Addressing Misspecification in Simulation-based Inference through Data-driven Calibration

A. Wehenkel[1], **J. L. Gamella**[2]*, O. Sener[1], J. Behrmann[1], G. Sapiro[1], J. Jacobsen*, M. Cuturi[1]

ICML 2025 · [1]: Apple — [2]: ETH Zürich — *: work done while being at Apple

Abstract

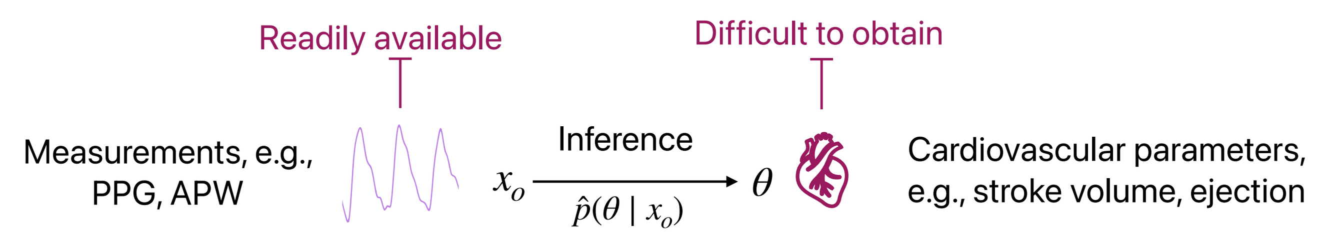
- Simulation-based inference (SBI) provides a powerful framework to estimate parameters and quantify uncertainty from black-box stochastic simulators. [1]
- While SBI has enabled breakthroughs across science and engineering, its real-world impact is limited by a key weakness: sensitivity to model misspecification. [3]

Question: *How can we make SBI algorithms more robust to misspecification?*

Our idea: *Jointly leverage the simulator, unlabelled data, and, if accessible, small amounts of labeled data to model misspecification with optimal transport.*

Motivation

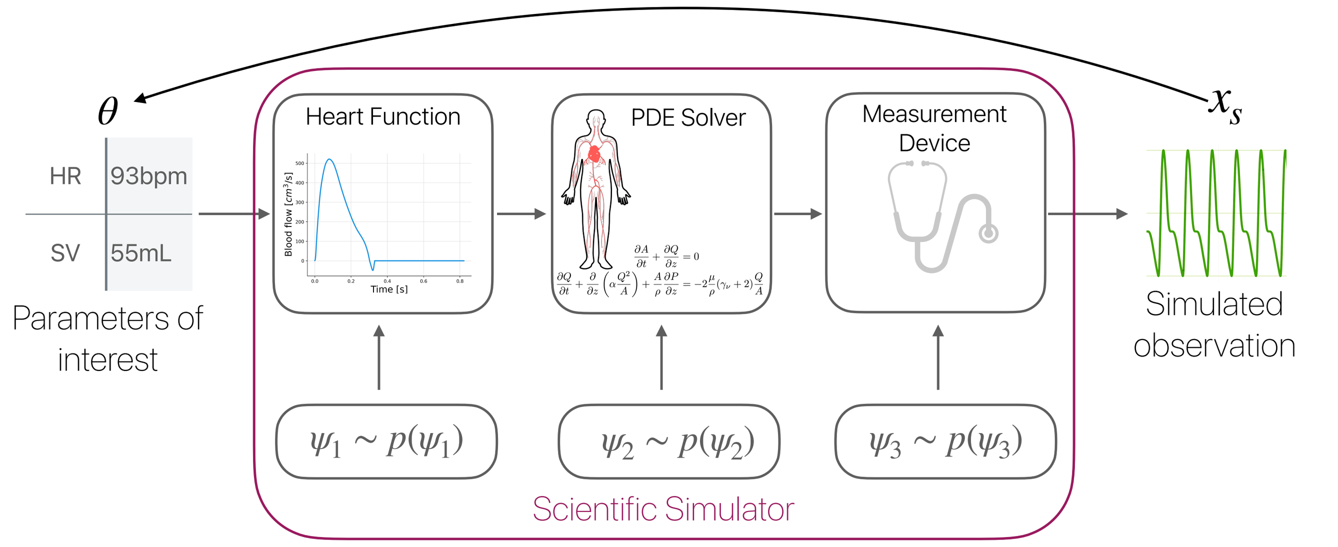
Uncertainty quantification under scarce data



