Does Bayes have to be slow?

A glimpse into amortized Bayesian inference

Paul Bürkner

TU Dortmund University, Germany

https://paul-buerkner.github.io/

Our Team


Stefan T. Radev


Marvin Schmitt


Florence Bockting


Daniel Habermann


Ullrich Köthe


Lasse Elsemüller


Lukas Schumacher


Philipp Reiser


Jacob Grytzka

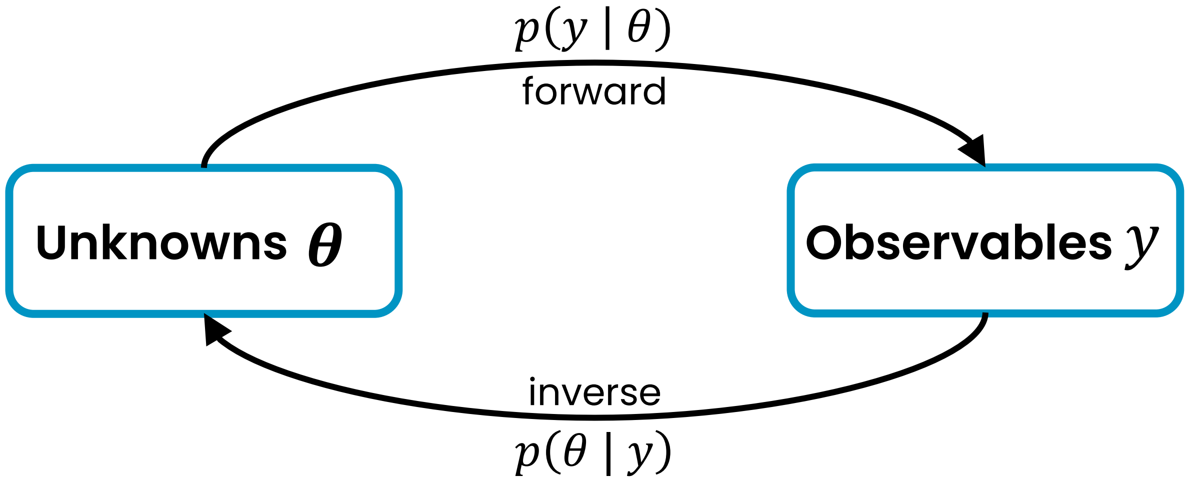

Soham Mukherjee


Maximilian Scholz


Luna Fazio

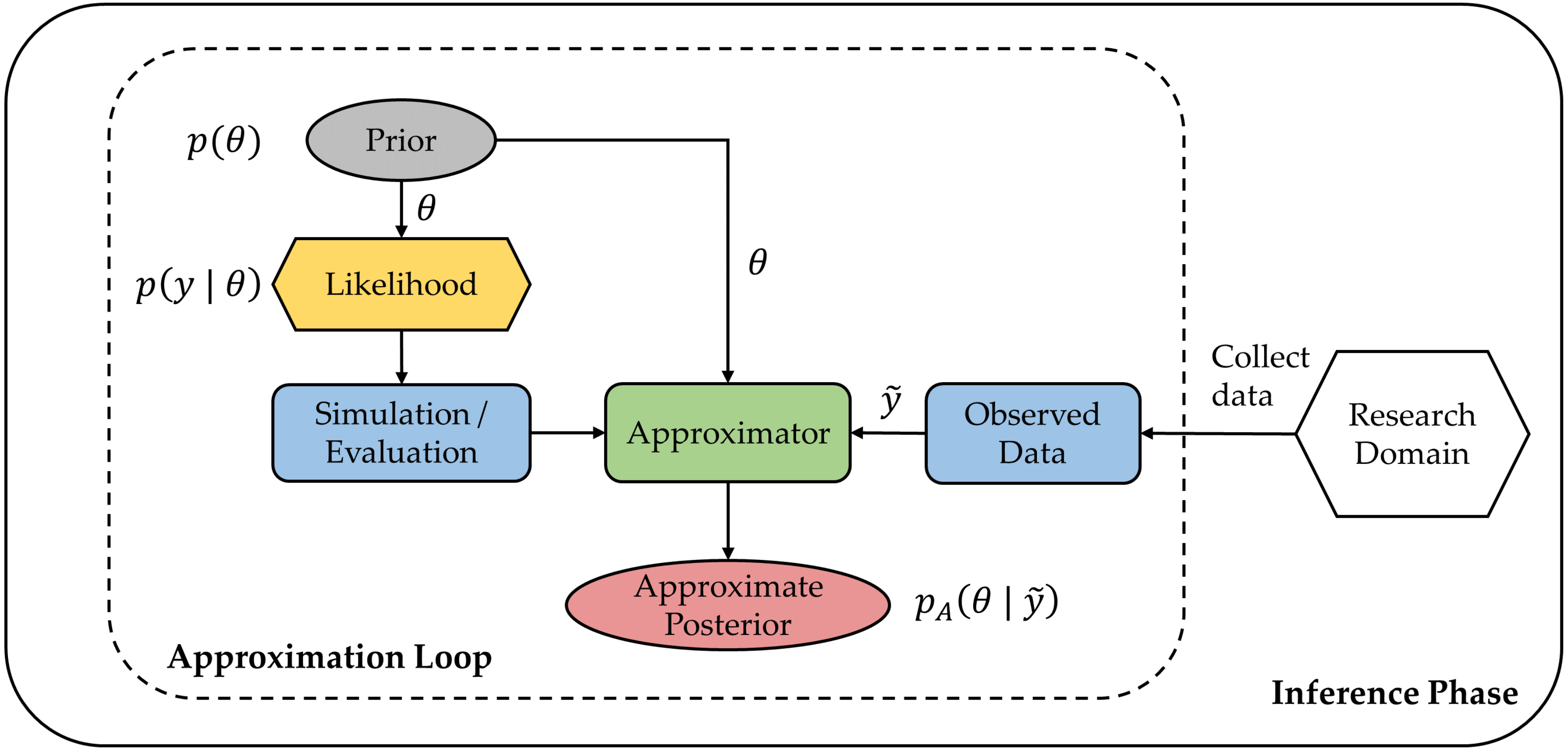

Javier Aguilar

