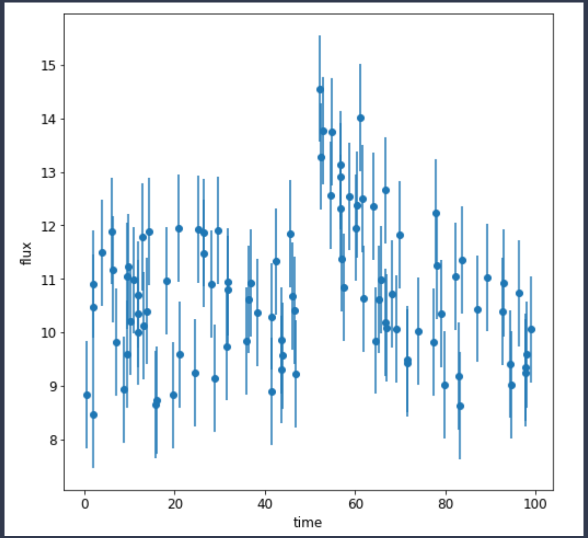
Simulation-Based Inference
Stakes and applications

Sacha Guerrini - Cosmology and Statistics Day
01/02/2024

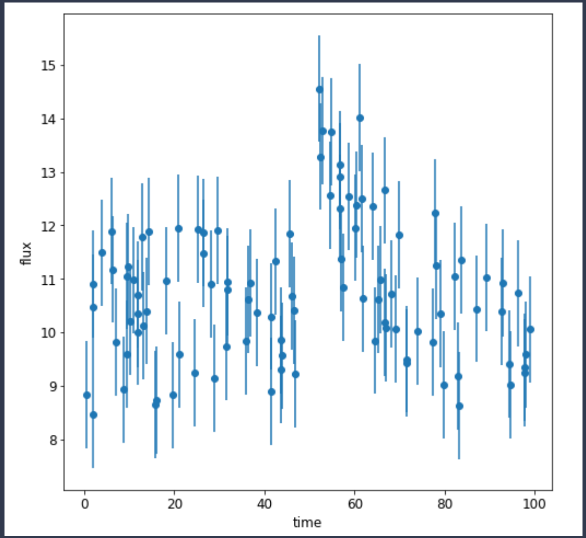
Table of content

What is inference?

What is inference?

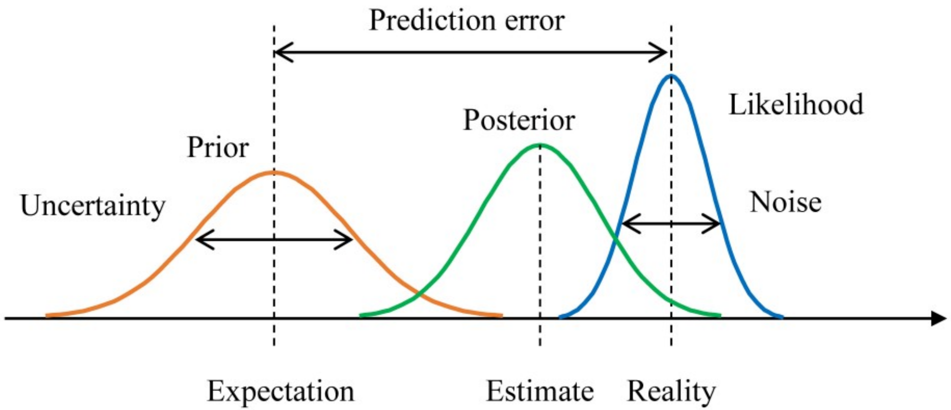


What is inference?



Parameters: θ ⟶ **Model** — Latent variables: z ⟶ Observation: d

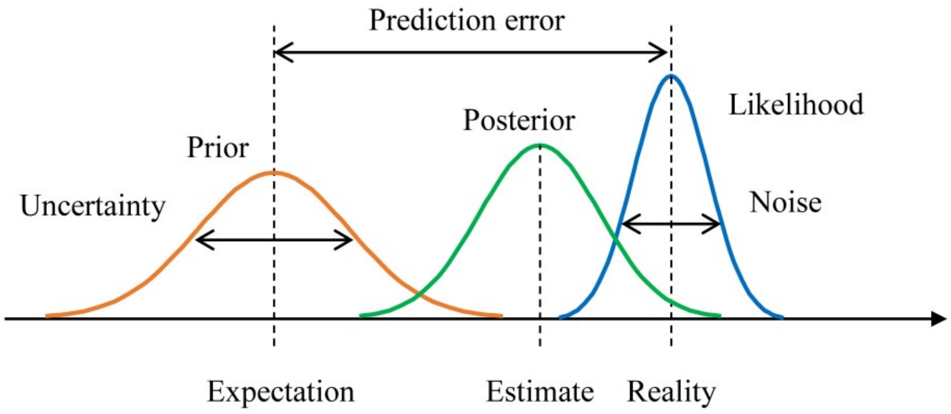
Likelihood-Based Inference



Bayes theorem:

$$\underbrace{p(\theta|d)}_{\text{Posterior}} \propto \underbrace{p(d|\theta)}_{\text{Likelihood}} \underbrace{p(\theta)}_{\text{Prior}}$$

Likelihood-Based Inference



Bayes theorem:

$$\underbrace{p(\theta|d)}_{\text{Posterior}} \propto \underbrace{p(d|\theta)}_{\text{Likelihood}} \underbrace{p(\theta)}_{\text{Prior}}$$

Requires a choice for the Likelihood function.

Likelihood-Based Inference

Usual assumption: the Gaussian likelihood

$$L(\mathbf{d}|\boldsymbol{\theta}) = (2\pi)^{-m/2}|\mathbf{C}(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\mu}(\boldsymbol{\theta})^T\mathbf{C}^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}(\boldsymbol{\theta})\right]$$

$$\boldsymbol{\mu} = \mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$$

Data vector   Model prediction

$\mathbf{C}(\boldsymbol{\theta})$: Covariance matrix

Likelihood-Based Inference

Usual assumption: the Gaussian likelihood

$$L(\mathbf{d}|\boldsymbol{\theta}) = (2\pi)^{-m/2}|\mathbf{C}(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\mu}(\boldsymbol{\theta})^T\mathbf{C}^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}(\boldsymbol{\theta})\right]$$

$$\boldsymbol{\mu} = \mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$$

Data vector   Model prediction

$\mathbf{C}(\boldsymbol{\theta})$: Covariance matrix

Can be challenging to estimate

Likelihood-Based Inference

Usual assumption: the Gaussian likelihood

$$L(\mathbf{d}|\boldsymbol{\theta}) = (2\pi)^{-m/2}|\mathbf{C}(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\mu}(\boldsymbol{\theta})^T\mathbf{C}^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}(\boldsymbol{\theta})\right]$$

$$\boldsymbol{\mu} = \mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$$

Data vector   Model prediction

$\mathbf{C}(\boldsymbol{\theta})$: Covariance matrix

Can be challenging to estimate

Covariance estimation methods:

- Jackknife/bootstrap resampling
- Sample variance from simulations
- Analytical model