Solution is often non-deterministic

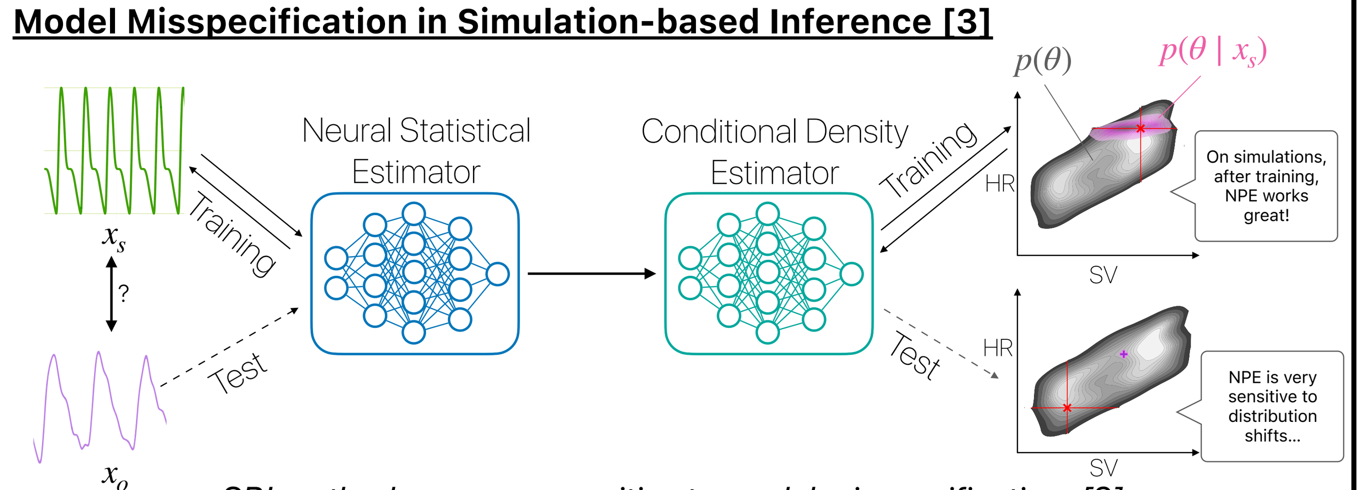
Scientific simulators as a source of data

Inverse Problem — Estimate $p(\theta \mid x_s)$



One can train a neural posterior estimator on simulated data!

Model Misspecification in Simulation-based Inference [3]

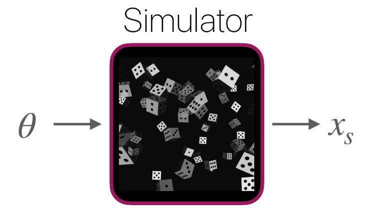


SBI methods are very sensitive to model misspecification. [3]

Setting considered:

Small amount of labelled data $\{(\theta^j, x_o^j)\}_{j=1}^{n_c}$

Large amount of measurements $\{(x_o^j)\}_{j=1}^{n_o}$



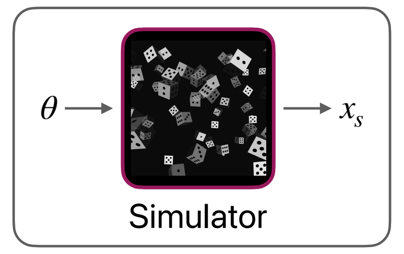
Method: *Optimal Transport for Robust Posterior Estimation (RoPE) in Simulation-based* Inference

Simulation-based Inference (SBI) [1]