Inverse Problems



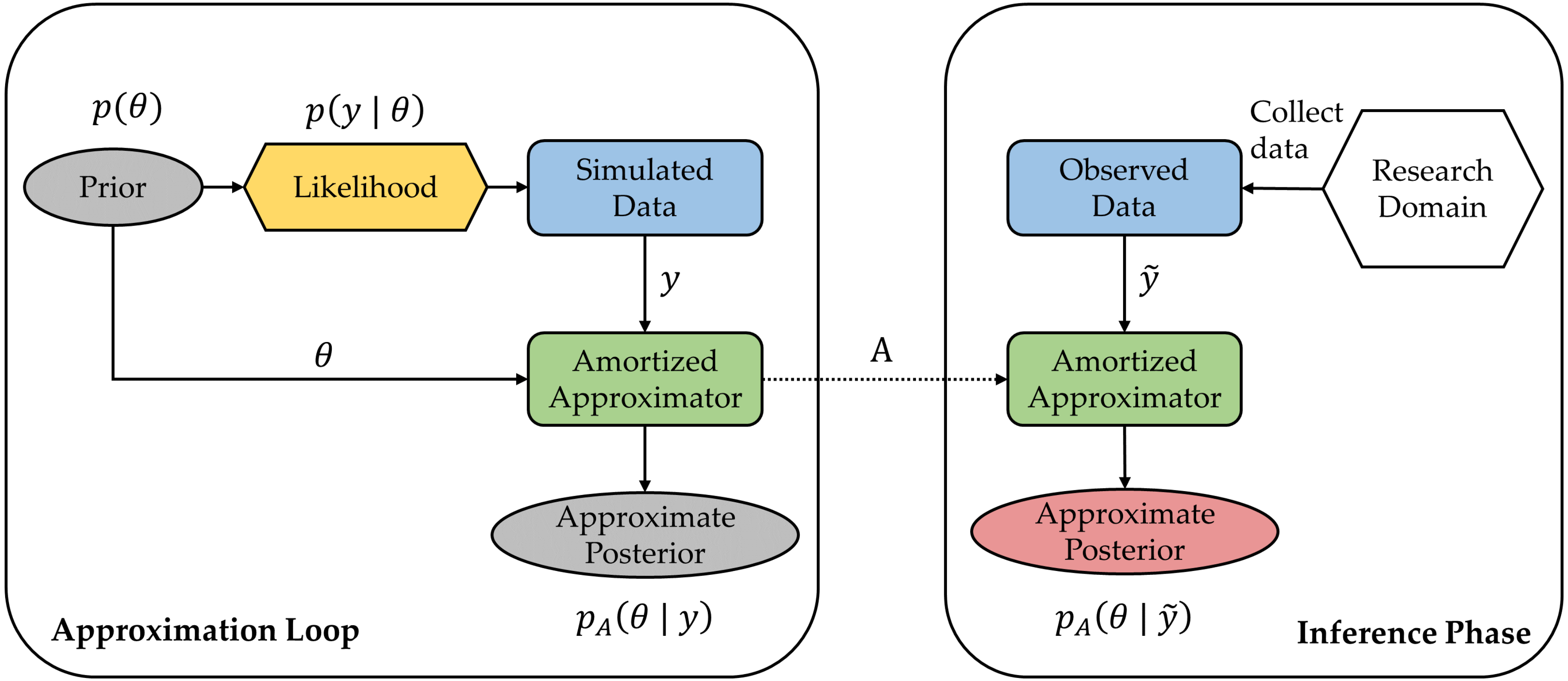
Statistical modeling:	**Parameters $\boldsymbol{\theta}$**	**Data $\boldsymbol{y}$**
Epidemiology:	Virus attributes	Infection curve (time series)
Image processing:	Crisp image	Blurry image
Psychology:	Cognitive parameters	Reaction times

Non-amortized Bayesian inference



Approximation and inference are **coupled**. No resource pooling.

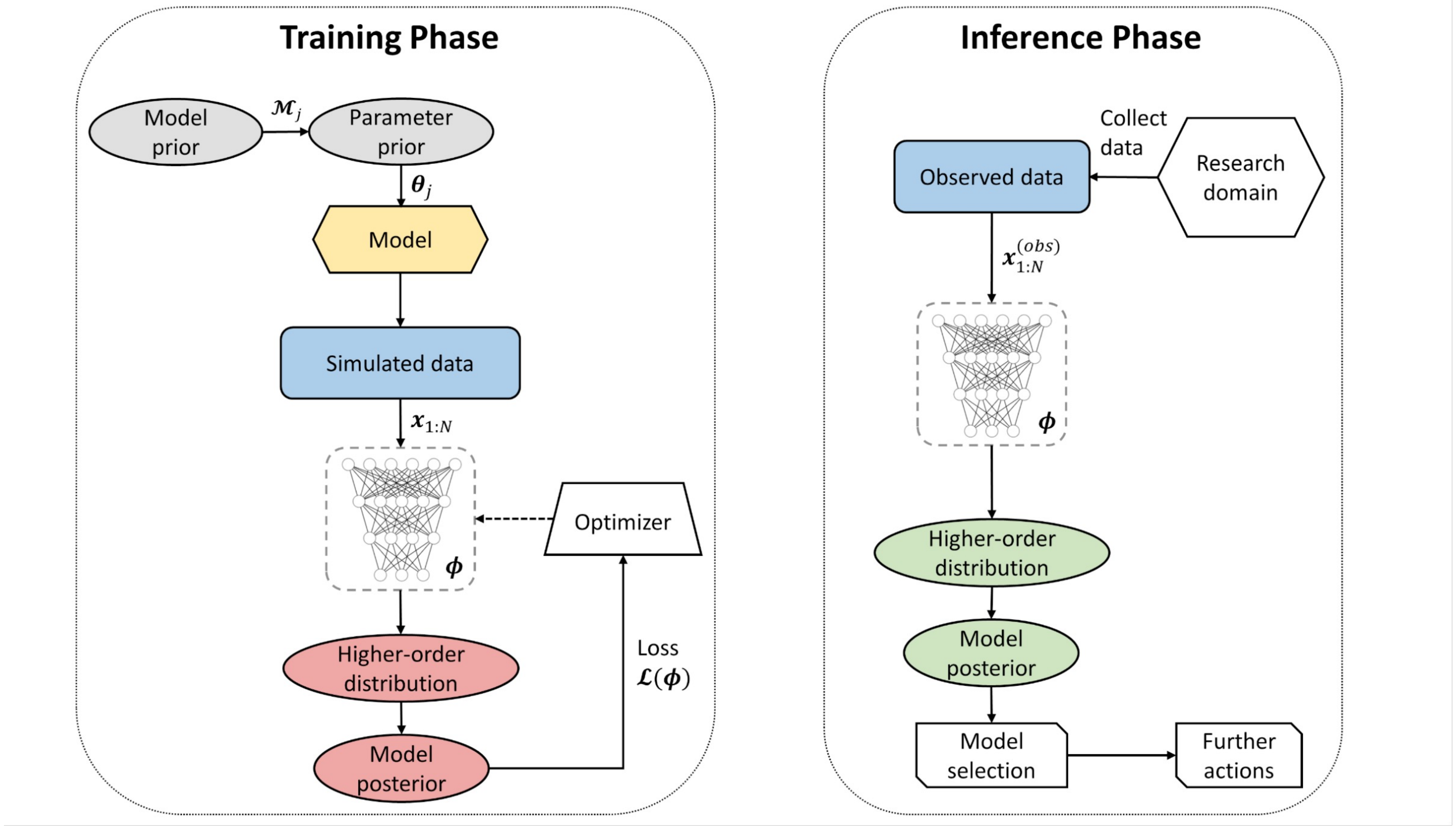
Amortized Bayesian inference (ABI)



