Simulation-Based Inference With Quantile Regression

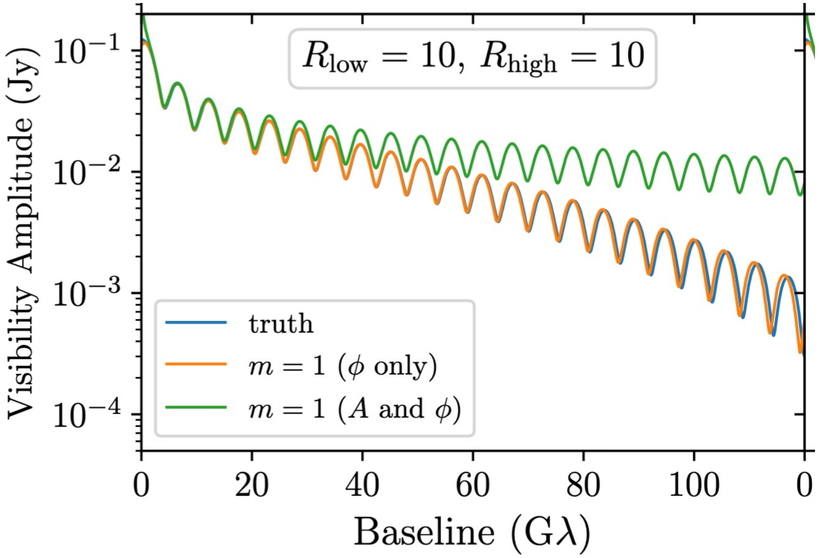
He Jia (贾赫)

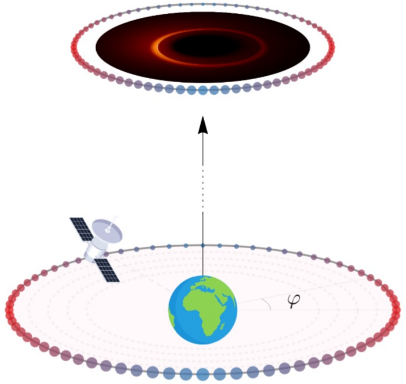
Princeton University

Apr 9, 2024 @ Berkeley

About Me

- I was a visiting undergrad with Uros a few years ago
- I'm now a 4th year grad student, working on BHs (with Eliot Quataert) for my thesis (how to measure e.g. the spin of M87*)
- But I'm still working on cosmology (with David Spergel) when I can find time
- Today I want to advertise a new SBI method for cosmology
- I will stay here for the rest of the week





Gralla et al, 2008.03879; **Jia** et al, to be submitted

Simulation-Based Inference

- Simulation-Based Inference (SBI), aka Likelihood-Free Inference (LFI), Implicit-Likelihood Inference (ILI)
- Given the model parameters θ and simulated data x
- For example, θ is $(\Omega_m,\ \sigma_8)$, x is the weak lensing map
- SBI: fits *something* in Bayes' theorem $p(\theta|x)\,p(x) = p(x|\theta)\,p(\theta)$ with Neural Networks (NN)

Simulation-Based Inference

- SBI: fits *something* in Bayes' theorem $p(\theta|x)\,p(x) = p(x|\theta)\,p(\theta)$ with Neural Networks (NN)
- Neural Posterior Estimation (NPE): fits posterior $p(\theta|x)$ with Normalizing Flows (NF)
- Neural Likelihood Estimation (NLE): fits likelihood $p(x|\theta)$ with NF
- Neural Ratio Estimation (NRE): fits the ratio $\frac{p(\theta,x)}{p(\theta)p(x)}$ with NN
- NEW: Neural Quantile Estimation (NQE)

Why NOT SBI? Your SBI can be biased because…

- You have the *correct* simulator, your simulation budget is limited
- You have the *correct* simulator (e.g. Illustris) and a fast emulator (e.g. n-body), you can only afford to run many simulations with the fast emulator
- You have a fast emulator, you assume the *correct* simulator is among several candidates (e.g. within CAMELS)
- ~~You do not know what the *correct* simulator is at all~~
- Our new NQE method helps in the first three scenarios!
- Guaranteed to be unbiased if you have 500-1000 runs from the *correct* simulator, regardless of the dimensionality of the problem

Neural Posterior Estimation (NPE)

- NPE: fits posterior $p(\theta|x)$ with Normalizing Flows (NF)
- NF: a special NN, for each x, outputs a bijective transformation between $p(\theta|x)$ and Gaussian

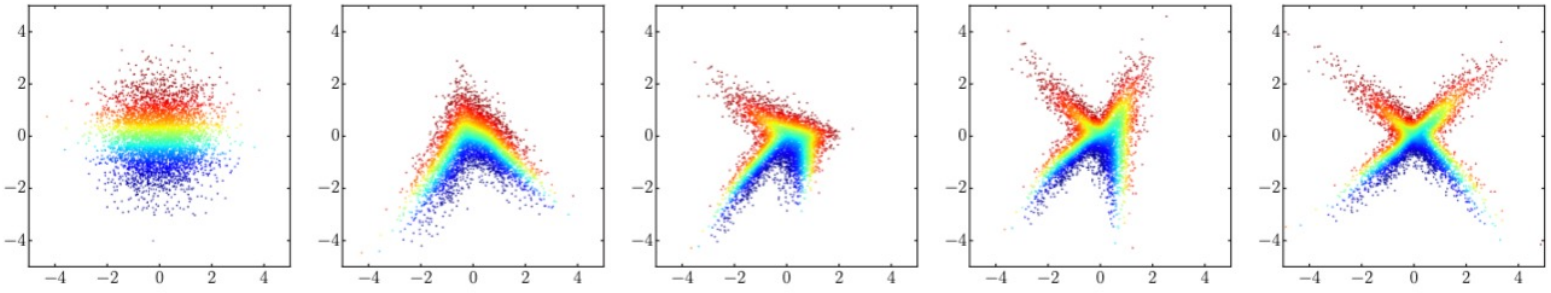
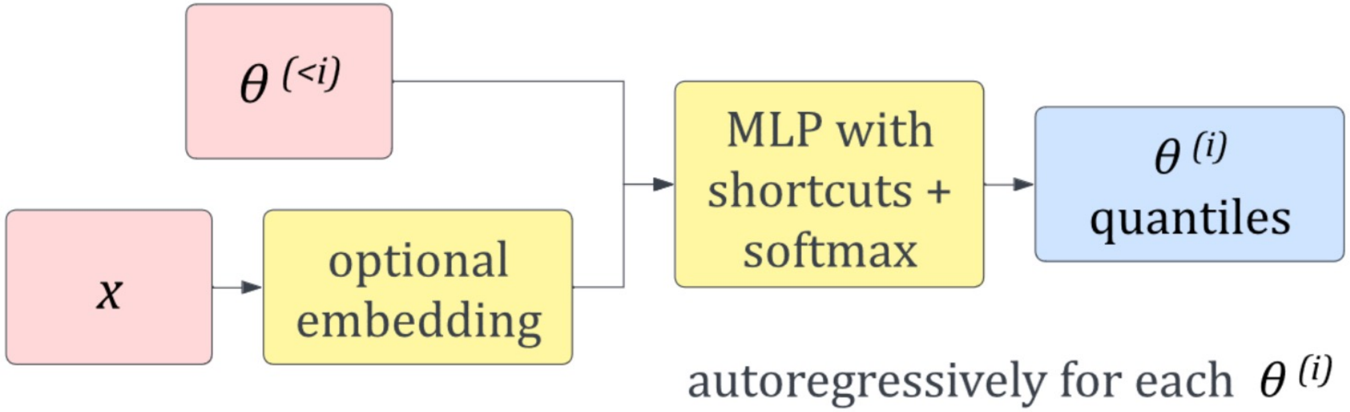


Figure 1: Example of a 4-step flow transforming samples from a standard-normal base density to a cross-shaped target density.