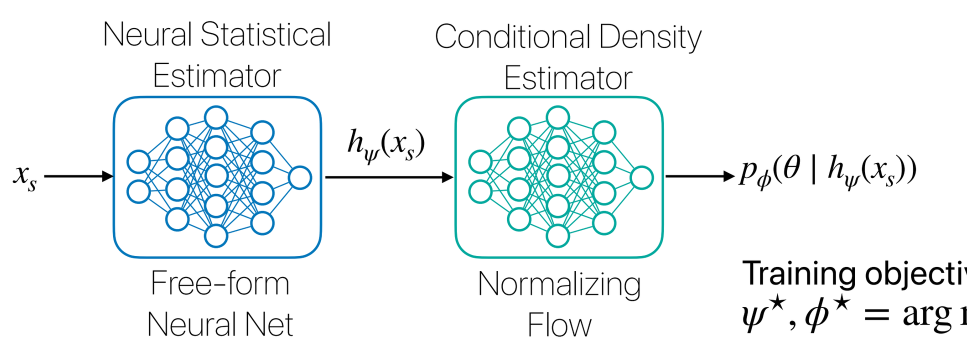
SBI enables statistical inference over the parameters of complex simulators.

1) Generate a large set of simulations.

One can generate as many samples from $p(\theta, x_s)$ as needed: $\{(\theta^j, x_s^j)\}_{j=1}^{n}$



2) Neural Posterior Estimation



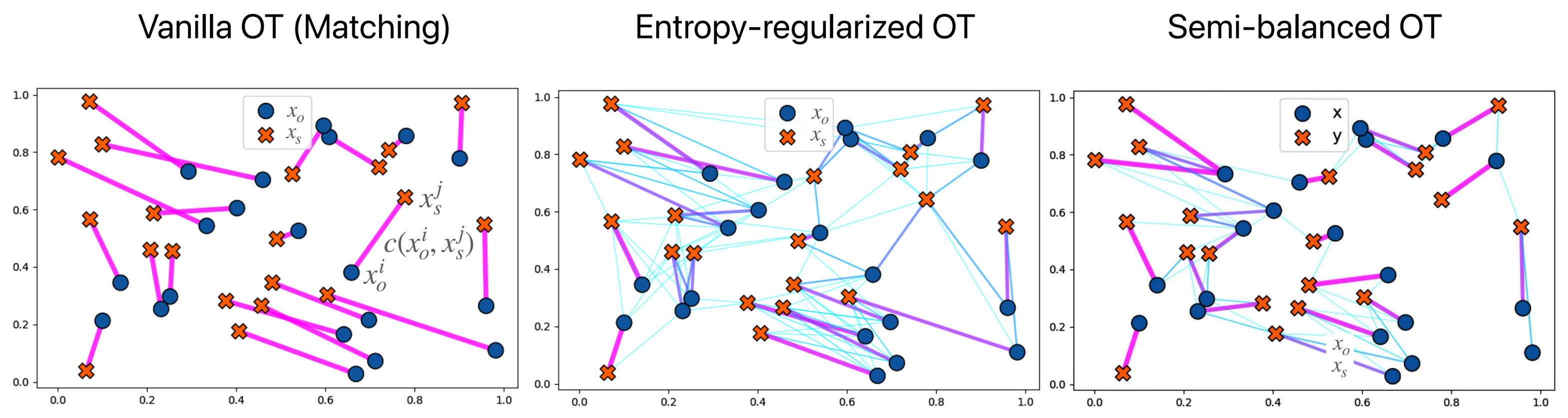
Training objective:

$$\psi^\star, \phi^\star = \arg\min_{\psi,\phi} KL[p(\theta, x_s) \,|\, p_\phi(\theta, h_\psi(x_s))]$$
$$= \arg\min_{\psi,\phi} \mathbb{E}_{p(\theta,x_s)}[\log p_\phi(\theta \mid h_\psi(x_s))]$$

One can make $h_\psi(x_s)$ arbitrarily close to a minimal sufficient statistics of θ.

Optimal Transport (OT) [2]

OT provides a flexible way to model the gap between simulated and real-world observations.