Approximation and inference are **decoupled**. Pooling of resources.

https://arxiv.org/abs/2209.02439

Amortized Model Comparison



https://arxiv.org/abs/2004.10629

Loss Functions: Variational Inference

<u>Backward KL divergence</u>

$$\text{minimize} \quad KL(p_A(\theta \mid \tilde{y}) \,||\, p(\theta \mid \tilde{y})) = \int \log\left(\frac{p_A(\theta \mid \tilde{y})}{p(\theta \mid \tilde{y})}\right) p_A(\theta \mid \tilde{y})\, d\theta$$

$$\Longleftrightarrow \quad \text{maximize} \quad \frac{1}{S}\sum_{s=1}^{S} \log\left(\frac{p(\theta^{(s)}, \tilde{y})}{p_A(\theta^{(s)} \mid \tilde{y})}\right) \quad \text{for} \quad \theta^{(s)} \sim p_A(\theta \mid \tilde{y})$$

The ELBO just requires the joint density of the model

Loss Functions: Simulation-Based Inference

Forward KL divergence

$$\text{minimize} \quad KL(p(\theta \mid \tilde{y}) \,||\, p_A(\theta \mid \tilde{y})) = \int \log \left(\frac{p(\theta \mid \tilde{y})}{p_A(\theta \mid \tilde{y})} \right) p(\theta \mid \tilde{y}) \, d\theta$$

$$\iff \quad \text{maximize} \quad \frac{1}{S} \sum_{s=1}^{S} \log p_A(\theta^{(s)} \mid \tilde{y}) \quad \text{for} \quad \boxed{\theta^{(s)} \sim p(\theta \mid \tilde{y})}$$

Minimizing the expected KL over the whole data space

$$\text{minimize} \quad E_{p(y)} \left[KL(p(\theta \mid y) \,||\, p_A(\theta \mid y)) \right]$$

$$\iff \quad \text{maximize} \quad \frac{1}{S} \sum_{s=1}^{S} \log p_A(\theta^{(s)} \mid y^{(s)}) \quad \text{for} \quad \boxed{(\theta^{(s)}, y^{(s)}) \sim p(\theta, y)}$$

How to obtain the posterior approximator?

Sample $z^{(s)} \sim \mathrm{multinormal}(0, I)$

Transform to $\theta^{(s)} = f_\phi(z^{(s)} \mid y)$ with an invertible neural network

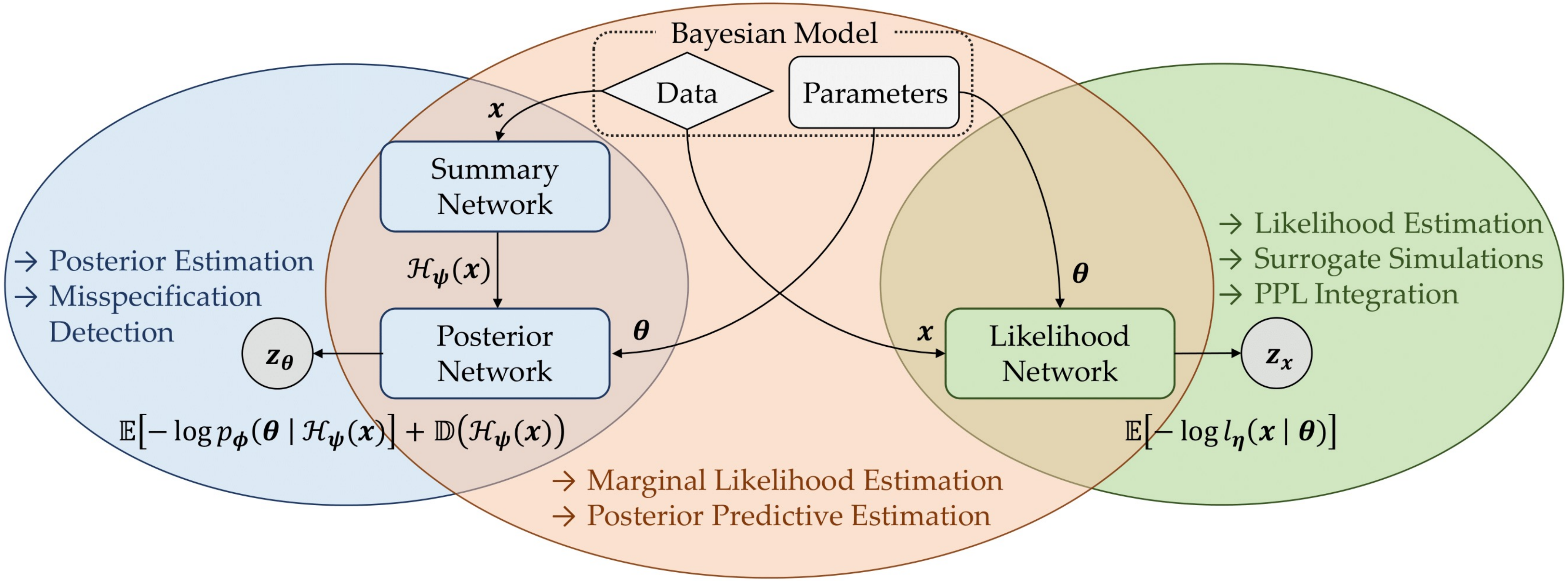
Obtain the approximator's density for training via expected KL divergence:

$$p_A(\theta \mid y) = p(z = f_\phi^{-1}(\theta \mid y)) \left| \det \left(\frac{\delta f_\phi^{-1}(\theta \mid y)}{\delta \theta} \right) \right|$$

