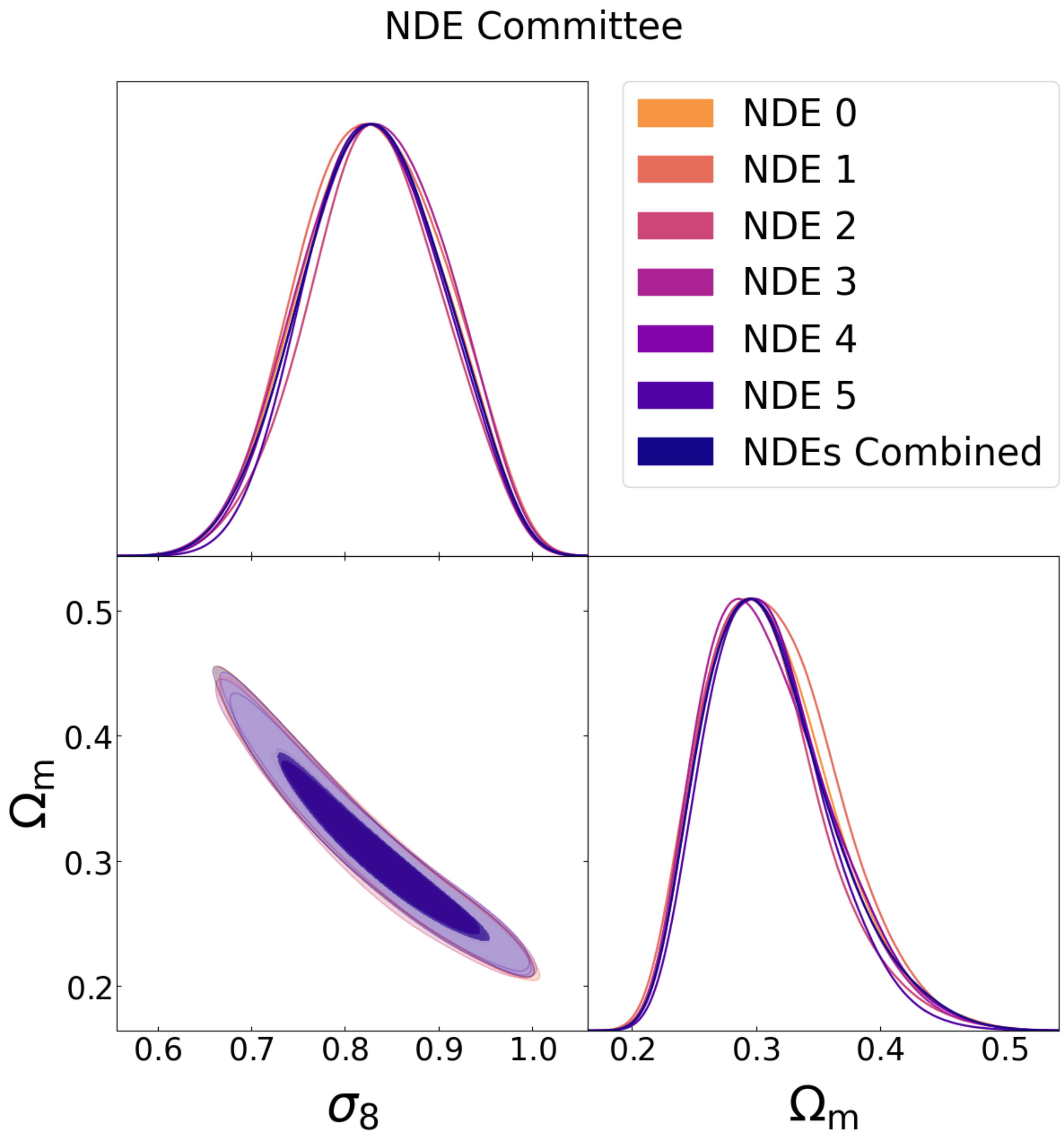
SBI – NDEs

- Mixture density networks
- Masked autoregressive flows
- Neural spline flows
- Real-valued non-volume preserving flows

Papamakarios et al. (2021)



Mock standard Gaussian likelihood
Mock learned Gaussian SBI
Mock full neural density SBI
Ω_m
S_8
σ_8
A_IA
0.2
0.4
0.6
0.7
0.8
0.5
1.0
−1
0
1