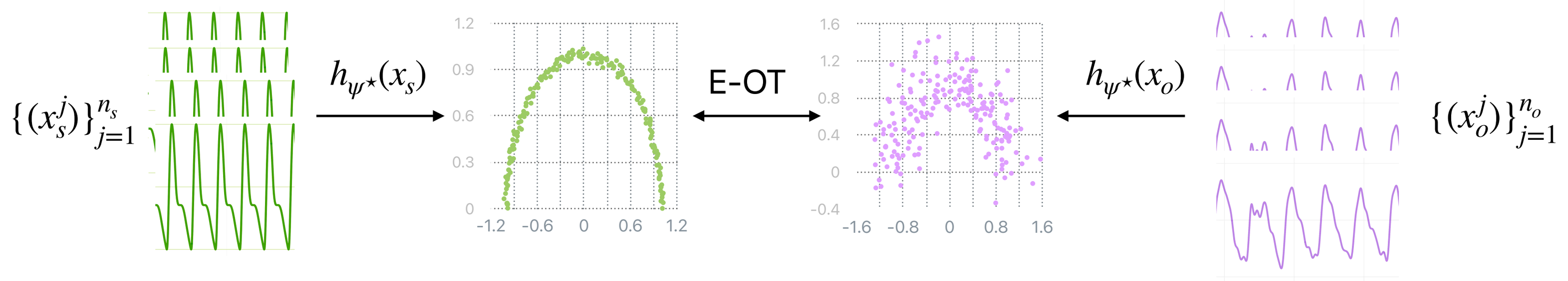


For arbitrary cost functions, solving a discrete OT problem costs $O(n_o^3)$

We use ***entropy-regularised semi-balanced*** OT, solved with a variant of the Sinkhorn algorithm in $O(n_o n_s)$

$$\mathbf{P}^\star = \arg\min_{\mathbf{P}\in\mathscr{B}_o} \left\langle \mathbf{P}, \mathbf{C} \right\rangle + \rho\, \mathrm{KL}\left(\mathbf{P}^\top \mathbf{1}_{n_o} \,\Big\|\, \frac{1}{n_s}\mathbf{1}_{n_s}\right) + \gamma \left\langle \mathbf{P}, \log \mathbf{P} \right\rangle,$$

The solution obtained depends strongly on the cost function chosen!
We propose to define it on statistics of θ rather than raw observations.