https://arxiv.org/abs/2003.06281

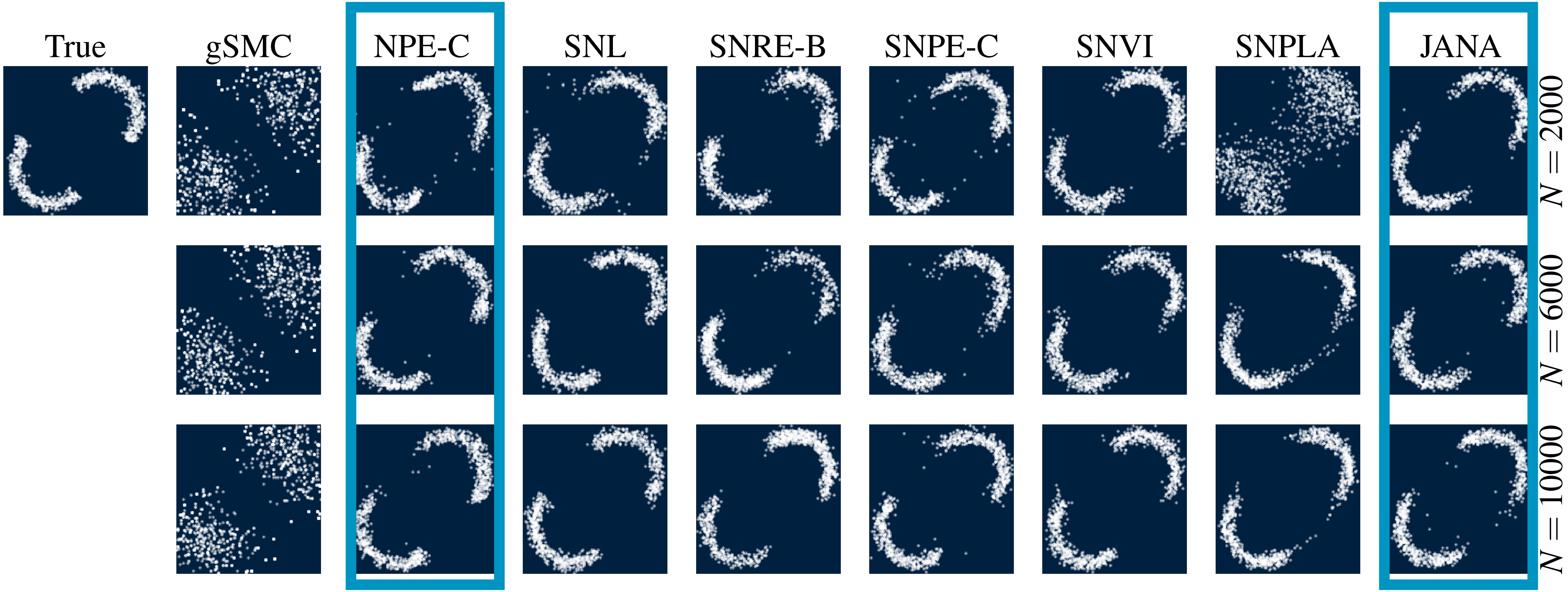
Jointly amortized learning

- Jointly amortized neural approximation (JANA)



https://arxiv.org/abs/2302.09125

Isn't amortized inference wasteful?



Amortized methods perform on-par with non-amortized counterparts

https://arxiv.org/abs/2302.09125

Potential of Amortized Bayesian Inference

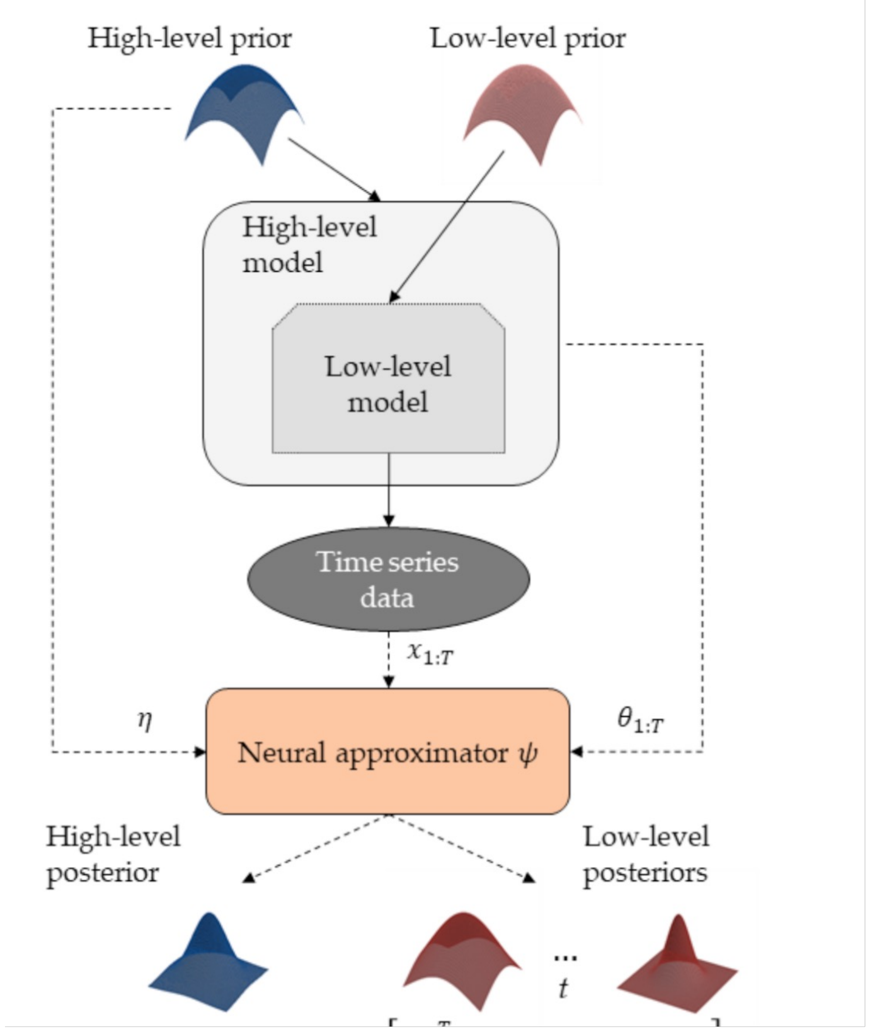
Massive number of inference repetitions

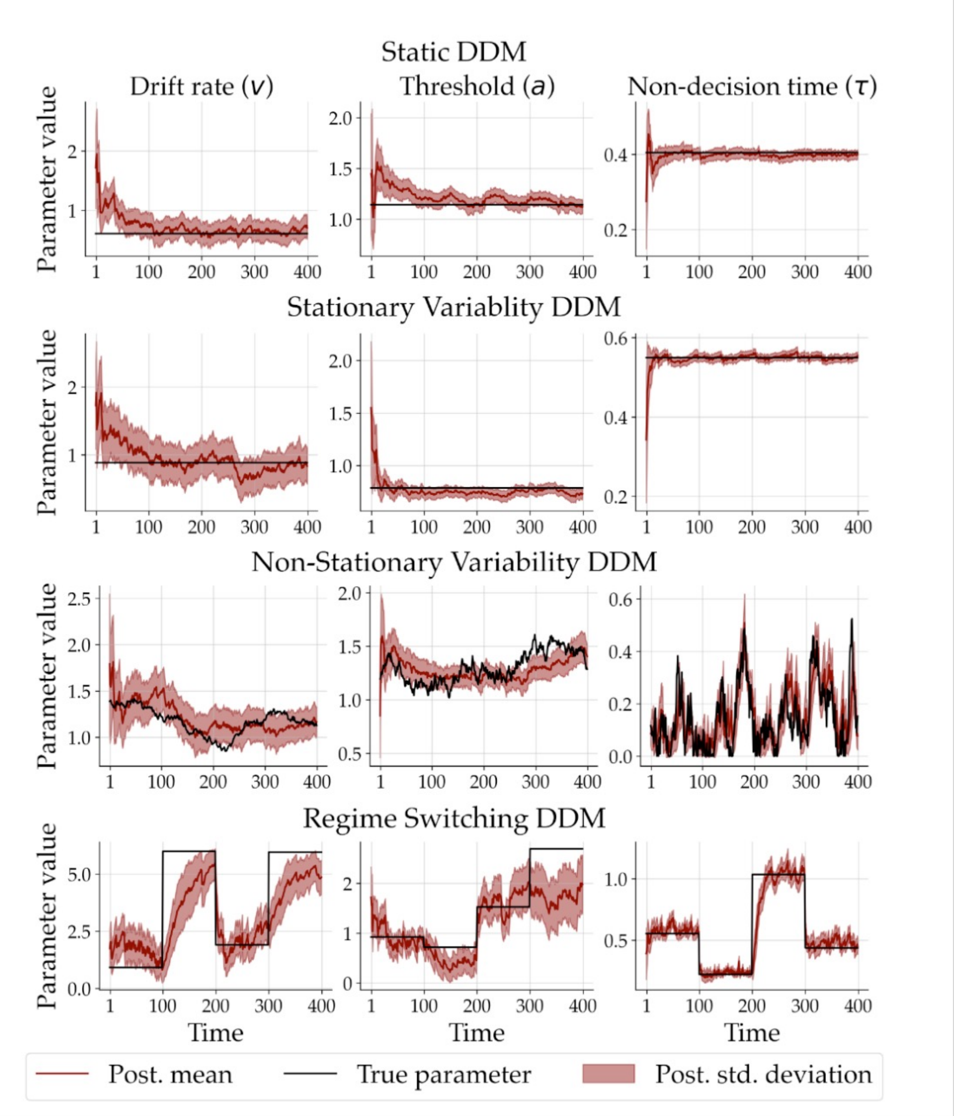
- Many data sets
- Cross-validation
- Sensitivity analyses, multiverse analyses