Likelihood-Based Inference

Usual assumption: the Gaussian likelihood

$$L(\mathbf{d}|\boldsymbol{\theta}) = (2\pi)^{-m/2}|\mathbf{C}(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\mu}(\boldsymbol{\theta})^T\mathbf{C}^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}(\boldsymbol{\theta})\right]$$

$$\boldsymbol{\mu} = \mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$$

Data vector — Model prediction

$\mathbf{C}(\boldsymbol{\theta})$: Covariance matrix

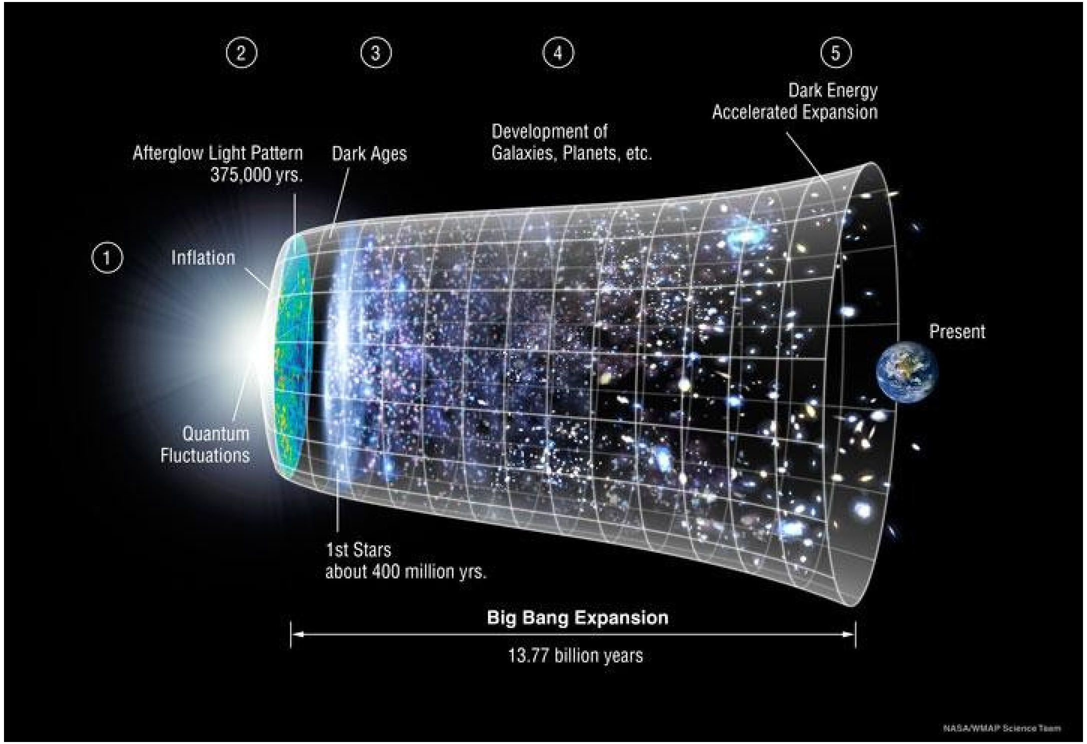
Can be challenging to estimate

Covariance estimation methods:

- Jackknife/bootstrap resampling
- Sample variance from simulations
- Analytical model

Can be cumbersome and/or computationally expensive.

Application: Cosmological parameters inference

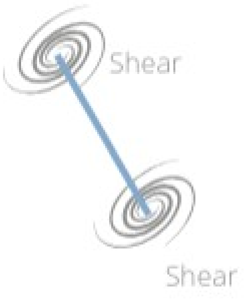


The history of the Universe is well described by a few parameters:

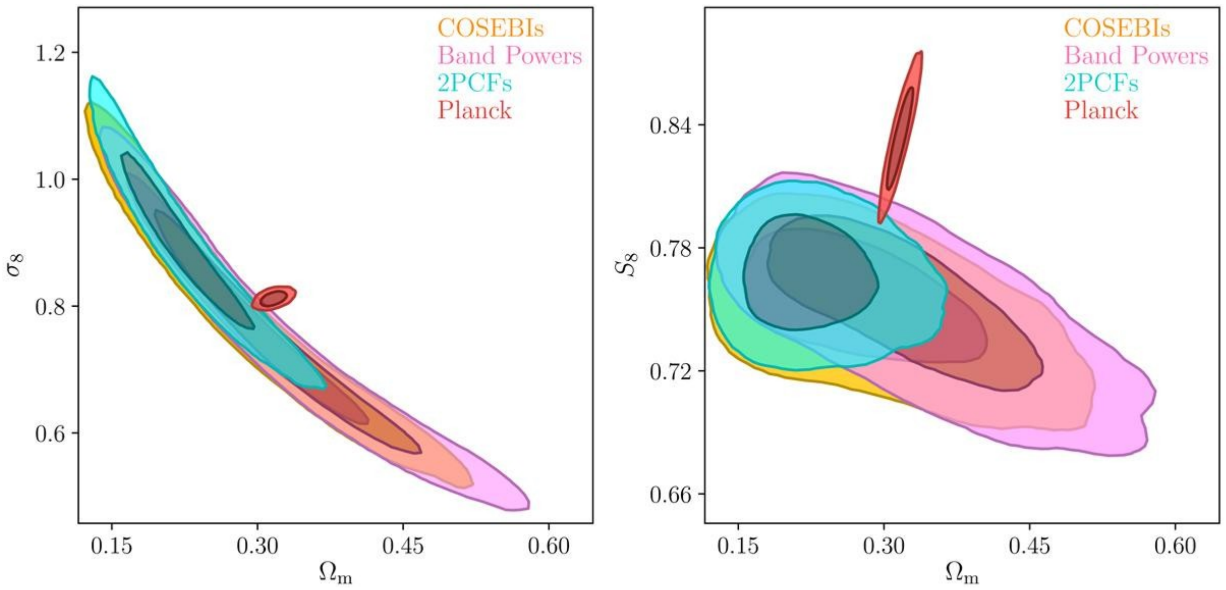
- H_0: Current expansion rate
- Ω_m: Matter density
- Ω_b: Baryon density
- Ω_Λ: Dark energy density
- σ_8: Clumpiness
- n_s: Scale index of initial density fluctuations
- w: Evolution of dark energy

Application: Cosmological parameters inference

Summary statistics:
2-point correlation function



KiDS-1000 cosmology result
(Asgari et al. (2021))



What is problematic with LBI?

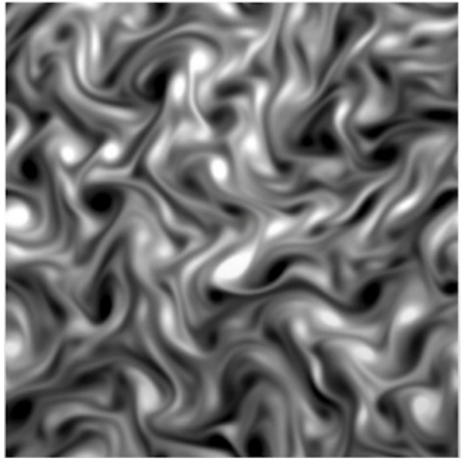
Caveats of Likelihood-Based Inference

- Covariance estimate can be cumbersome.
- Latent variables are intractable and can add systematic effects that needs to be taken into account.
- The Likelihood is not necessarily Gaussian or analytical. (Gaussian assumption do not capture all the information, namely the interactions between scales.)

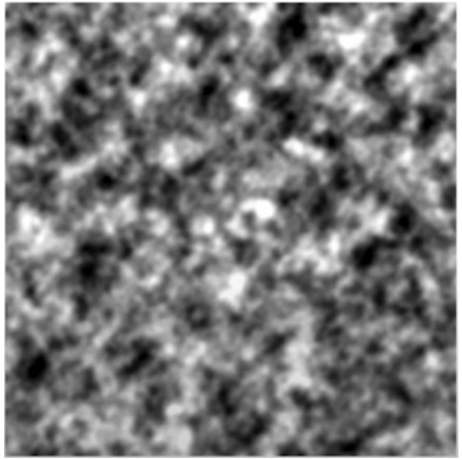
Caveats of Likelihood-Based Inference

- Covariance estimate can be cumbersome.
- Latent variables are intractable and can add systematic effects that needs to be taken into account.
- The Likelihood is not necessarily Gaussian or analytical. (Gaussian assumption do not capture all the information, namely the interactions between scales.)