Papamakarios et al, 1912.02762

Neural Quantile Estimation (NQE)

- Learns quantiles for each 1-dim conditional $p(\theta^{(i)}|\theta^{(j<i)}, x)$
- Autoregressive structure: $p(\theta|x) = p\left(\theta^{(1)}\middle|x\right) \times p\left(\theta^{(2)}\middle|\theta^{(1)}, x\right) \times \cdots$
- L2 loss => mean ; L1 loss => median ; weighted L1 loss => arbitrary quantiles



Jia, 2401.02413

Neural Quantile Estimation (NQE)

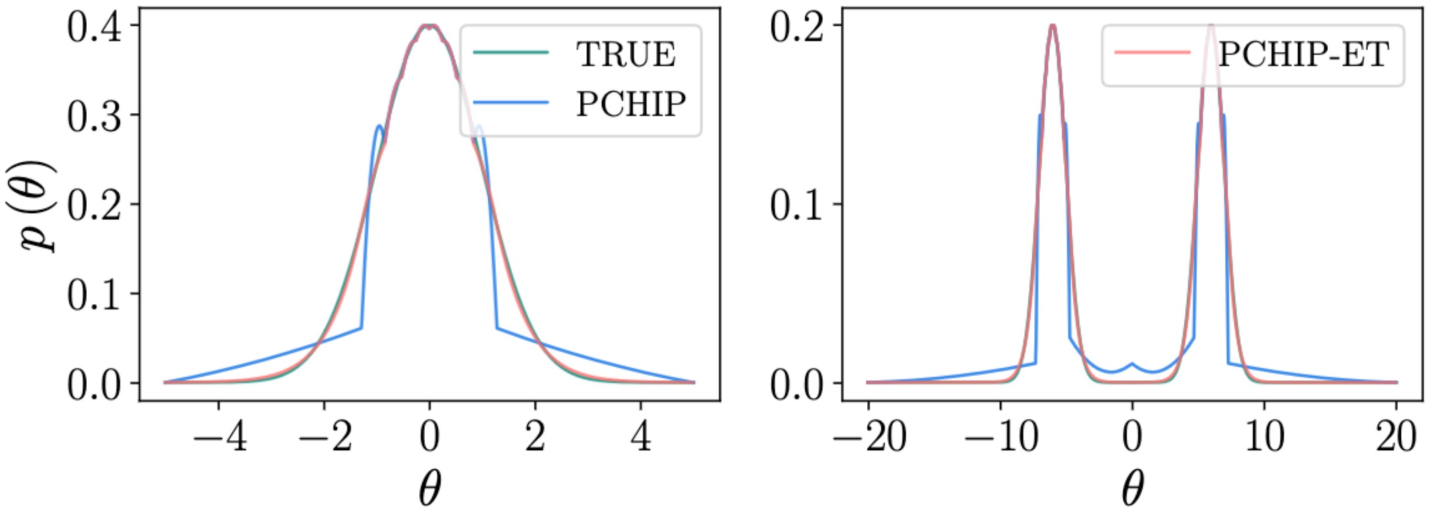
- WL example: weak lensing map => CNN => Ω_m
- With L1 loss ➔ median of Ω_m posterior
- With L2 loss ➔ mean of Ω_m posterior
- With weighted L1 loss ➔ arbitrary quantiles of Ω_m posterior
- We can reconstruct a 1-dim distribution with ~15 quantiles

$$\mathcal{L}_\tau[\theta,\, F_\phi(\mathbf{x})] \equiv (\tau - 1) \sum_{\theta < F_\phi(\mathbf{x})} w(\mathbf{x})\, [\theta - F_\phi(\mathbf{x})] + \tau \sum_{\theta \geq F_\phi(\mathbf{x})} w(\mathbf{x})\, [\theta - F_\phi(\mathbf{x})]\,. \tag{1}$$

Jia, 2401.02413

Neural Quantile Estimation (NQE)

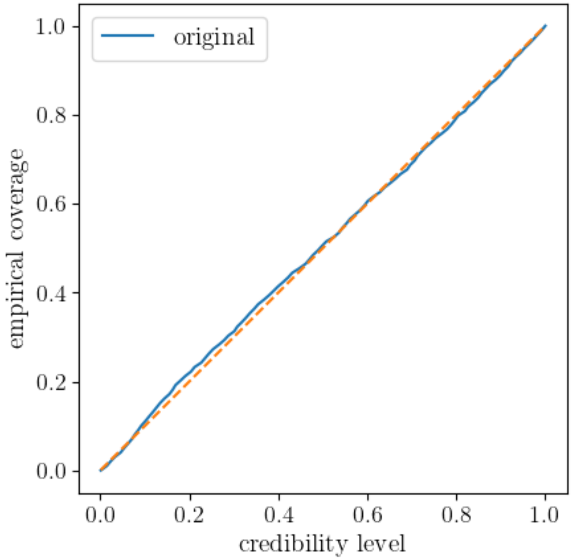
- We interpolate the CDF, which should be monotonic and continuous
- Piecewise Cubic Hermite Interpolating Polynomial with Exponential Tails (PCHIP-ET)
- Perfectly reconstructs a 1-dim distribution with ~15 quantiles

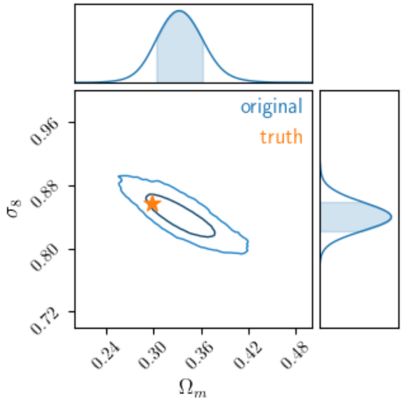


Jia, 2401.02413

NQE Calibration

- Why NQE? NQE can be easily calibrated to be unbiased
- Does your $\alpha\%$ credible region really contain $\alpha\%$ of the truth?
- Empirical Coverage: the probability of the truth to fall within the $\alpha\%$ credible region
- Above diagonal => over-conservative, below diagonal => biased
- The Bayesian optimal posterior has "diagonal" coverage, but the opposite is not always true
- The goal: "diagonal" > "above diagonal" > "below diagonal"

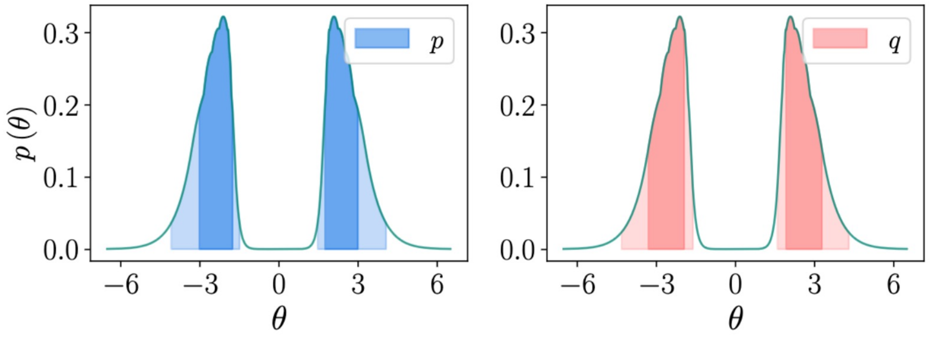




Jia, 2401.02413

How to define credible regions?