Real-time inference

- Neurological monitoring
- Adaptive experimental design
- Disease surveillance

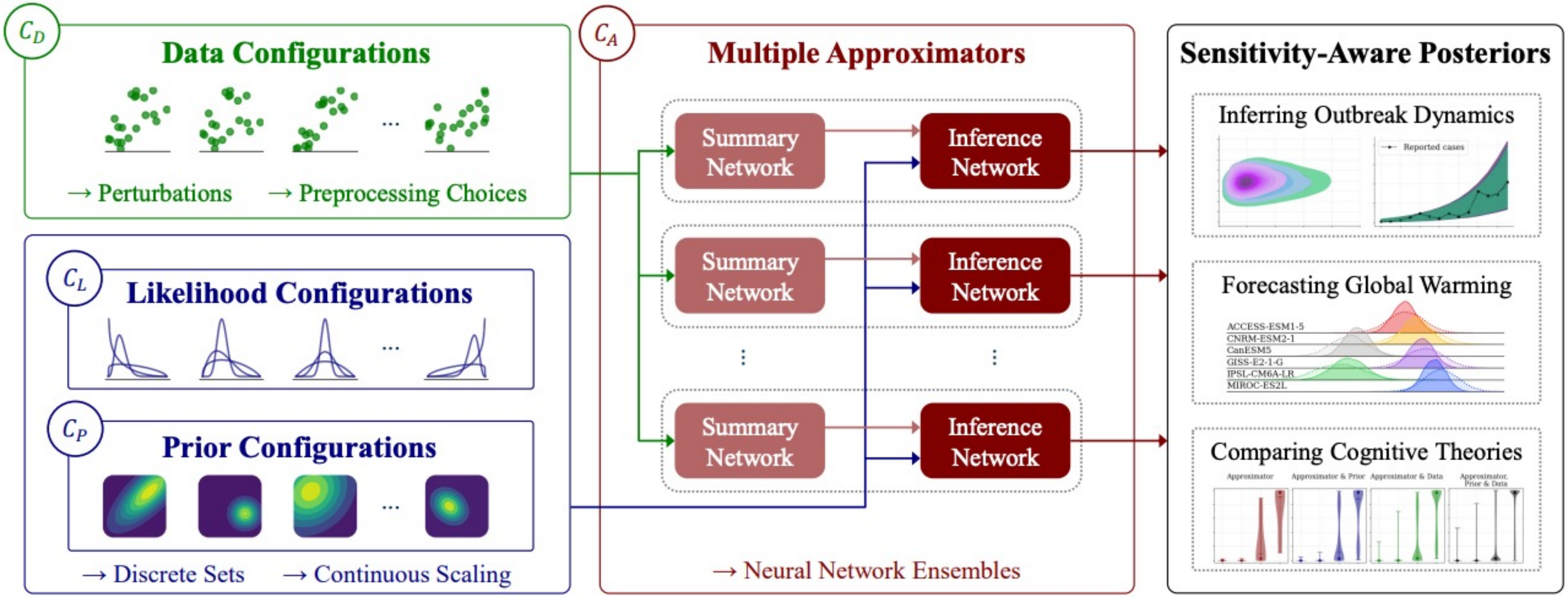
Dynamic Hierarchical Modeling





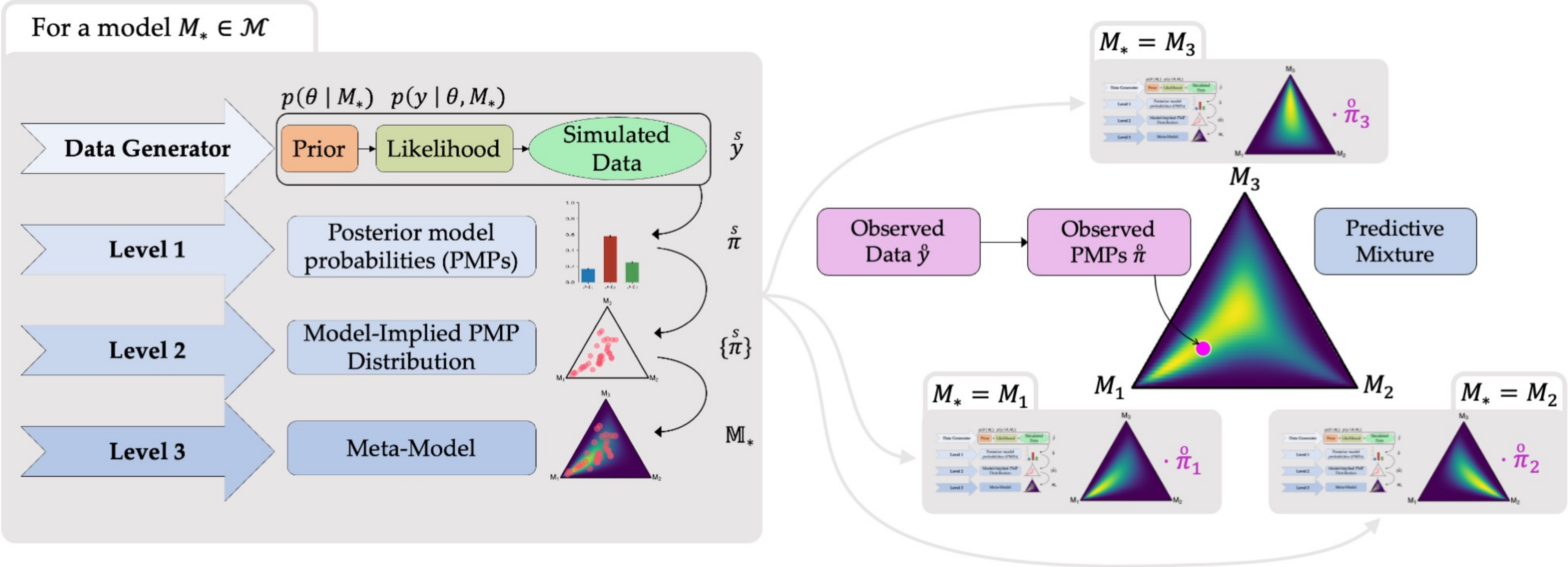
 https://arxiv.org/abs/2211.13165

Amortized sensitivity analyses



https://arxiv.org/abs/2310.11122

Meta-Uncertainty in Bayesian Model Comparison



https://arxiv.org/abs/2210.07278

Key Challenges of ABI

- Neural networks have a bad user experience
- Heaps of simulated training data necessary
- Constrained neural network architecture of normalizing flows
- Model misspecification invalidates simulation-based training

ABI library: BayesFlow

BayesFlow

Tests passing | License MIT | JOSS 10.21105/joss.05702 | contributions welcome