Turbulent flow



Gaussian



Simulation-Based Inference - Principles

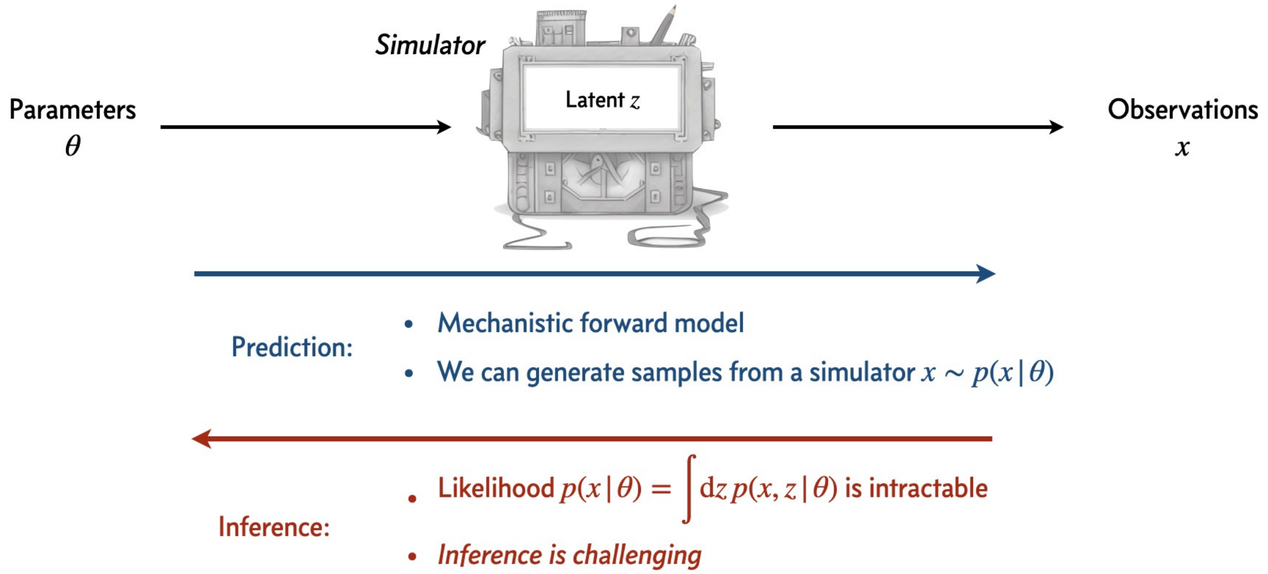
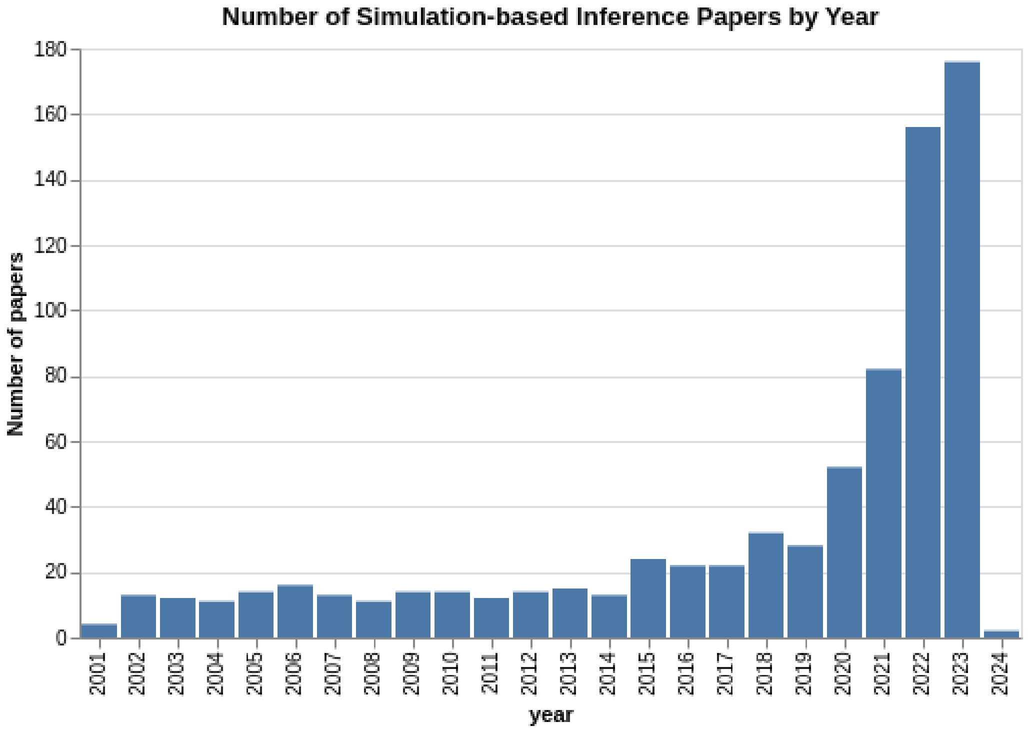


Image Credit: Siddharth Mishra-Sharma

Simulation-Based Inference - Principles



Simulation-Based Inference - Principles

Tools and Resources

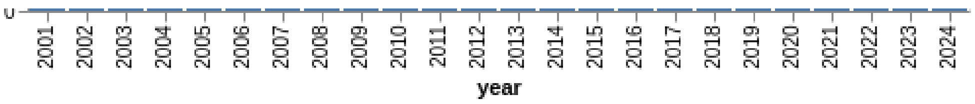
Neuroscience

Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience

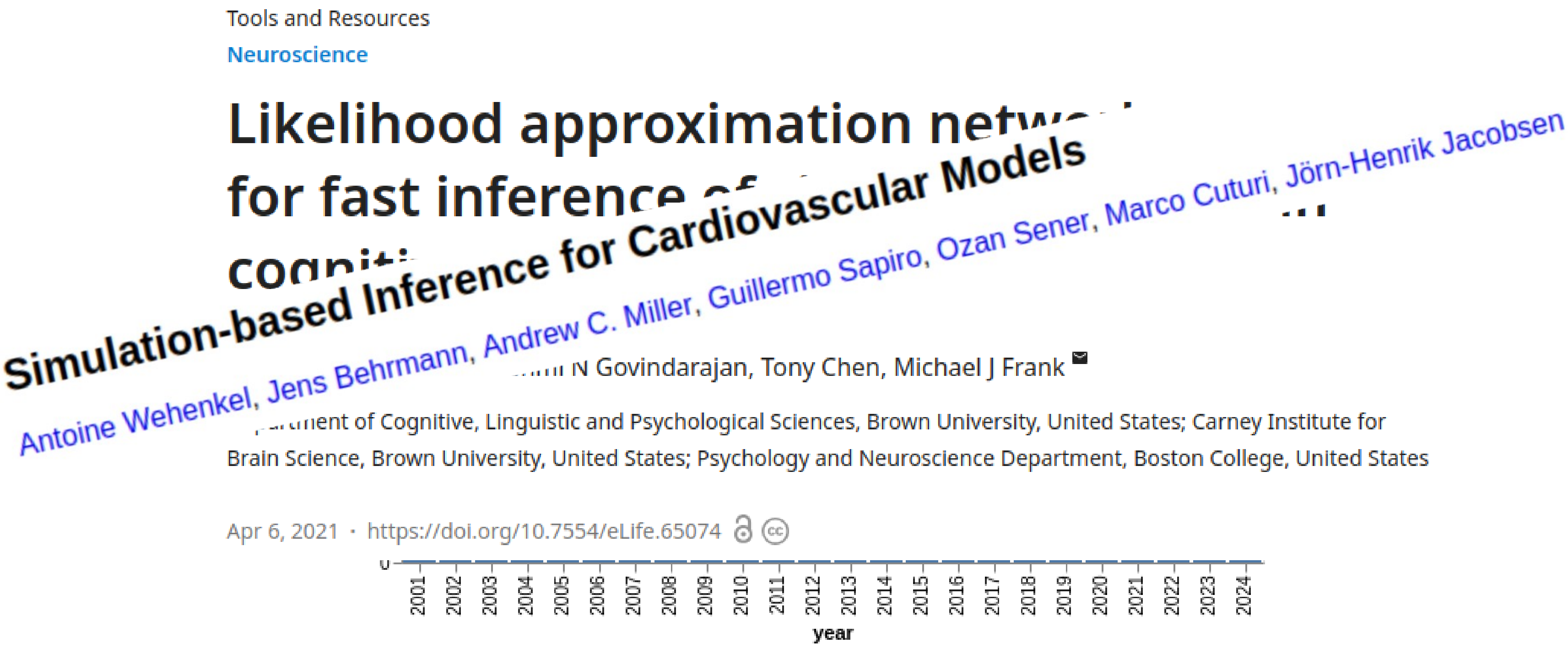
Alexander Fengler, Lakshmi N Govindarajan, Tony Chen, Michael J Frank