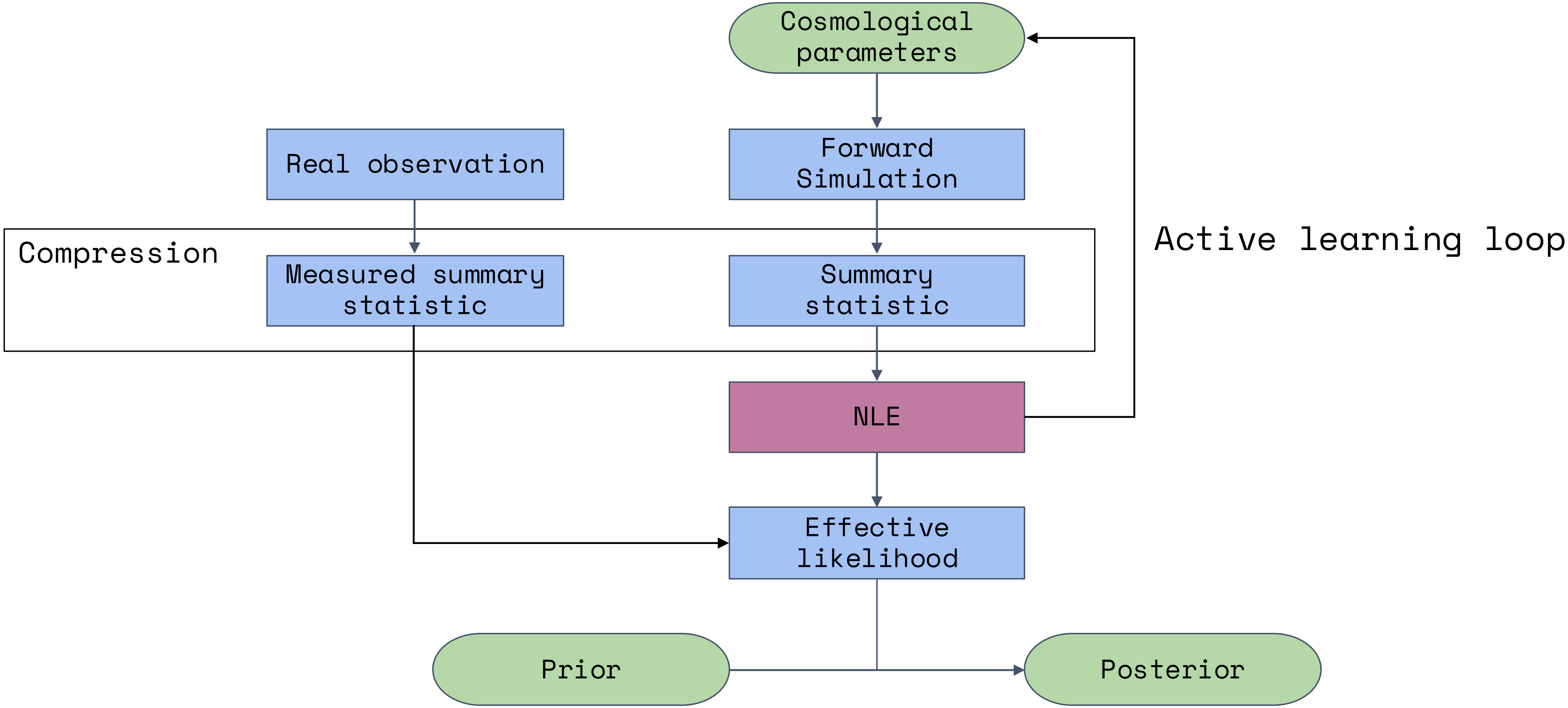


NDE Committee
NDE 0
NDE 1
NDE 2
NDE 3
NDE 4
NDE 5
NDEs Combined
0.5
0.4
0.3
0.2
Ω_m
0.6
0.7
0.8
0.9
1.0
σ_8
0.2
0.3
0.4
0.5
Ω_m

SBI – Compression

- Analytical compression
  - MOPED
  - Score
  - CCA
  - PCA
- Neural compression
  - IMNN
  - Fishnets

SBI – The loop


Cosmological parameters
Real observation
Forward Simulation
Compression
Measured summary statistic
Summary statistic
Active learning loop
NLE
Effective likelihood
Prior
Posterior

SBI – Challenges

- Dimensionality
- Speed
- Validation and Accuracy

SBI – Evidence

- $p(\theta|d,\mathcal{M}) = \frac{p(d|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(d|\mathcal{M})}$
- $\frac{p(\mathcal{M}_1|d)}{p(\mathcal{M}_2|d)} = \frac{p(d|\mathcal{M}_1)p(\mathcal{M}_1)}{p(d|\mathcal{M}_2)p(\mathcal{M}_2)}$
- Nested Sampling
- Learned Harmonic Mean (McEewen et al 2021)
- Evidence Networks (Jeffrey & Wandelt 2023)