- For a 1-dim distribution, the 68% credible region is…
- (Standard definition) the 68% samples with largest posterior, need to sample many θ for each x to get the rank of $p(\theta|x)$
- (Alternative definition) between 16% and 84% quantiles of the distribution, directly from CDF
- Multimodal distribution: local CDF within the peak



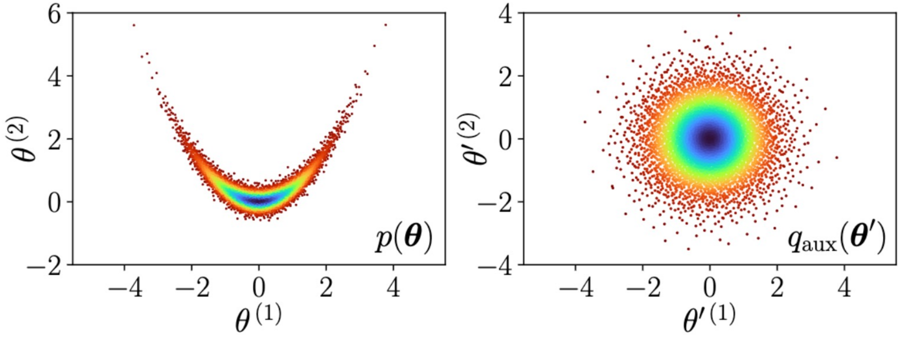
Jia, 2401.02413

How to define credible regions?

- Multimodal distribution: local CDF within the peak

- Multidimensional distribution: map 1-dim conditional quantiles to Gaussian, then use the rank of Gaussian PDF (calculated analytically)

- Advantages: similar results, orders of magnitude faster to evaluate, exclusive to NQE

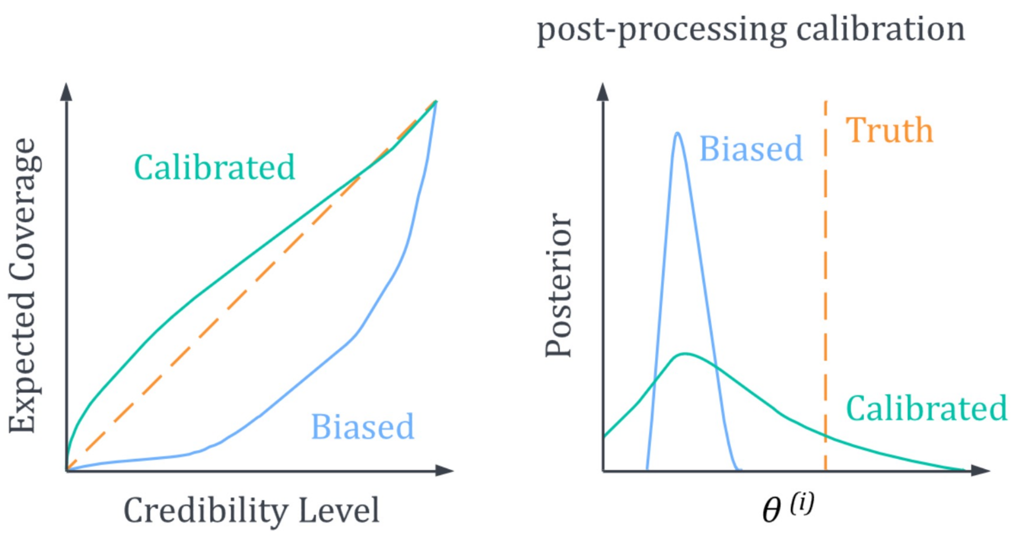


	coverage	simulations	network calls
NQE	$\boldsymbol{q}$	$\boldsymbol{N_o}$	$\boldsymbol{\mathcal{O}(N_o)}$
NQE	p	N_o	$\mathcal{O}(N_i N_o N_r)$
NLE	p	N_o	$\mathcal{O}(N_i N_o N_r N_m)$
NPE	p	N_o	$\mathcal{O}(N_i N_o N_r N_m)$
NRE	p	N_o	$\mathcal{O}(N_i N_o N_r N_m)$

Jia, 2401.02413

NQE Calibration

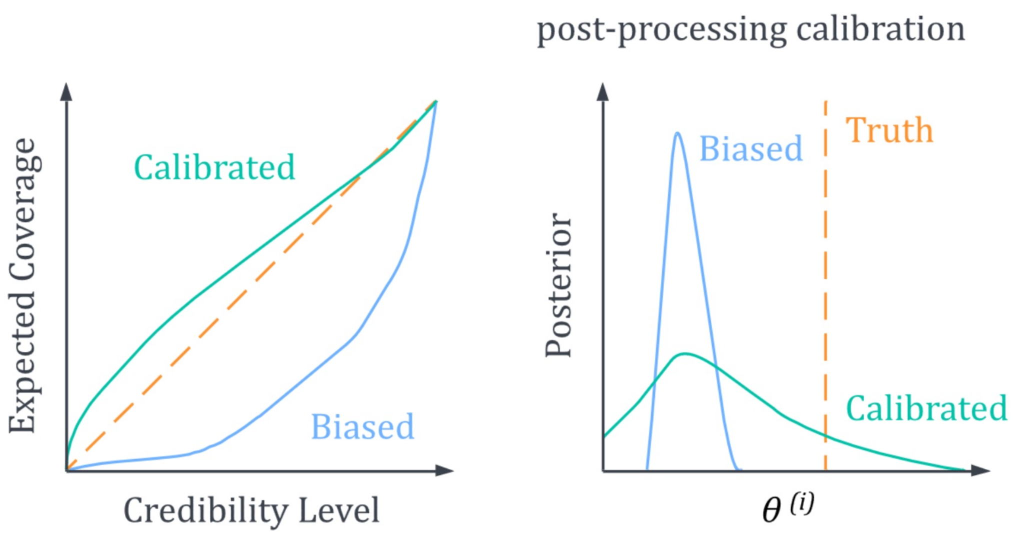
- Biased posterior: posterior is too narrow to cover the truth
- Simple fix: make the posterior broader by post-processing



Jia, 2401.02413

NQE Calibration

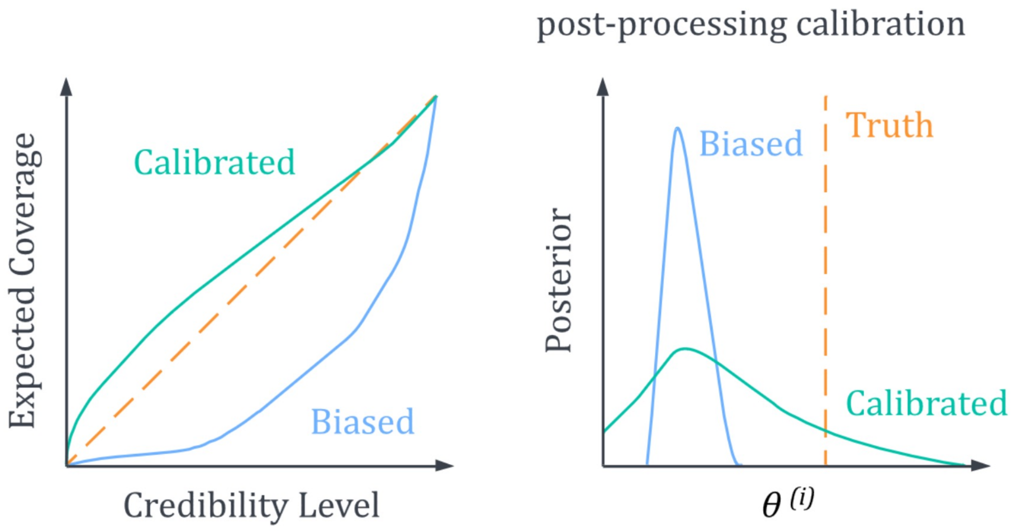
- Only 1 parameter to learn: fix the 1-dim conditional medians, expand all other quantiles by a common “broadening factor”
- Can be done as long as you can accurately calculate the coverage