Department of Cognitive, Linguistic and Psychological Sciences, Brown University, United States; Carney Institute for Brain Science, Brown University, United States; Psychology and Neuroscience Department, Boston College, United States

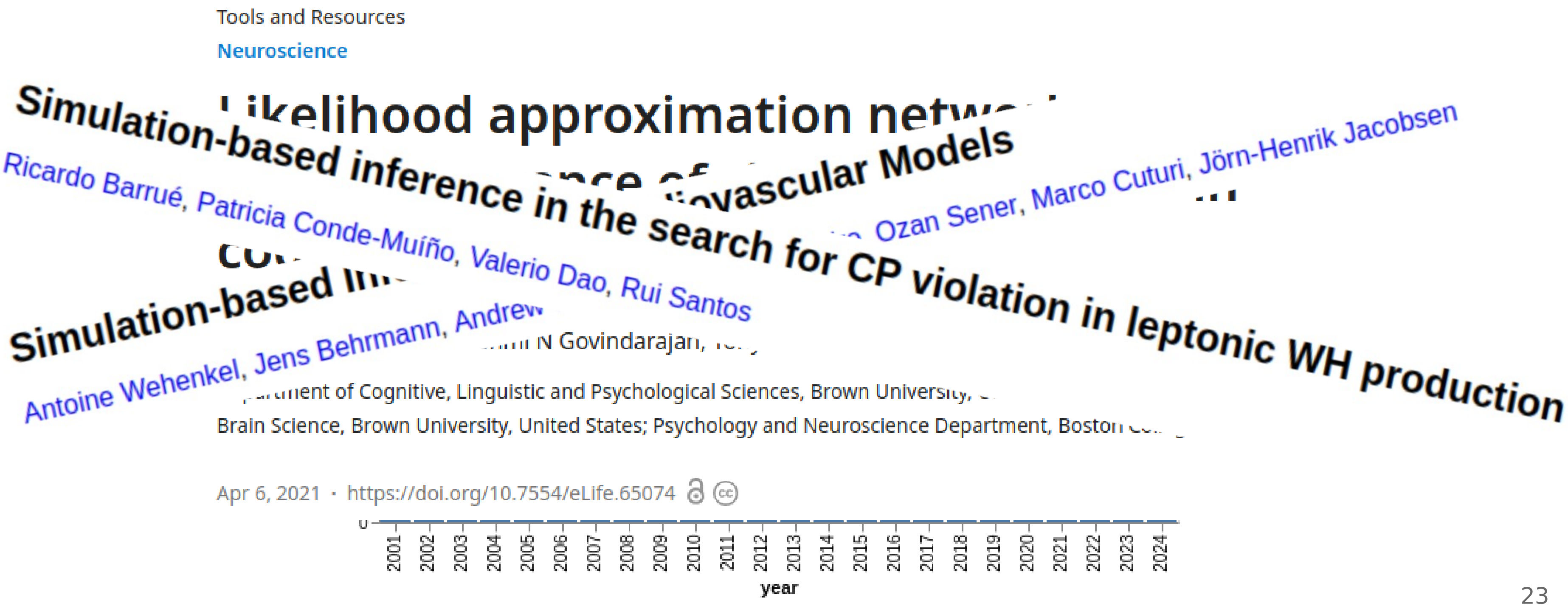
Apr 6, 2021 · https://doi.org/10.7554/eLife.65074



