Simulation-based inference of epidemic and phylodynamic models via Neural Posterior Estimation (NPE)

Francesco Pinotti[1], Julien Thézé[2], Xavier Bailly[2] and Guillaume Fournié[1,2,3]

[1] Université de Lyon, INRAE, VetAgro Sup, UMR EPIA, France
[2] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, France
[3] Department of Pathobiology and Population Sciences, Royal Veterinary College, UK

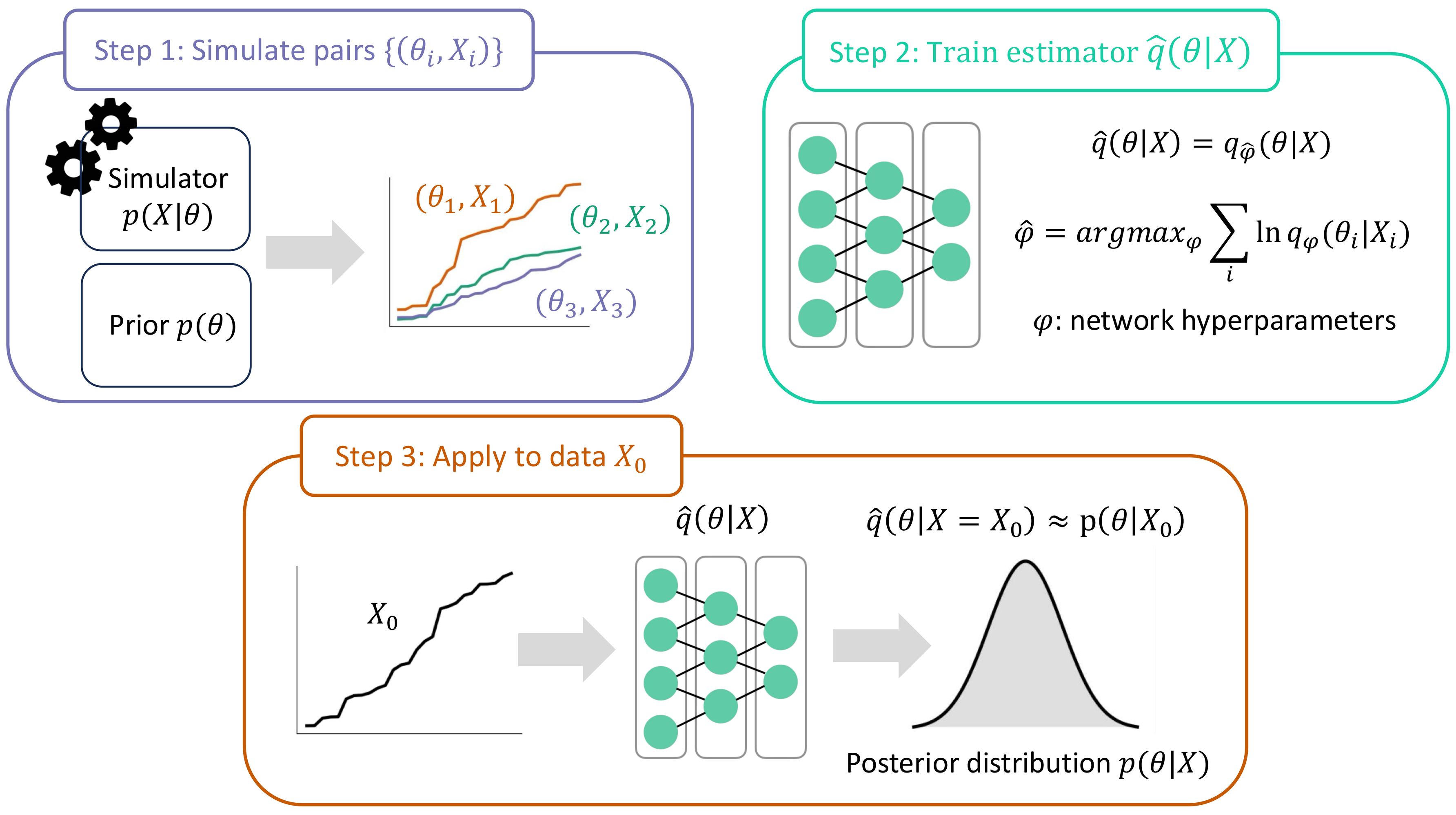
francesco.pinotti@inrae.fr

Background

Bayesian inference of transmission models can be made difficult by complicated likelihood functions. Simulation-based inference (SBI) methods use simulations to perform inference[a].