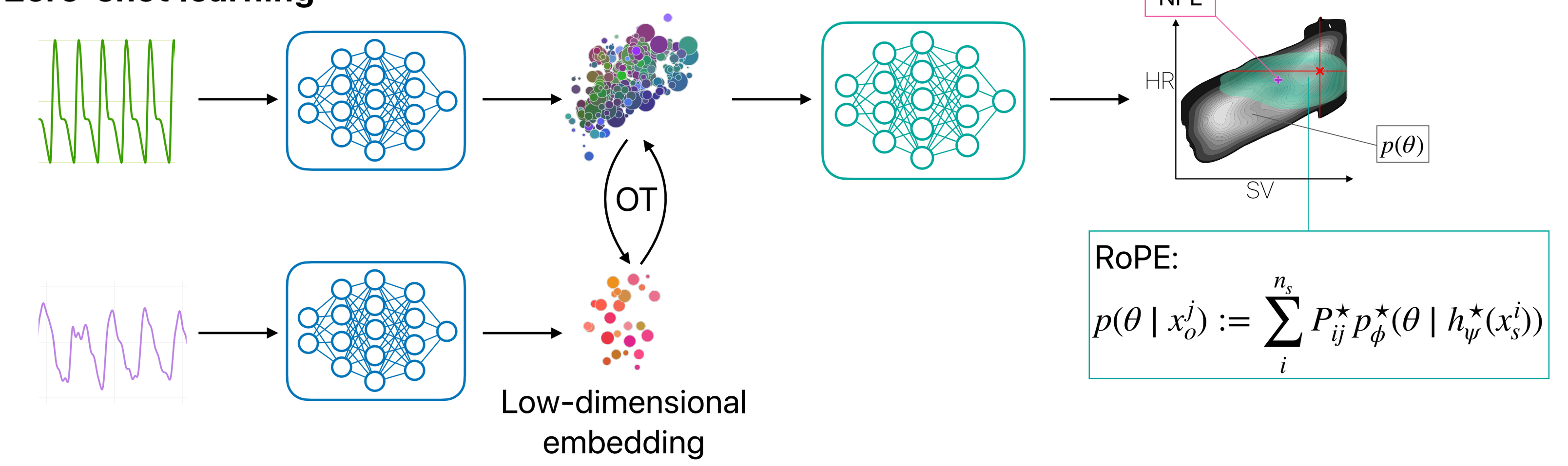


RoPE — A framework for Robust Posterior Estimation

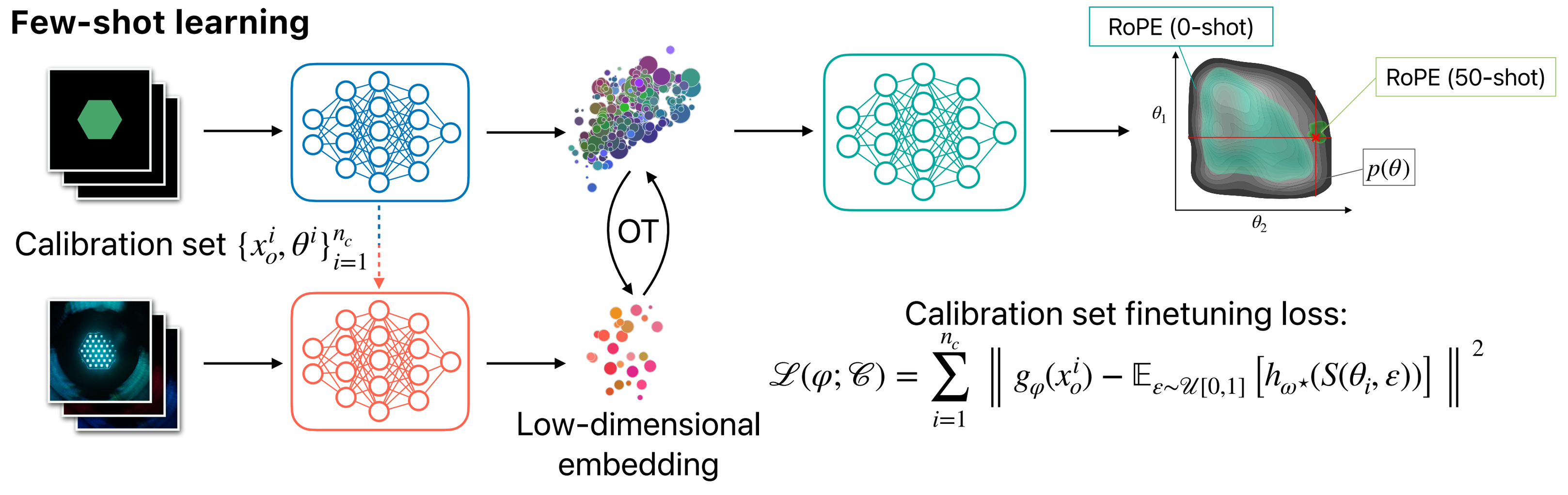
Modeling assumption: $x_o \perp \theta \mid x_s$

$$\theta \xrightarrow{\text{Simulator}} x_s \xrightarrow{\text{OT}} x_o \longleftrightarrow p(\theta \mid x_o) := \int p(\theta \mid x_s) p(x_s \mid x_o) \mathrm{d}x_s$$