Jia, 2401.02413

NQE Calibration

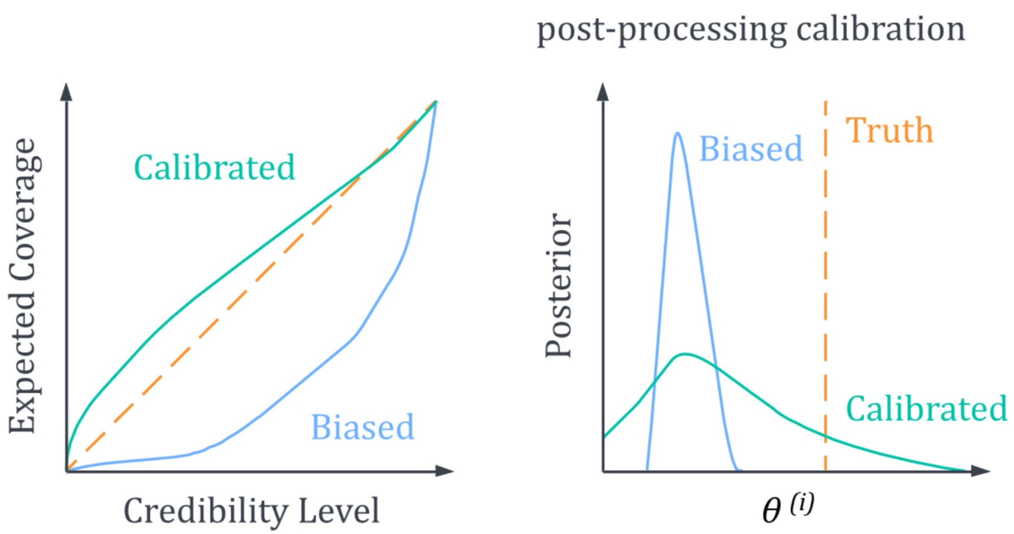
- Errorbar of coverage can be estimated with the Binomial distribution
- <1.6% with 1000 simulations, **regardless of dim x and dim θ**
- In other words, you can always make your estimator unbiased, with 1000 simulations



Jia, 2401.02413

NQE Calibration

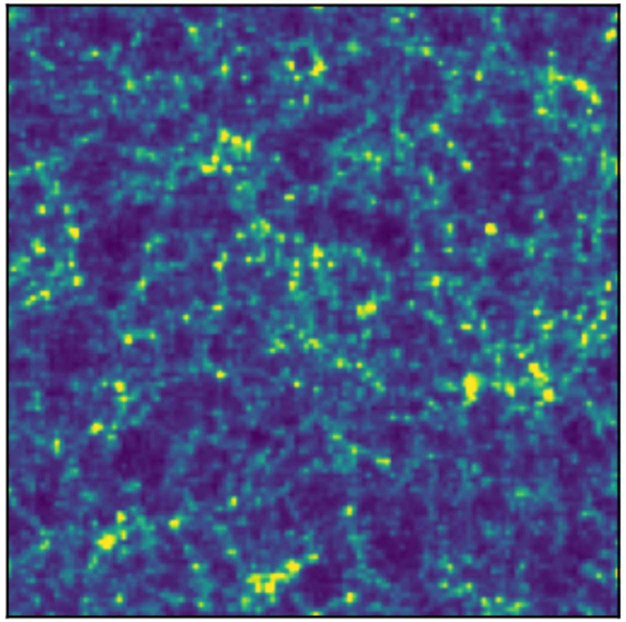
- The global "broadening factor" guarantees unbiasedness, but can be suboptimal
- We will see a better way to do the calibration in a few minutes (with a WL example)



Jia, 2401.02413

WL Example

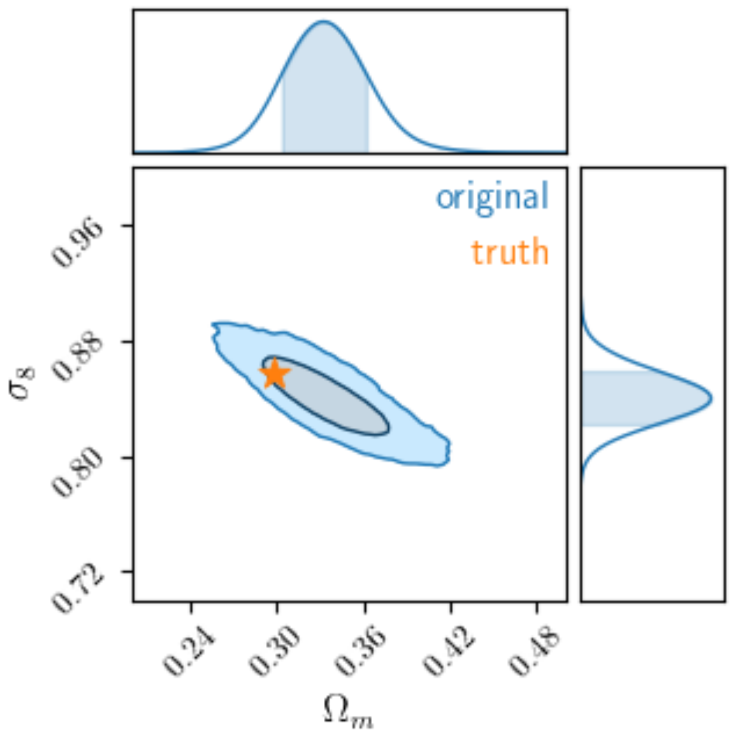
- Infer $(\Omega_m,\ \sigma_8)$ from projected 2-dim density fields
- PM as forward simulator
- Modified ResNet as embedding network
- Field level SBI with NQE
- Can also be applied to summary statistics

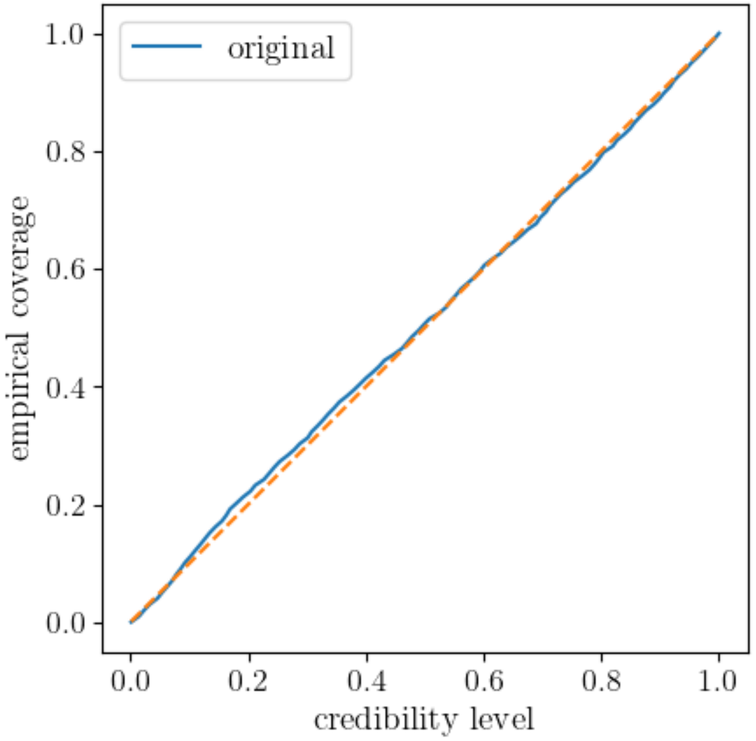


Jia, in prep

WL Example

- Trained on PM, applied to PM ➔ seems to work well!