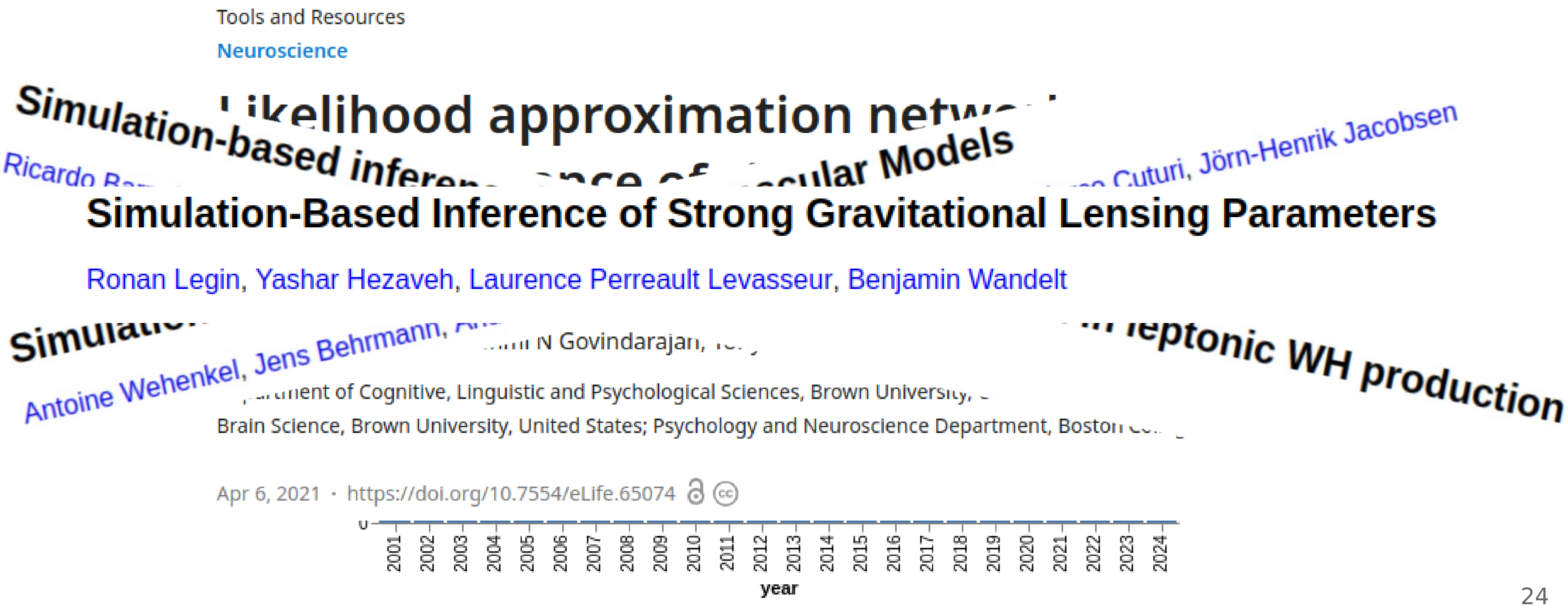
Simulation-Based Inference - Principles



Simulation-Based Inference - Principles



Simulation-Based Inference - Principles



Simulation-Based Inference - ABC

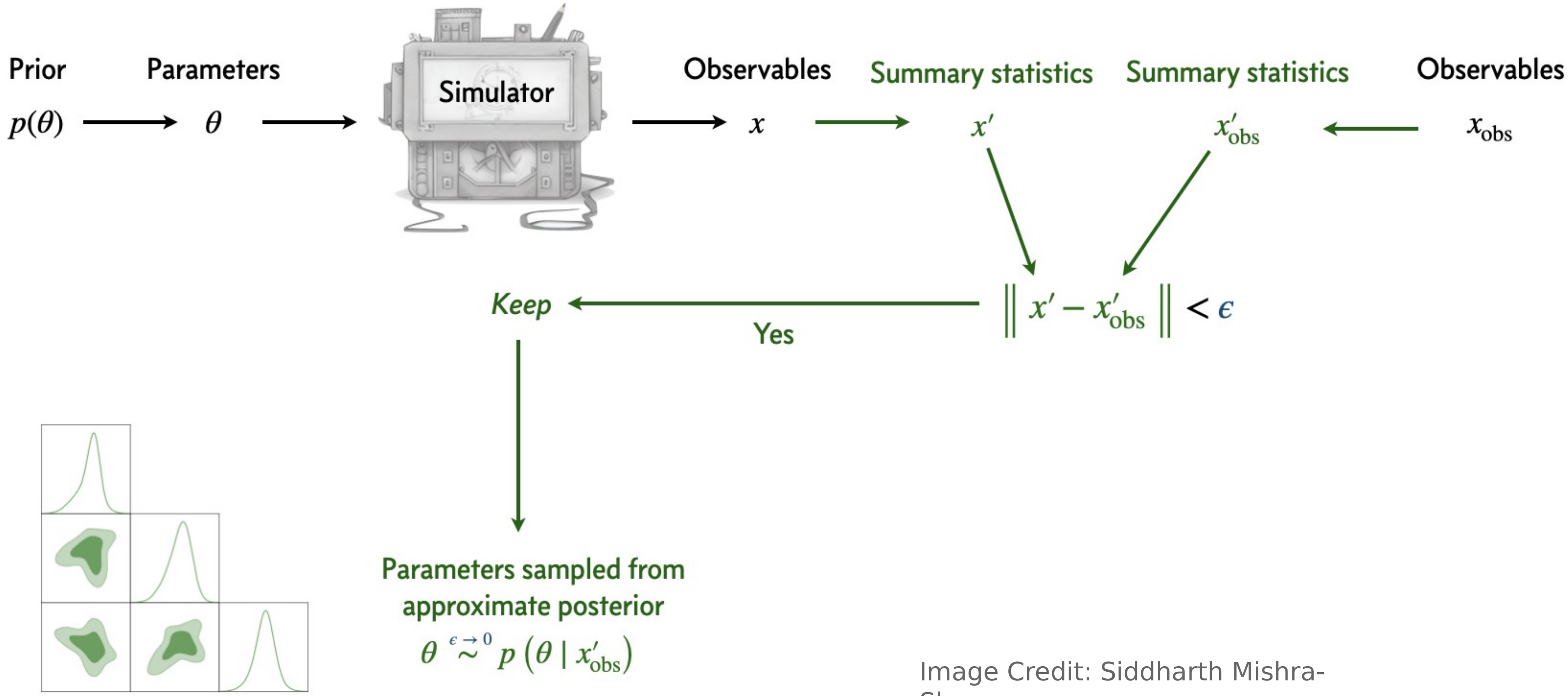
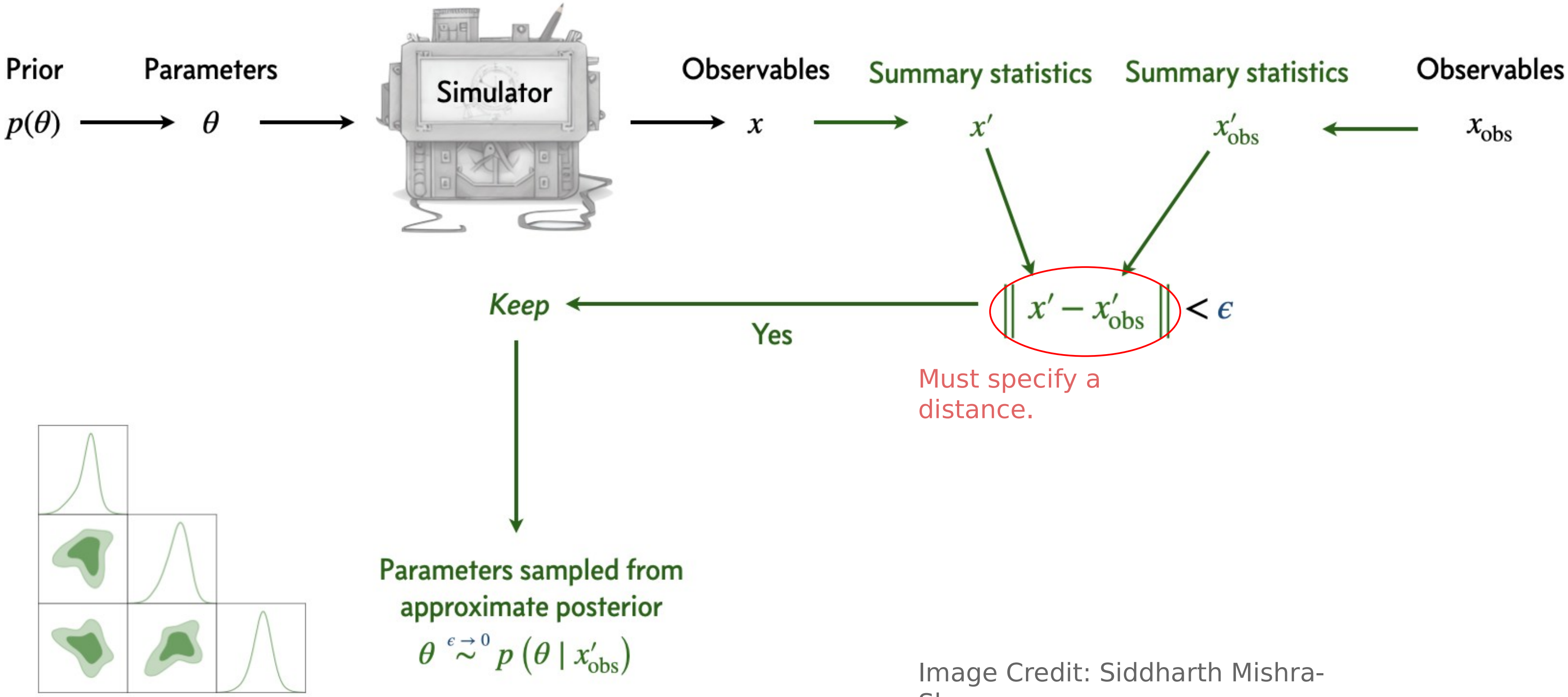


Image Credit: Siddharth Mishra-Sharma

Simulation-Based Inference - ABC



Must specify a distance.