Zero-shot learning



Few-shot learning

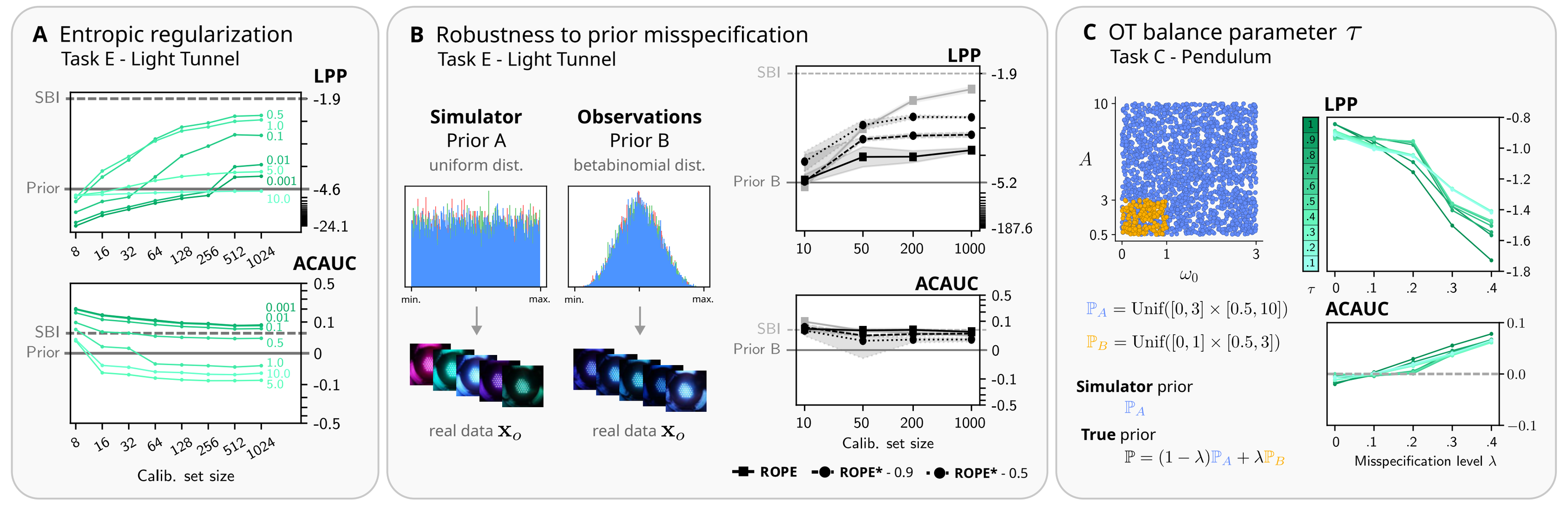


Calibration set finetuning loss:

$$\mathscr{L}(\varphi; \mathscr{C}) = \sum_{i=1}^{n_c} \left\| g_\varphi(x_o^i) - \mathbb{E}_{\epsilon\sim\mathscr{U}[0,1]}\left[h_{\omega^\star}(S(\theta_i, \epsilon))\right] \right\|^2$$