Welcome to our BayesFlow library for efficient simulation-based Bayesian workflows! Our library enables users to create specialized neural networks for *amortized Bayesian inference*, which repay users with rapid statistical inference after a potentially longer simulation-based training phase.

BayesFlow

Installation

```
pip install bayesflow
```

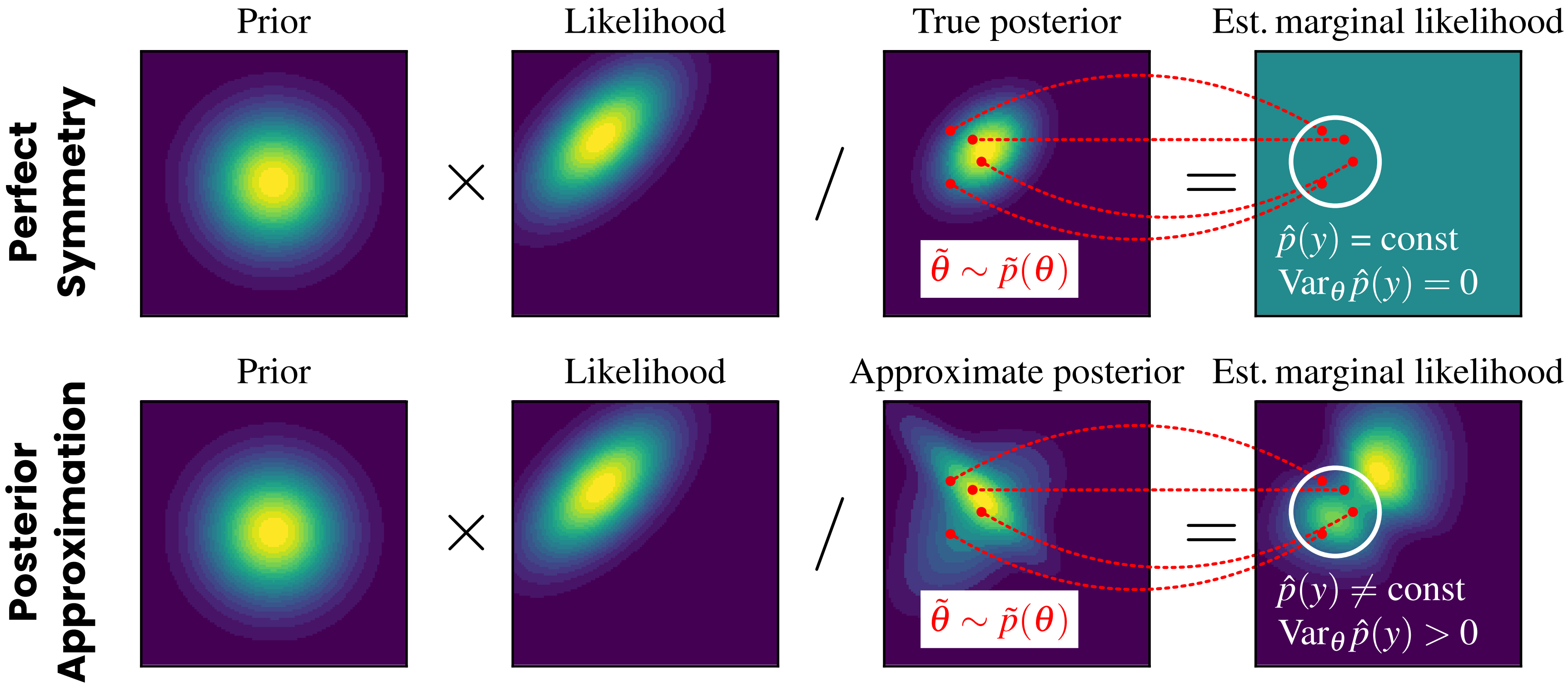
Documentation + Support

- www.bayesflow.org
- discuss.bayesflow.org

https://arxiv.org/abs/2306.16015

Low data → self-consistency

$$\boxed{p(y) = p(\theta)\, p(y \,|\, \theta) / p(\theta \,|\, y)} \implies \boxed{\frac{p(\tilde{\theta}_1)\, p(y \,|\, \tilde{\theta}_1)}{p(\tilde{\theta}_1 \,|\, y)} = \ldots = \frac{p(\tilde{\theta}_K)\, p(y \,|\, \tilde{\theta}_K)}{p(\tilde{\theta}_K \,|\, y)} \quad \tilde{\theta}_1, \ldots, \tilde{\theta}_K \in \Theta}$$