Image Credit: Siddharth Mishra-Sharma

Simulation-Based Inference - ABC

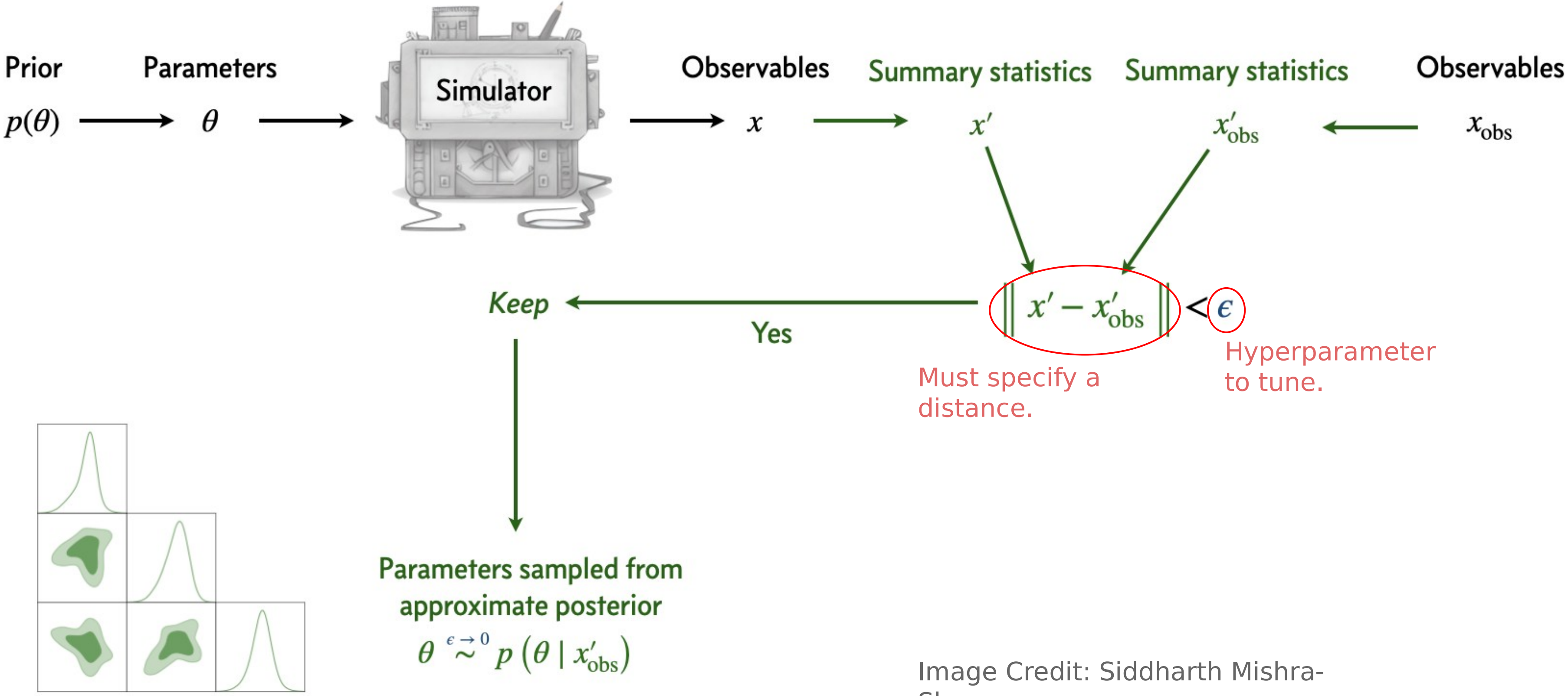
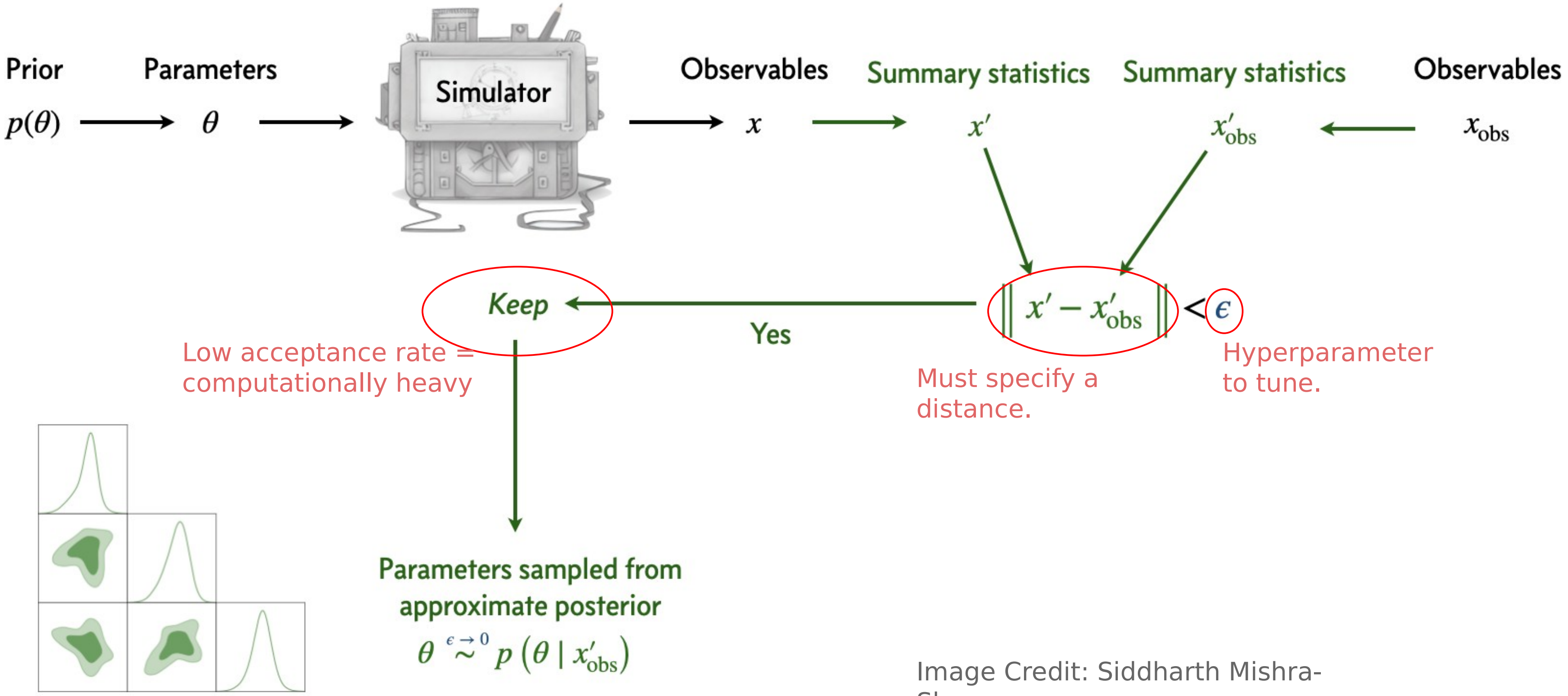


Image Credit: Siddharth Mishra-Sharma

Simulation-Based Inference - ABC



Low acceptance rate = computationally heavy

Must specify a distance.

Hyperparameter to tune.

Image Credit: Siddharth Mishra-Sharma

Simulation-Based Inference - NDE

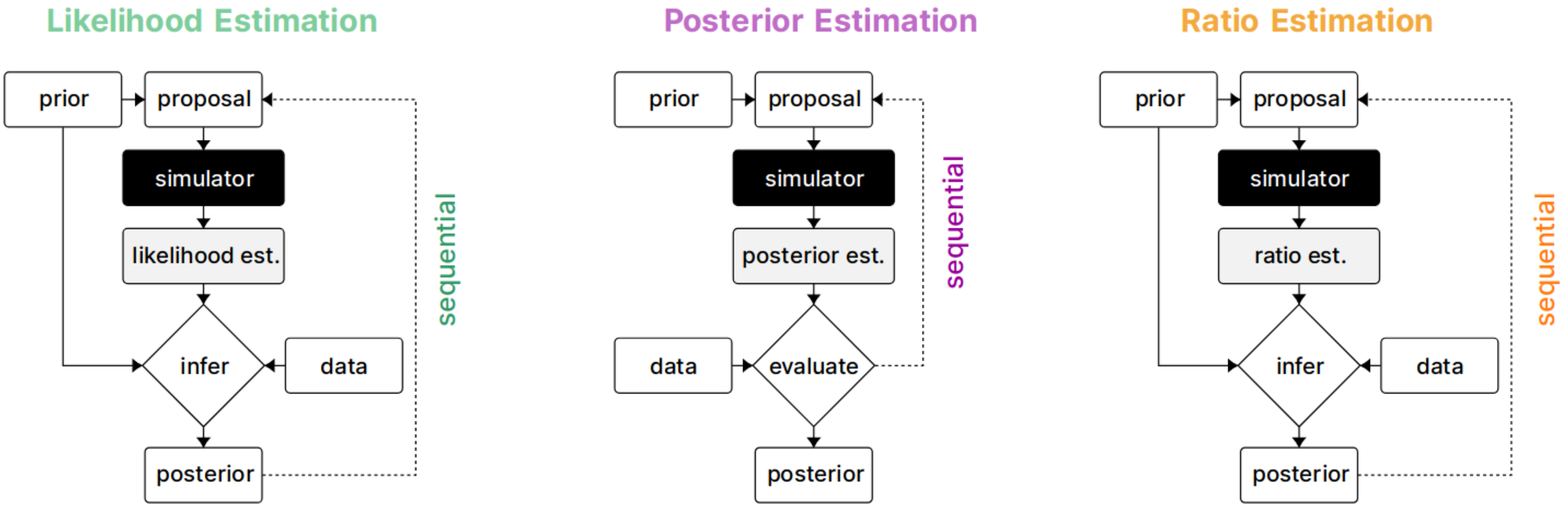


Image Credit: Lueckmann et al.