Jia, in prep

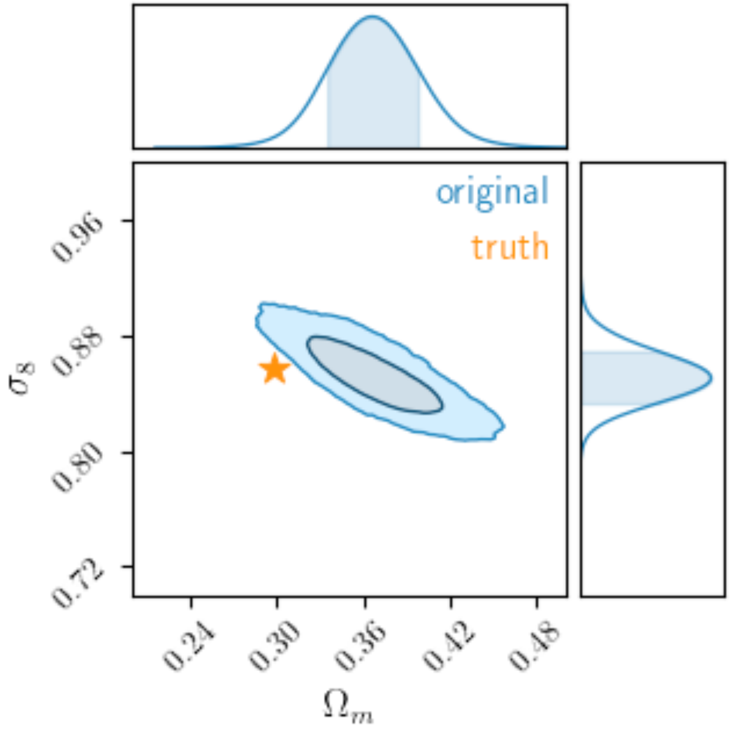
WL Example

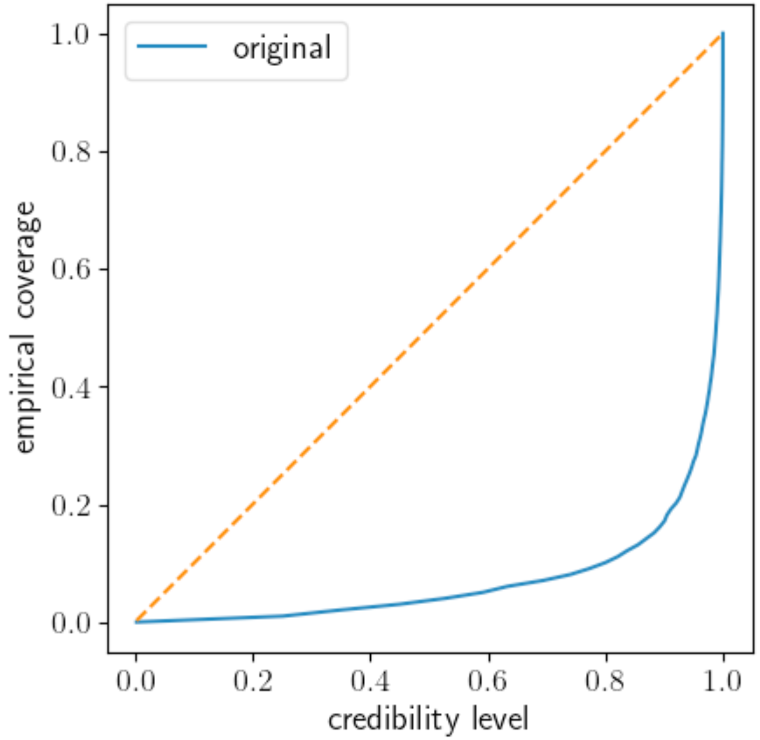
- Trained on PM, applied to “hydro”
- NB: as a proof-of-concept example, I’m not doing real hydro here
- It’s actually PM with scale-independent bias b=1.02

Jia, in prep

WL Example

- Trained on PM, applied to “hydro” ➔ posterior is biased!