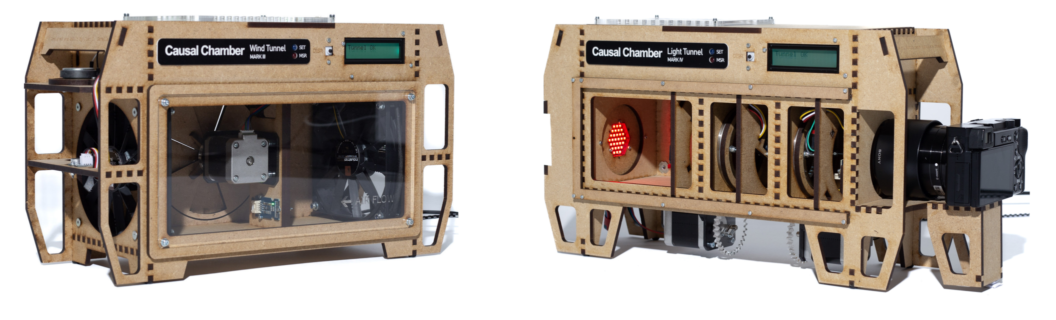
Properties of RoPE

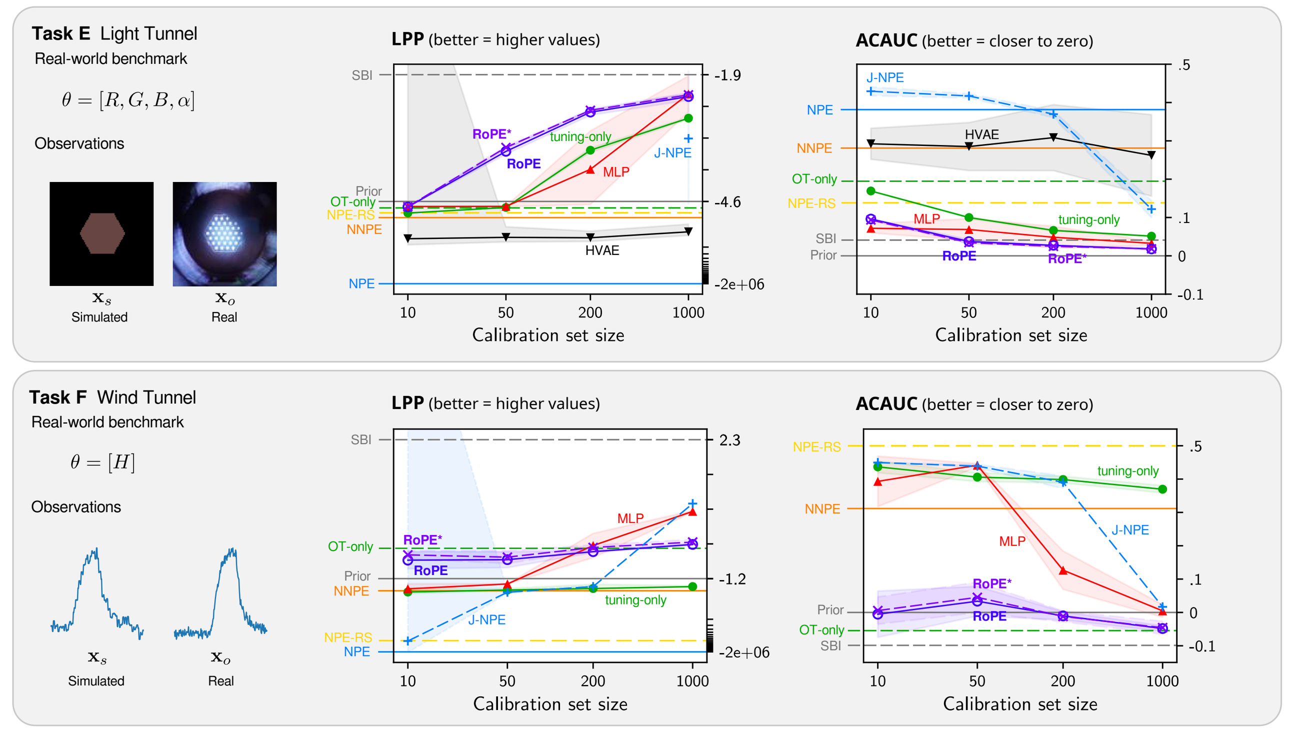
1. Self-calibration via entropy regularisation $E_{p(x_o)}[\tilde{p}(\theta \mid x_o)] = p(\theta)$
2. Control Mechanism for the posteriors' confidence $\gamma \to \infty, \quad \tilde{p}(\theta \mid x_o) \approx \frac{1}{n_s}\sum_{j=1}^{n_s} \tilde{p}(\theta \mid x_s^j) \approx p(\theta)$
3. Robustness to prior misspecification



Results

First benchmark with labeled real-world data





Take-home Messages

- RoPE enables reliable few-shot uncertainty quantification under model misspecification, leveraging a small set of real data to calibrate simulation-based inference.
- Optimal Transport (OT) provides a flexible, assumption-light mechanism to model misspecification, enabling control over calibration vs informativeness and robustness to prior shifts.
- We keep Neural Posterior Estimation (NPE) unchanged — preserving amortization and scalability — and layer on top a lightweight OT-based correction.
- No free lunch: any correction model creates a new posterior — RoPE makes this explicit and allows practitioners to tune the tradeoff with interpretable hyperparameters (γ and τ).
- Minimal assumptions, maximal reuse: RoPE treats simulators as imperfect but valuable priors, correcting only what's needed without discarding domain expertise.

Sources

[1]: Cranmer, Kyle, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference." Proceedings of the National Academy of Sciences 117.48 (2020): 30055-30062.
[2]: Peyré, Gabriel, and Marco Cuturi. "Computational optimal transport: With applications to data science." *Foundations and Trends® in Machine Learning* 11.5-6 (2019): 355-607.
[3]:Cannon, Patrick, Daniel Ward, and Sebastian M. Schmon. "Investigating the impact of model misspecification in neural simulation-based inference." *arXiv preprint arXiv:2209.01845* (2022).

© 2025 Apple Inc. All rights reserved.