Simulation-Based Inference - NDE

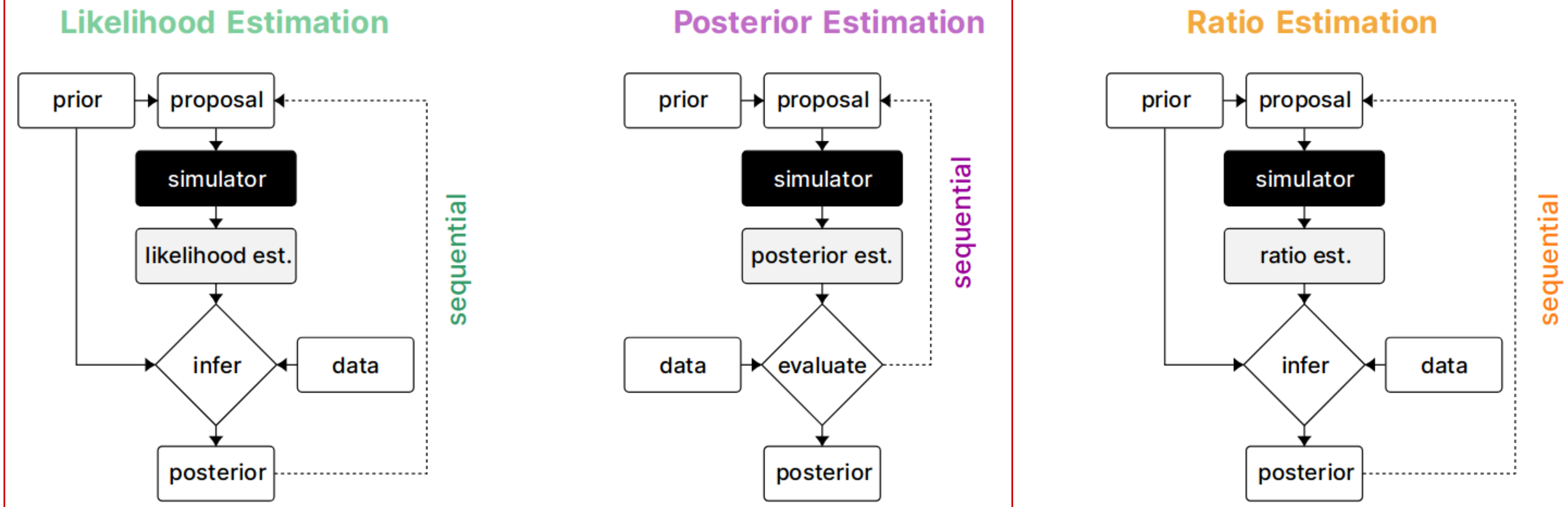
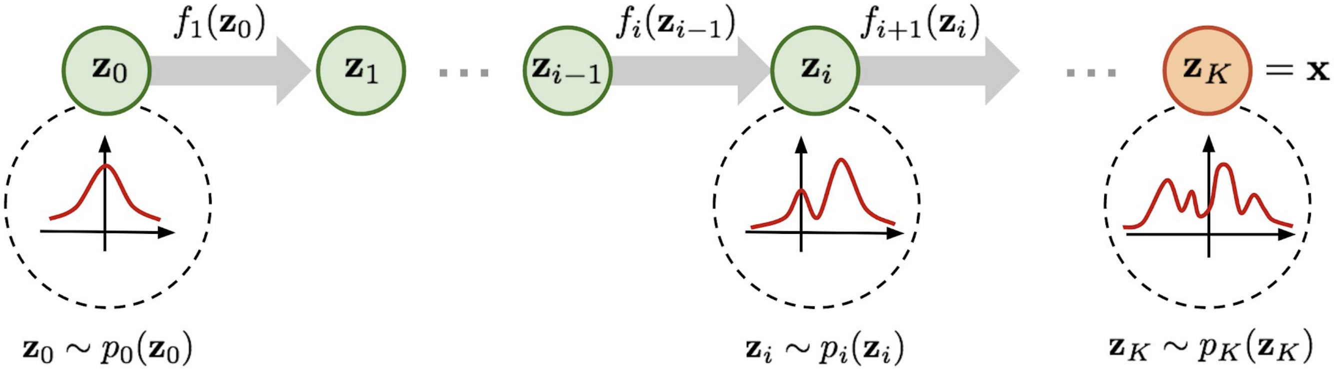


Image Credit: Lueckmann et al.

Normalizing Flows



$$p_i(z_i) = p_{i-1}(f^{-1}(z_i)) \left| \det \frac{df^{-1}}{dz_i} \right|$$

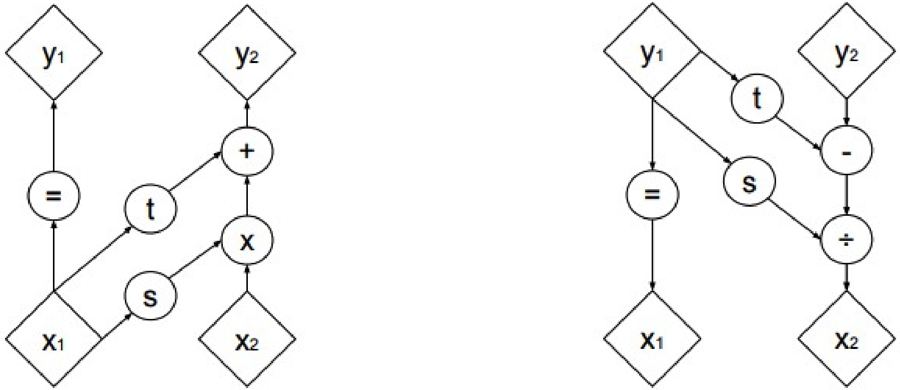
An example: Real NVP

Training data: (θ, d_{sim})

Loss function: $-\ln U = -\sum_i \ln p(d_i|\theta_i; w)$

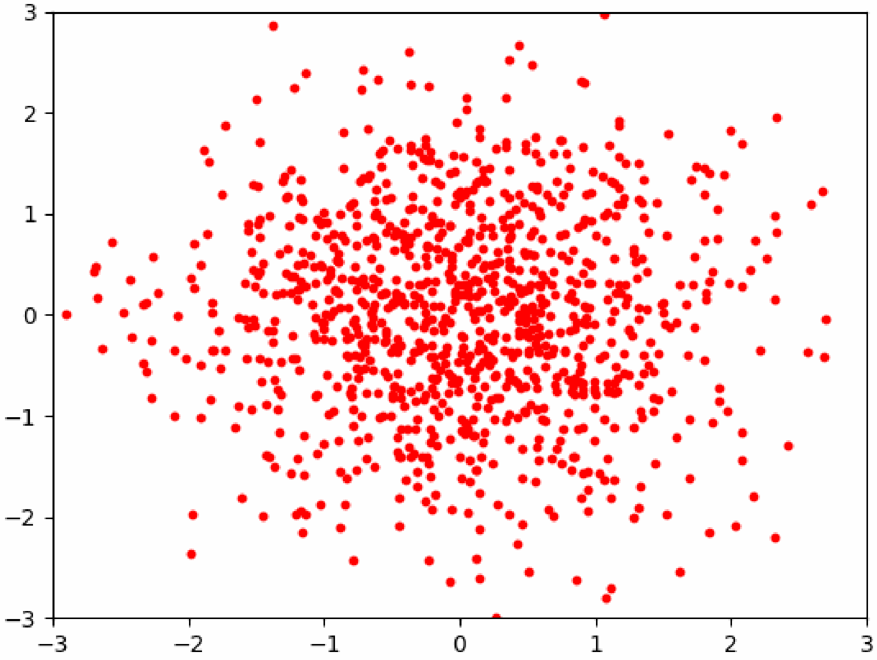
Minimizes KL-divergence w.r.t a target density:

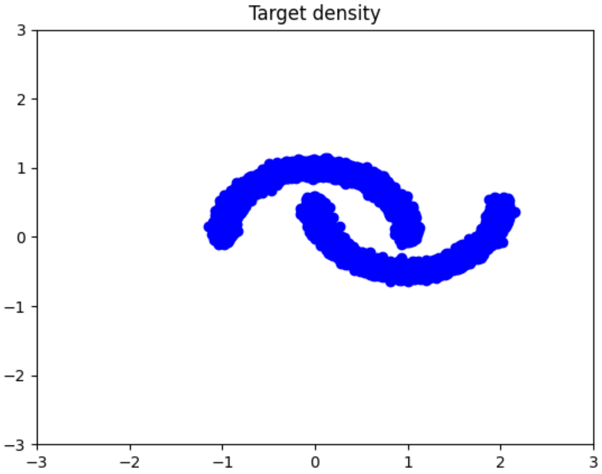
$$D_{\text{KL}}(p^*(d|\theta)||p(d|\theta;w)) = \int p^*(d|\theta) \log\left(\frac{p^*(d|\theta;w)}{p(d|\theta;w)}\right)$$



(a) Forward propagation (b) Inverse propagation

An example: Real NVP





Simulation-Based Inference: Summary

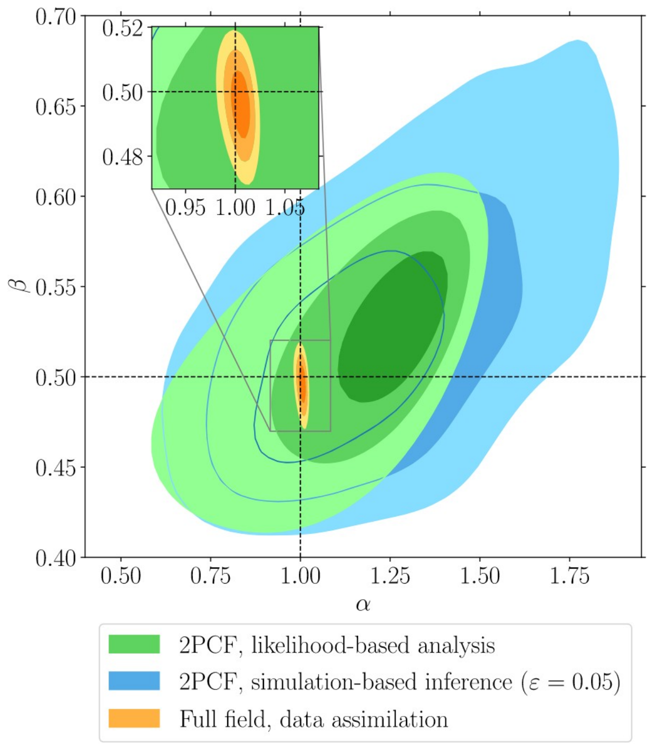
Advantages:

- Does not require any assumption on the form of the likelihood, you learn it.
- Systematics effects can be included in the forward model.
- Straightforward method.
- Can be trained and exploited on GPUs and the model is fully differentiable.

Drawbacks:

- Requires high-quality simulated data. The latter is difficult to assess.
- Uncertainty with neural network is difficult to quantify.
- Training can be long and cumbersome (e.g. compression and pre-training steps)

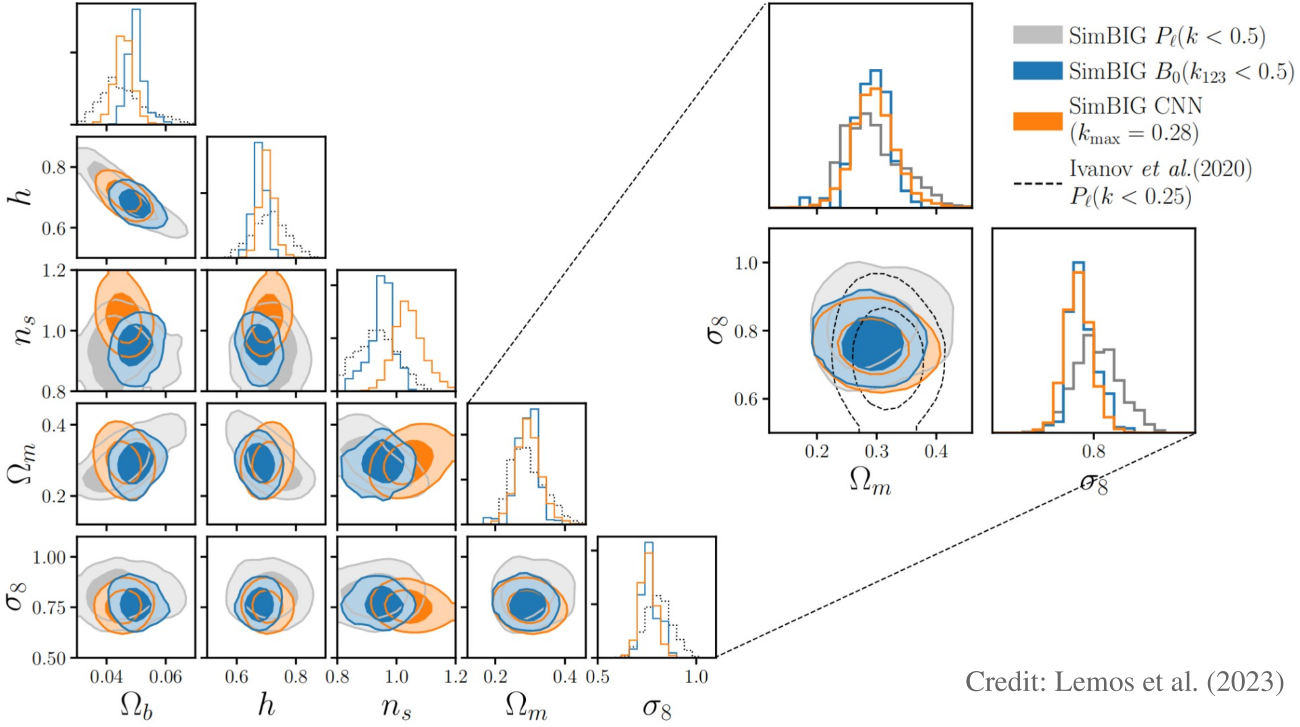
Application: Cosmological parameters inference



Leclercq and Heavens (2021):

- Simplified toy model for cosmology application.
- Likelihood-based analysis does not peak at the accurate mode.

Application: Cosmological parameters inference



Credit: Lemos et al. (2023)

Conclusion and summary

Simulation-based inference allows:

- To sample parameters from the posterior without any assumption on the form of the likelihood.
- To include all systematics effects in the forward model.
- To efficiently perform those computations on GPUs.

But:

- One must assess the realism of the simulations used.
- Uncertainty quantification is difficult.

Ressources:

- https://github.com/smsharma/awesome-neural-sbi
- https://simulation-based-inference.org/