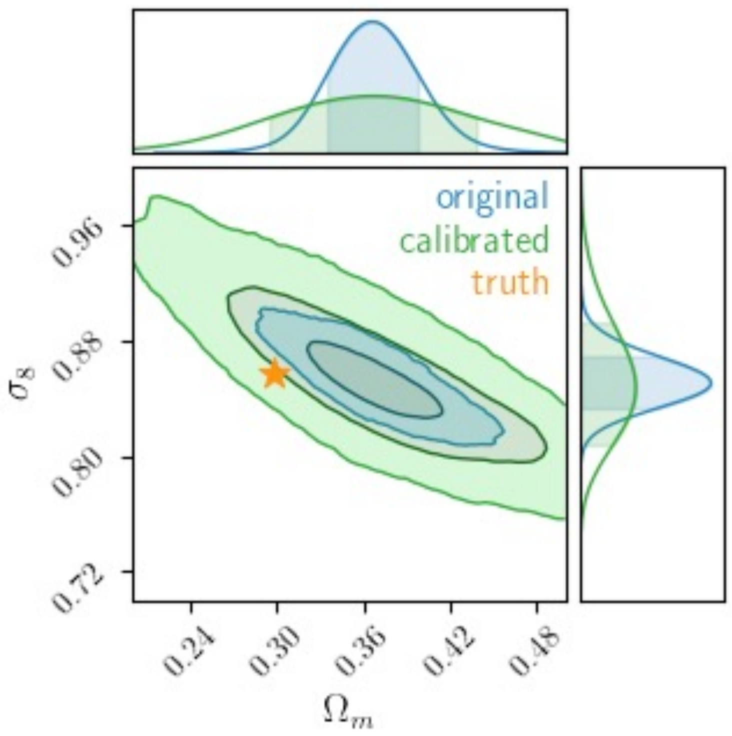


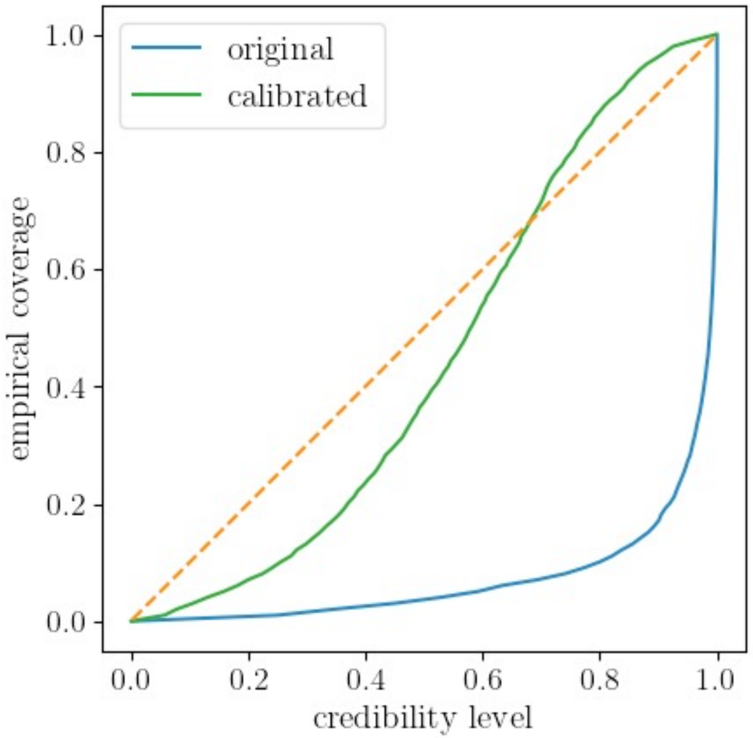


Jia, in prep

WL Example

- Trained on PM, applied to “hydro”, calibrated at 68% and 95%

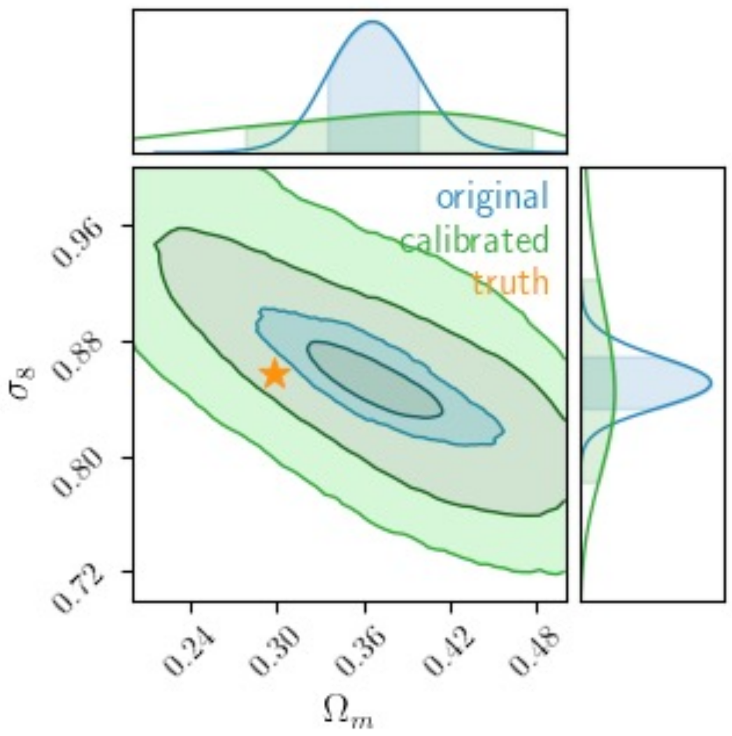


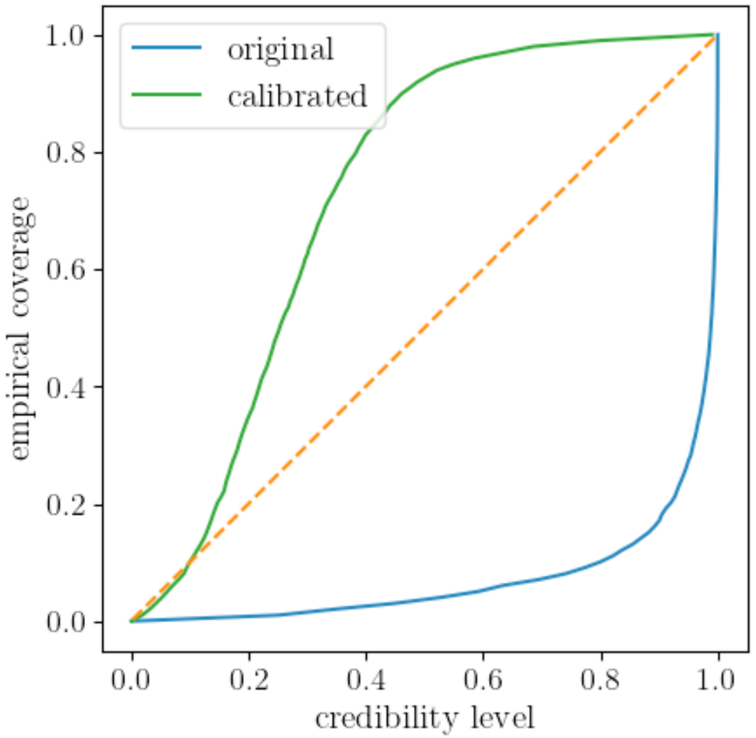


Jia, in prep

WL Example

- Trained on PM, applied to “hydro”, calibrated at 10%, 50% and 90%

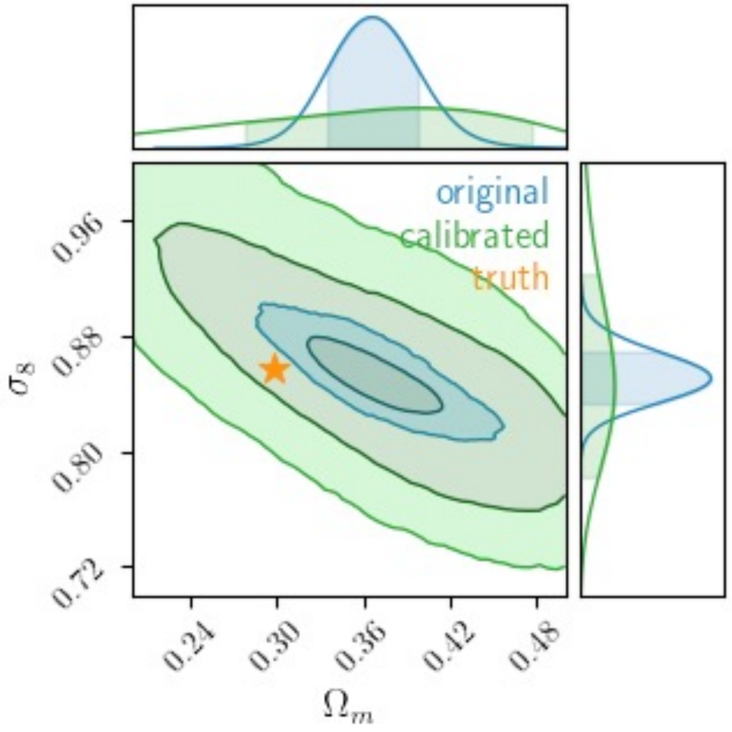




Jia, in prep

Better Way to Calibrate

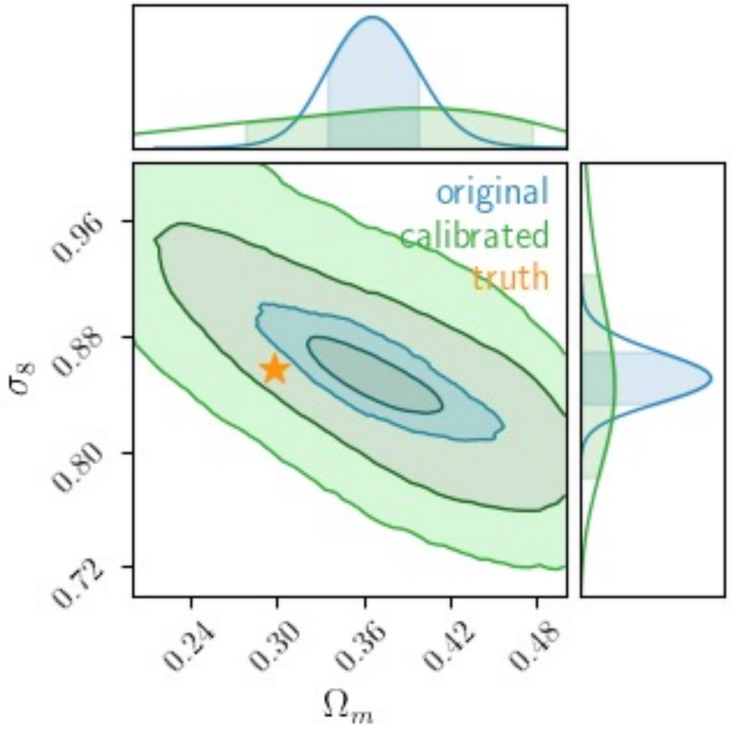
- The isotropic broadening removes the bias, but also makes the posterior taaaaaa broad
- If we know the truth tends to be at one direction, we do not need to broaden the posterior in the other direction
- There is a cleverer way to do the calibration, possible (and only possible) with NQE