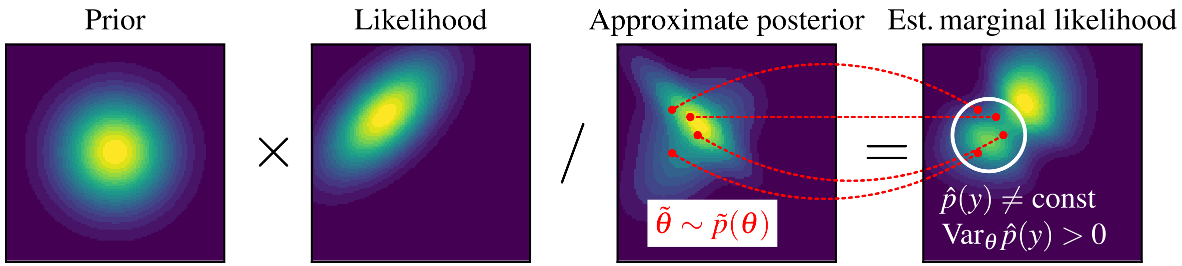
Low data → self-consistency

- **Idea:** Violations of self-consistency property as loss function

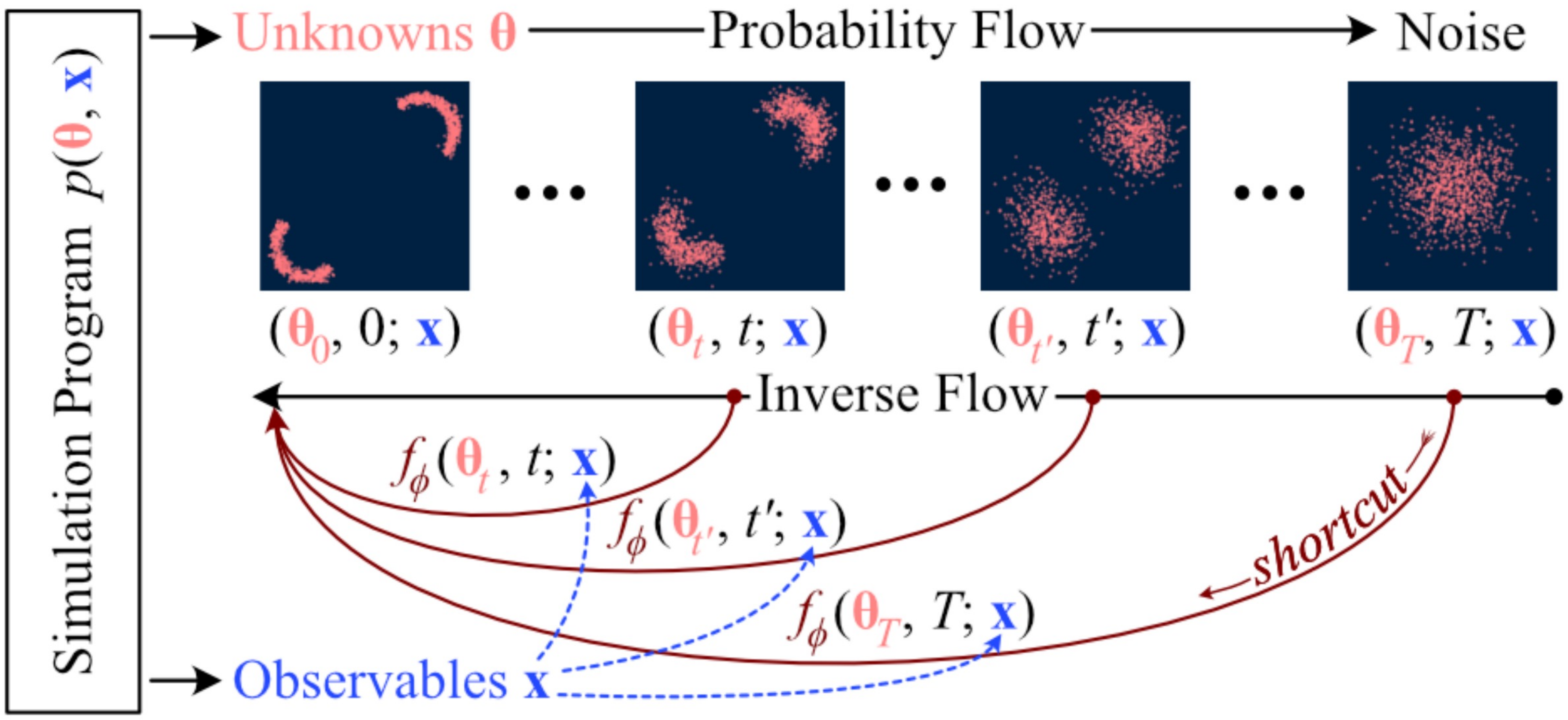
$$\mathcal{L}_{\text{SC}} := \mathbb{E}_{p(y)}\left[\operatorname{Var}_{\tilde{\theta}\sim\tilde{p}(\theta)}\left(\log p(\tilde{\theta}) + \log p(y \mid \tilde{\theta}) - \log q_\phi(\tilde{\theta} \mid y)\right)\right]$$

- **Integration** into standard neural posterior estimation loss

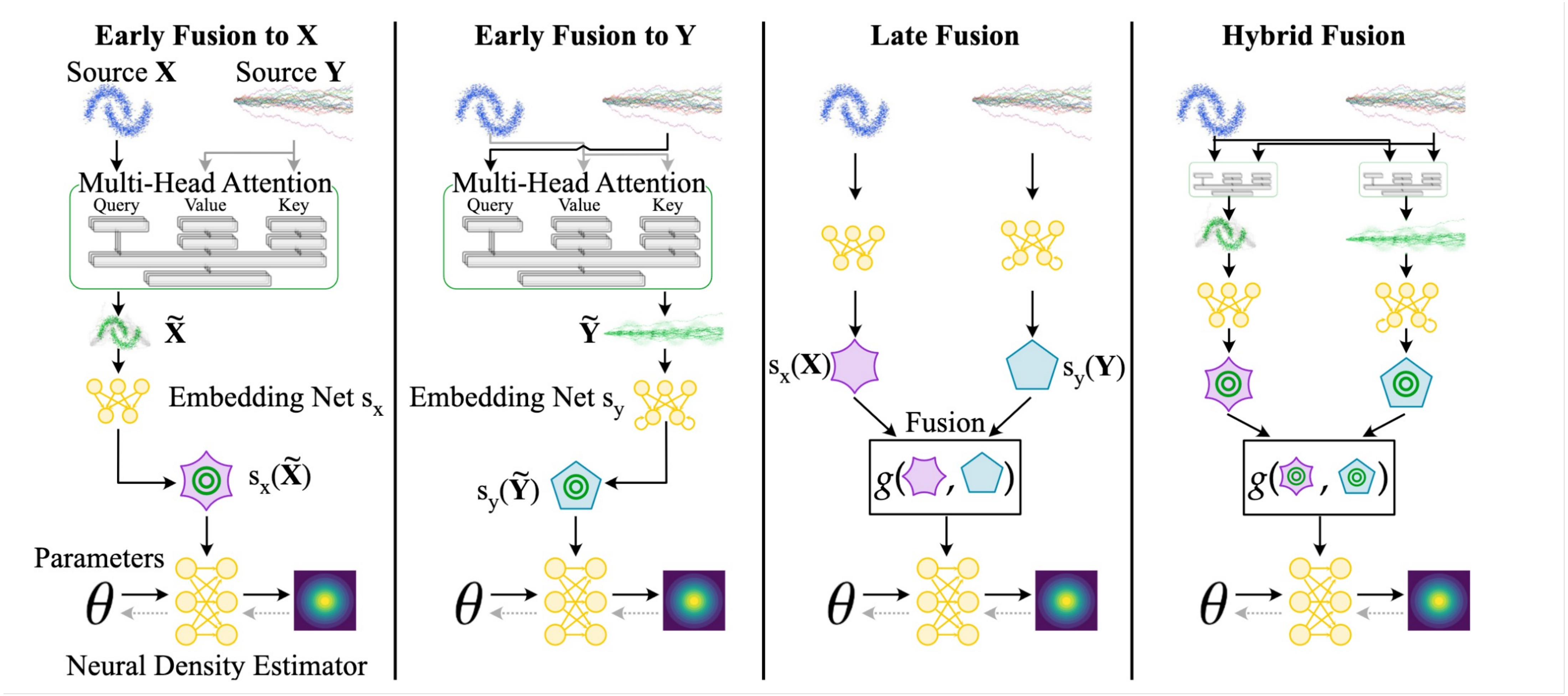
$$\mathcal{L}_{\text{NPE-SC}} := \mathbb{E}_{p(y)}\Bigg[\underbrace{\mathbb{E}_{p(\theta \mid y)}\left[-\log q_\phi(\theta \mid y)\right]}_{\text{NPE loss}} + \underbrace{\lambda \operatorname{Var}_{\tilde{\theta}\sim\tilde{p}(\theta)}\left(\log p(\tilde{\theta}) + \log p(y \mid \tilde{\theta}) - \log q_\phi(\tilde{\theta} \mid y)\right)}_{\text{self-consistency loss } \mathcal{L}_{\text{SC}} \text{ with weight } \lambda\in\mathbb{R}_+}\Bigg]$$