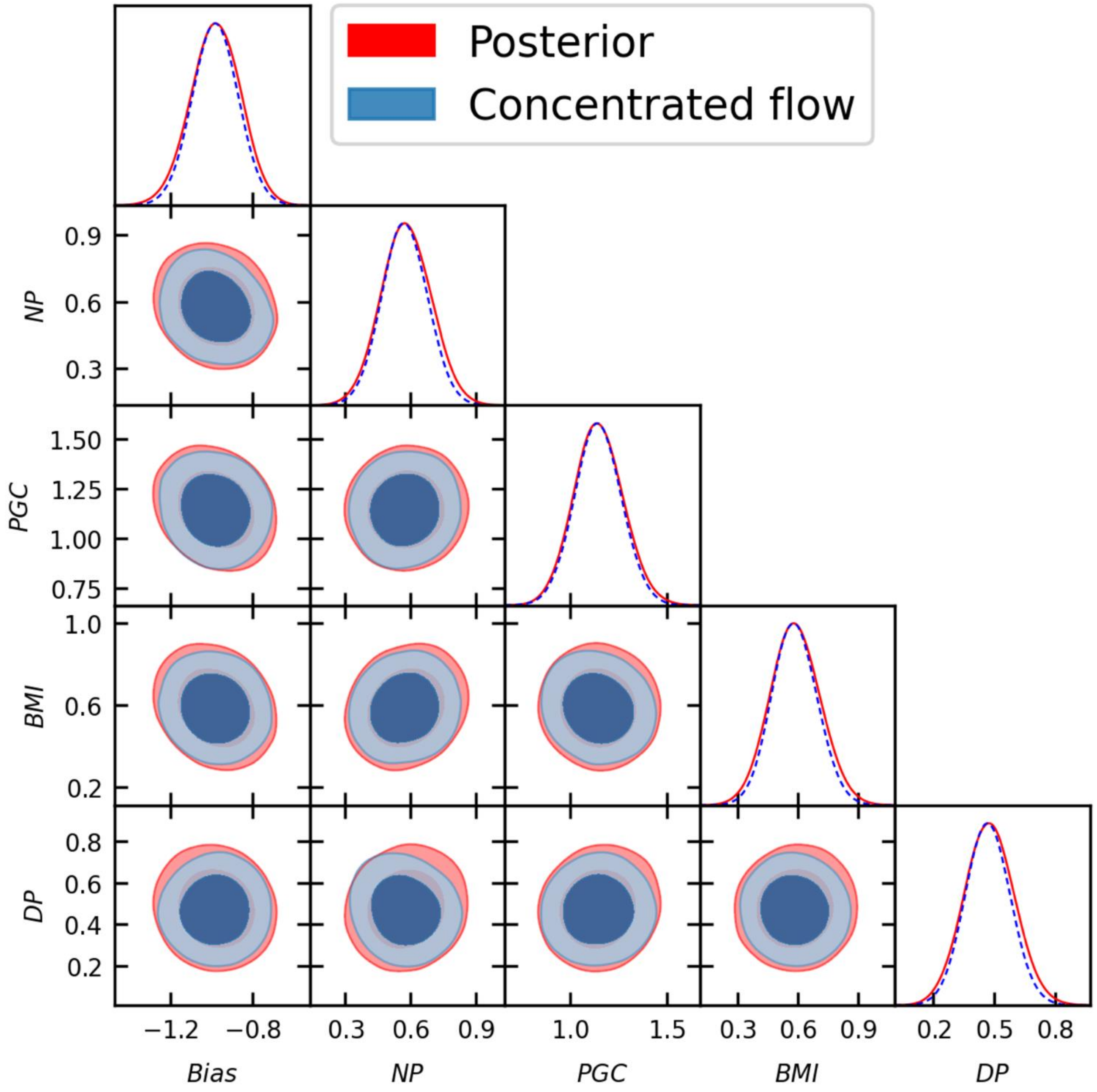
SBI – Evidence

- Learned Harmonic mean –
  - Get evidence directly from posterior samples!

In practice, a shrunk posterior!

$$\frac{1}{z} = \frac{1}{N}\sum_{i=1}^{N} \frac{\varphi(\theta_i)}{\mathcal{L}(d|\theta_i)\pi(\theta_i)}, \qquad \theta_i \sim p(\theta|d)$$

Mcewen et al. (2021), Polanska et al. (2024)



Polanska et al. (2024)

Savage-Dickey density ratio

- SDDR for nested models

$$\frac{z_1}{z_2} = \frac{p(\eta_1|d, M_2)}{\pi(\eta_1|M_2)}$$

- For high dimensional marginals, can make use of normalizing flows!
- Both implemented in harmonic package:
  - https://github.com/astro-informatics/harmonic

Demo notebook

https://github.com/Kiyam/astrodat-2025

Thank you for listening, questions?