Jia, in prep

Better Way to Calibrate

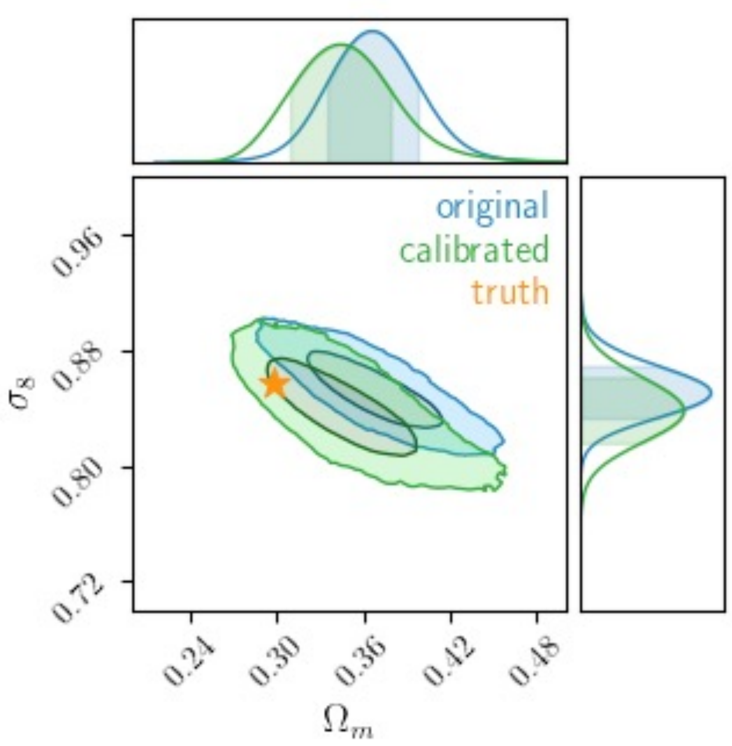
- There is a cleverer way to do the calibration, possible (and only possible) with NQE
- For each $\theta^{(i)}$ dimension, and for each quantile τ
- We compute the residual between the true $\theta^{(i)}$ and the predicted τ-th quantile
- The τ-th quantile of this residual (over all mocks) should be 0
- If not, we can correct the posterior by shifting the predicted quantile (same shift for all mocks)