Neural network constraints → consistency models

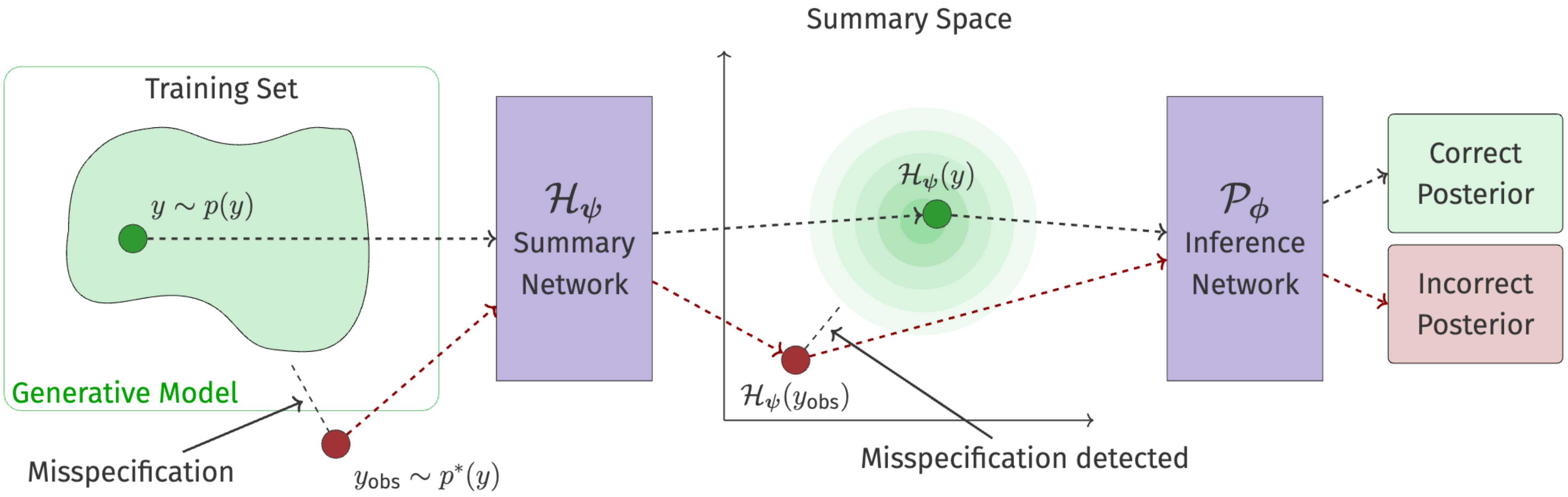


Fusing heterogeneous data sources

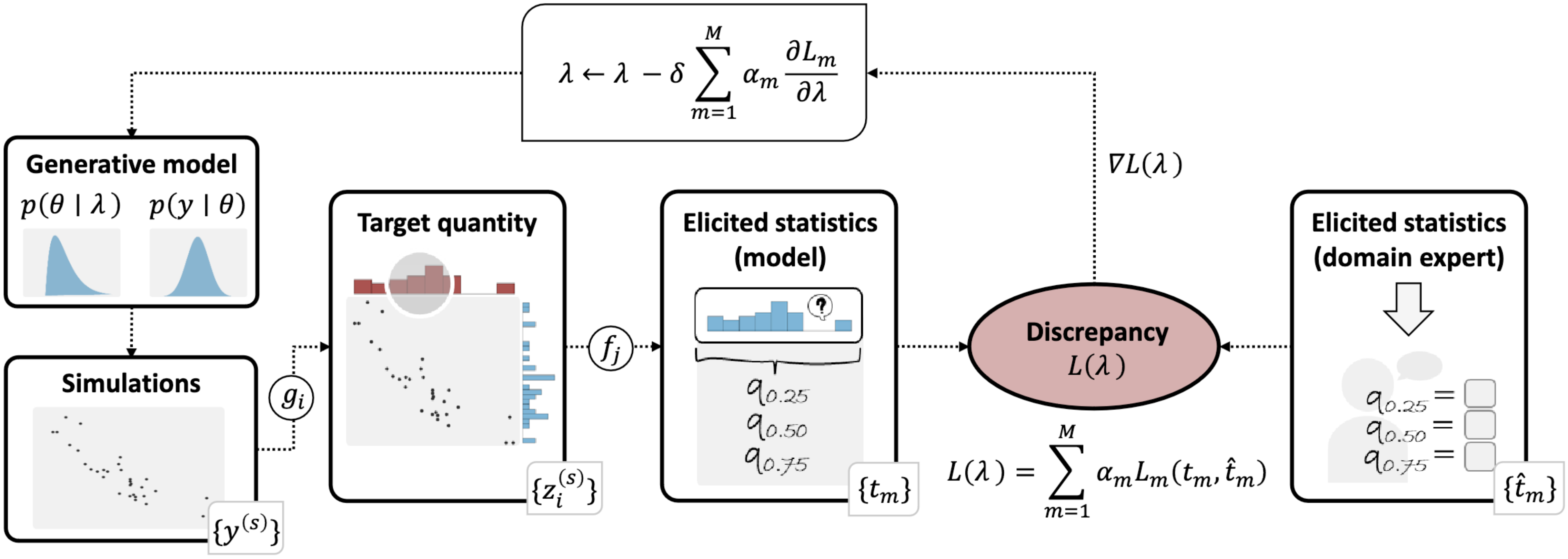


https://arxiv.org/abs/2311.10671

Model misspecification → detection



Simulation-Based Prior Elicitation



Summary

Amortized Bayesian inference

- High potential for large scale applicability
- Biggest issues: Reliability and trustworthiness
- Some questions have been tackled
- A lot of questions remain open

If you are interested in working with us, please reach out!