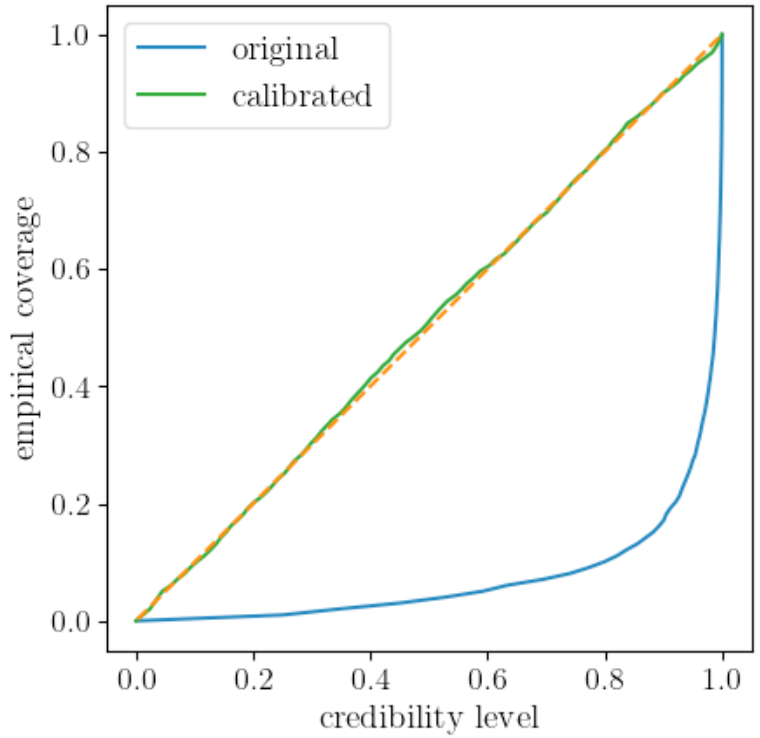


Jia, in prep

WL Example

- Trained on PM, applied to “hydro”, calibrated at all levels

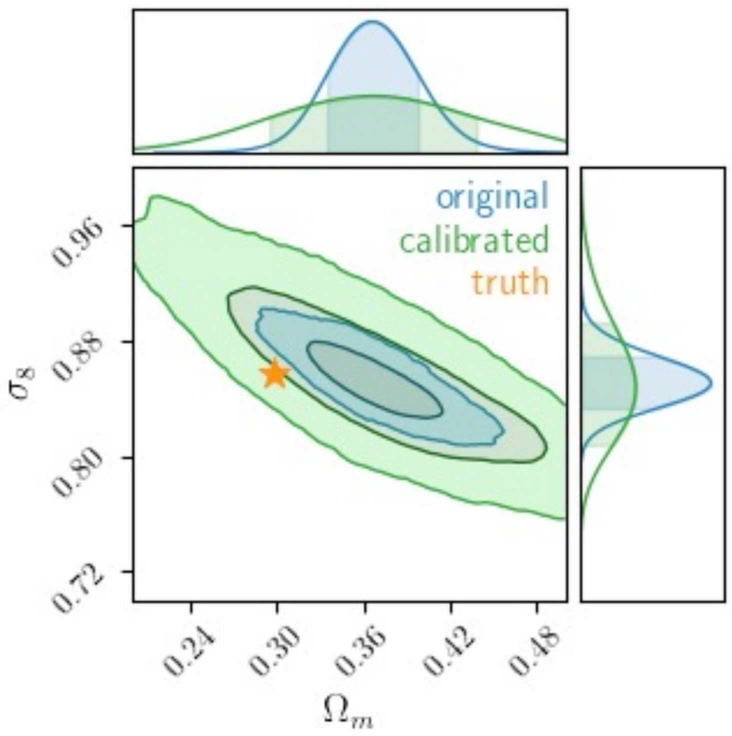


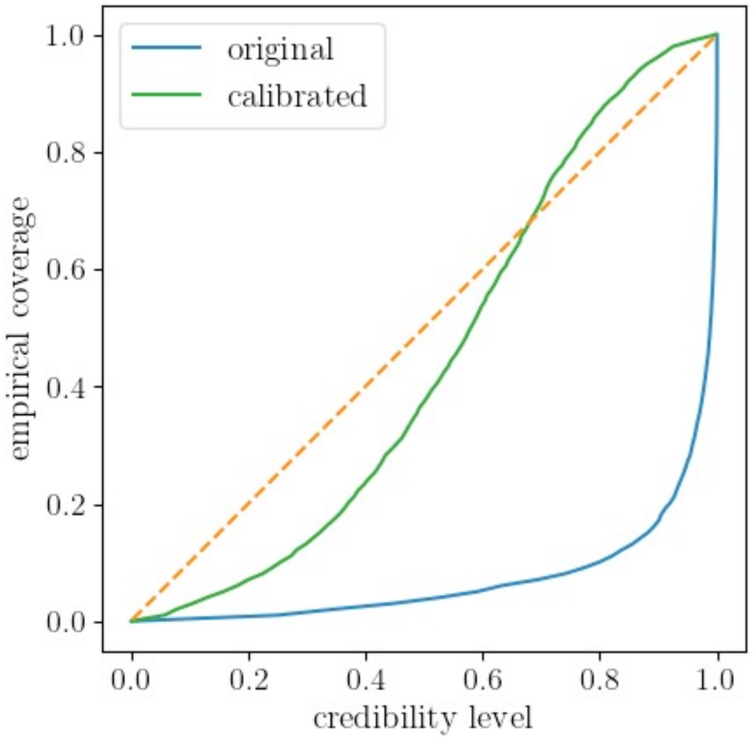


Jia, in prep

WL Example

- Trained on PM, applied to “hydro”, calibrated at 68% and 95%

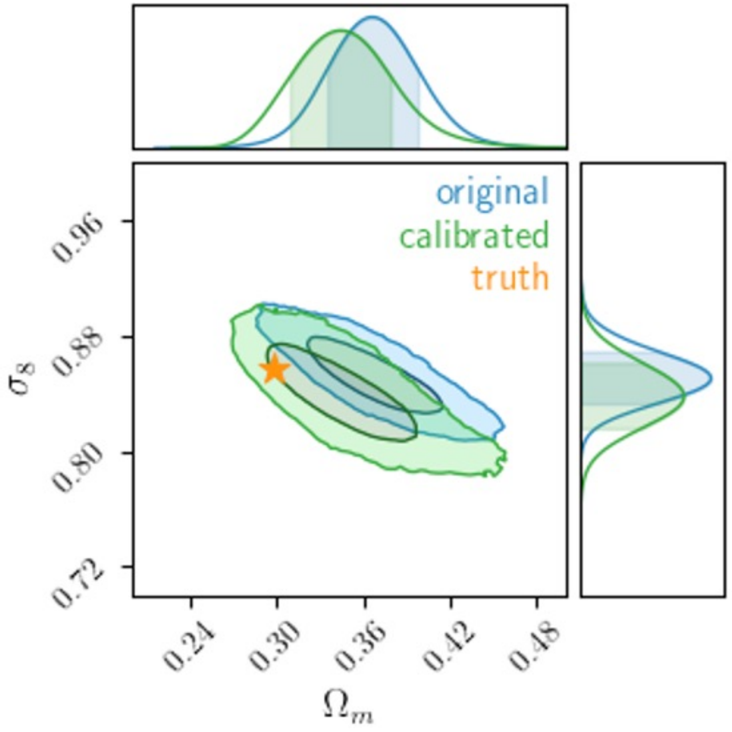




Jia, in prep

Better Way to Calibrate

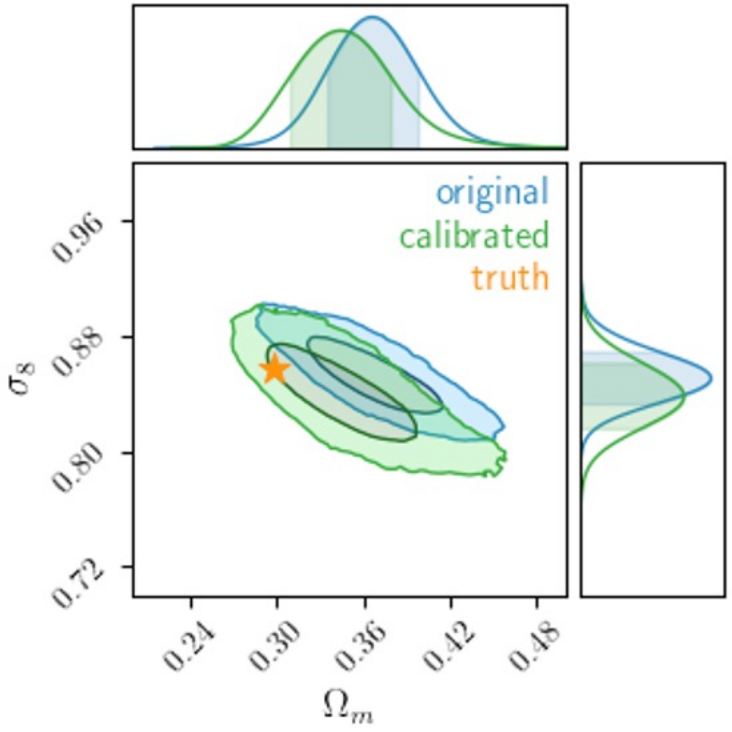
- Effectively, I'm averaging the posterior bias over all mocks
- This is optimal, if and only if the inferred posterior is always biased (relative to Bayesian optimal posterior) in the same way
- Otherwise, some information is lost
- However, you only need 500-1000 *correct* simulations (with which you want to calibrate NQE) to do this



Jia, in prep

Better Way to Calibrate

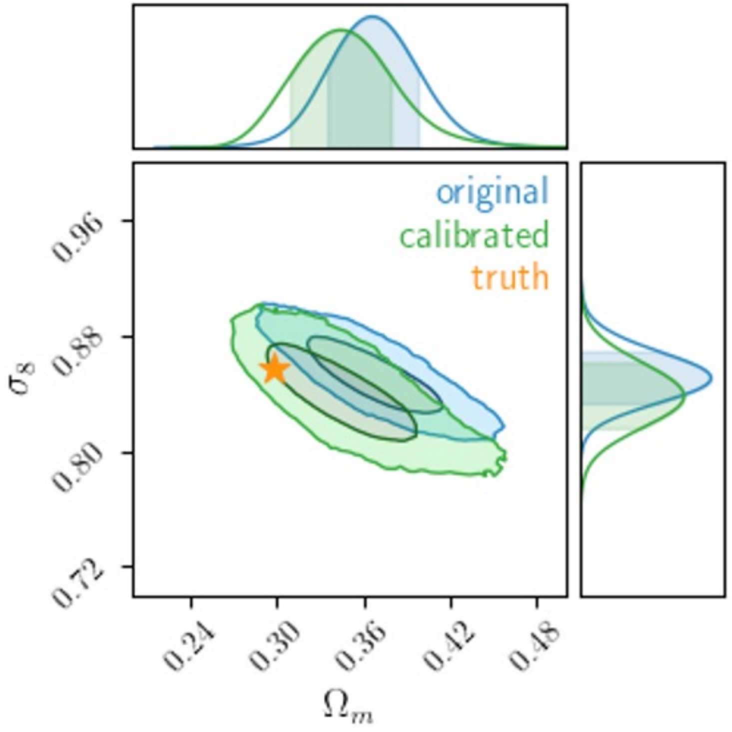
- This is only possible with NQE
- NQE predicts global information (quantiles) of the posterior: you know why your posterior is biased
- Existing methods like NPE predicts only local information (the PDF of the posterior)



Jia, in prep

Possible Applications (Emulators)

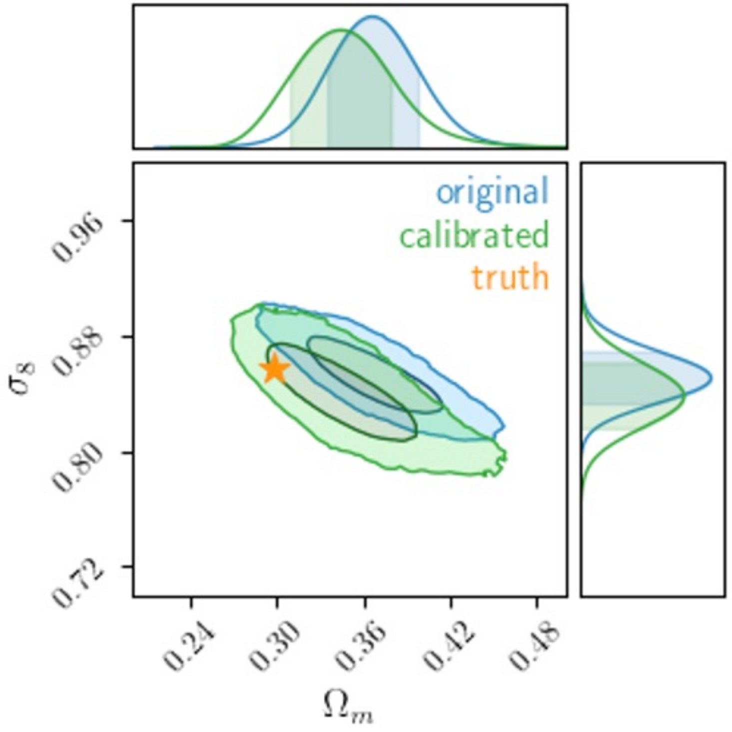
- A more direct way to evaluate how good the emulator is: how much calibration is required to remove the bias?
- To do inference, emulators do not need to be perfect
- Better emulators lead to more optimal posteriors
- Bias can always be removed with calibration



Jia, in prep

Possible Applications (Baryon Uncertainties)

- What do you do if you find your SBI results different on different hydro simulations?
- Before: manually pick some subset of observables that are less sensitive to the hydro models
- Now: accept it, train your SBI on some baseline model, then calibrate it against all the other hydros
- Explicitly marginalizing over baryon uncertainties in the posterior space



Jia, in prep

Thanks & Questions?

- Neural Quantile Estimation (NQE), a new SBI method
- Guaranteed to be unbiased if you have 500-1000 runs from the *correct* simulator
- Code is public on GitHub (h3jia/nqe), although no documentation yet
- ML methodology paper: 2401.02413
- Let me know if you want to try it on your examples!

Jia, in prep