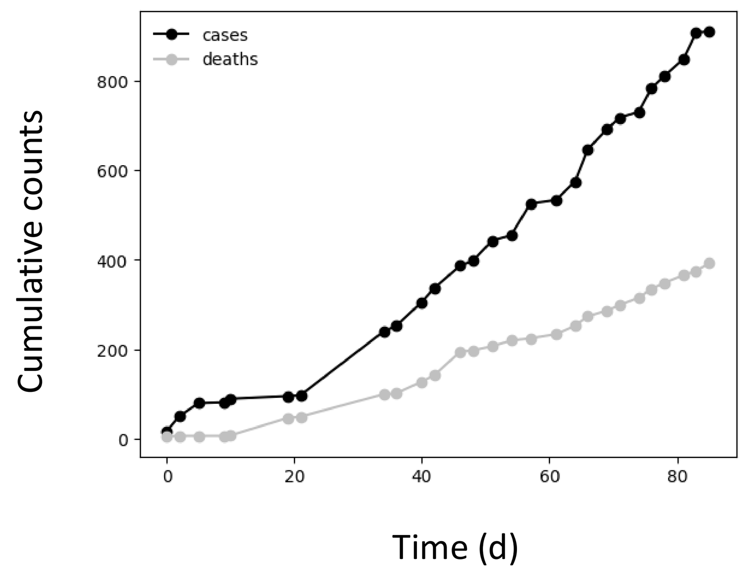
Traditional SBI methods	NPE
Target approximate posterior	Flow-based learning of exact posterior[c,d]
Ad-hoc selection of summary statistics[b]	Automatic detection of summary statistics from raw data via dedicated neural network
Potentially expensive to run and to calibrate	Training is costly but sampling is extremely fast. Inference is **amortized**, enabling powerful calibration checks

This study aims to assess the effectiveness of flow-based NPE in two case studies of epidemiological relevance.

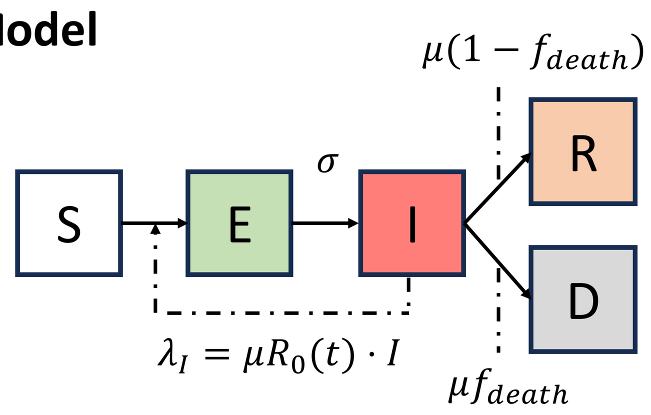


SEIRD model

Data: Time series of reported cases C_t and deaths D_t during the 2014 Ebola virus outbreak in Sierra Leone[e].



Model



$R_0(t) = R_0 e^{-k_\beta t}$

Reported cases/deaths follow a Negative Binomial distribution with expected counts set by the model:

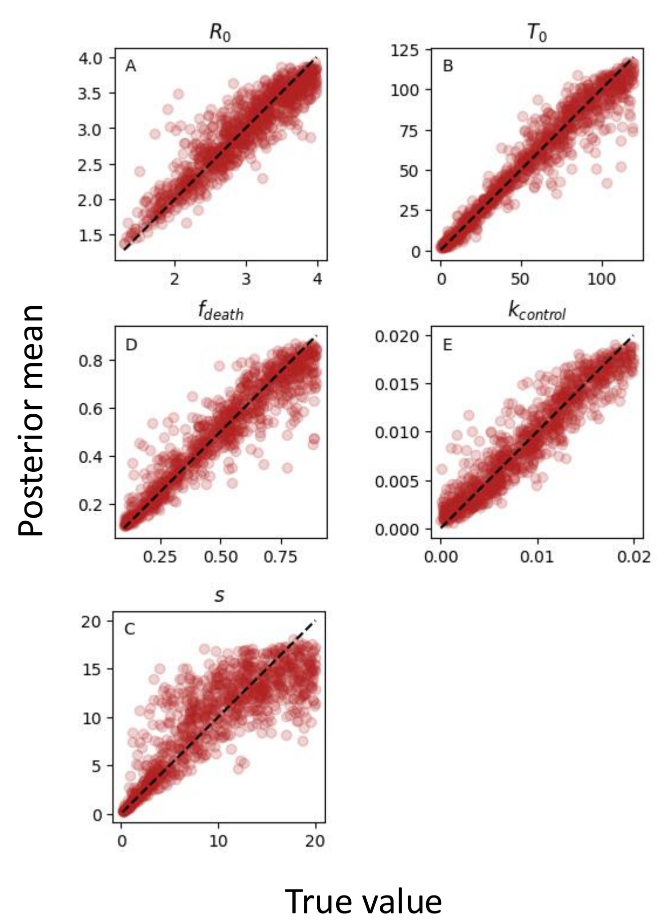
$C_t, D_t \sim NegBin(\langle C_t, D_t \rangle, s)$

Inferred parameters:

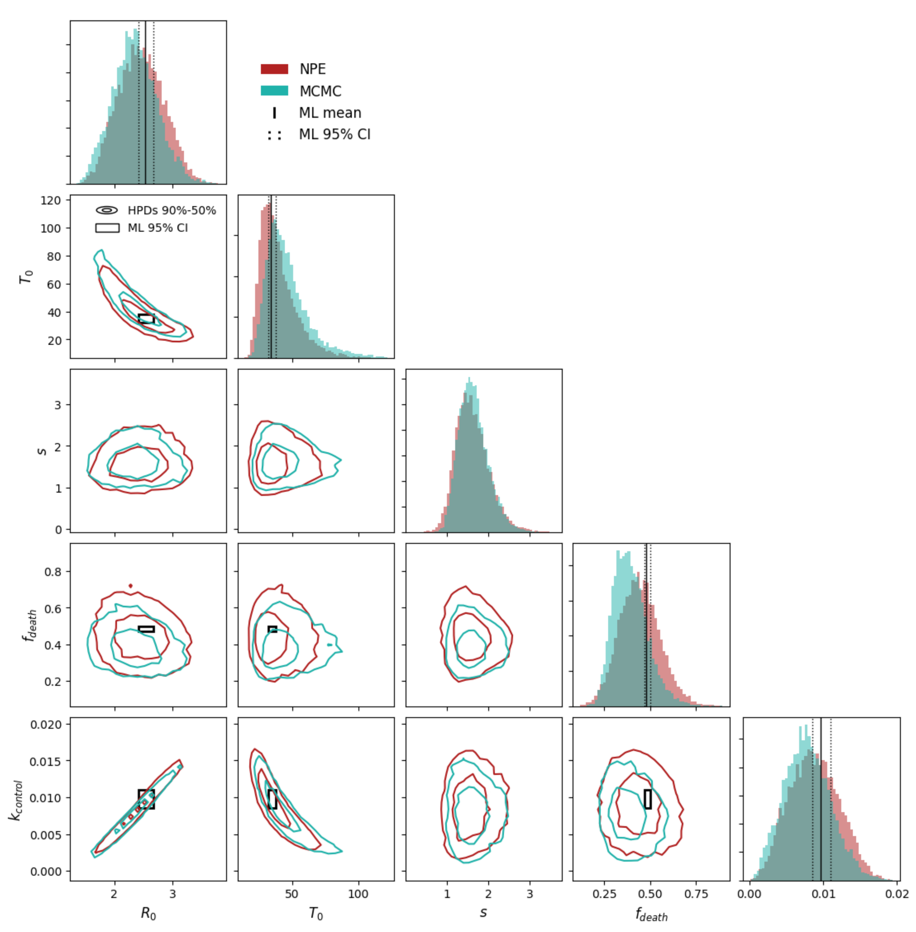
- R_0: basic reproduction number
- T_0: time to first reporting
- s: reporting distribution shape
- f_{death}: infection fatality rate
- k_β: effectiveness of interventions

Results

Simulated data



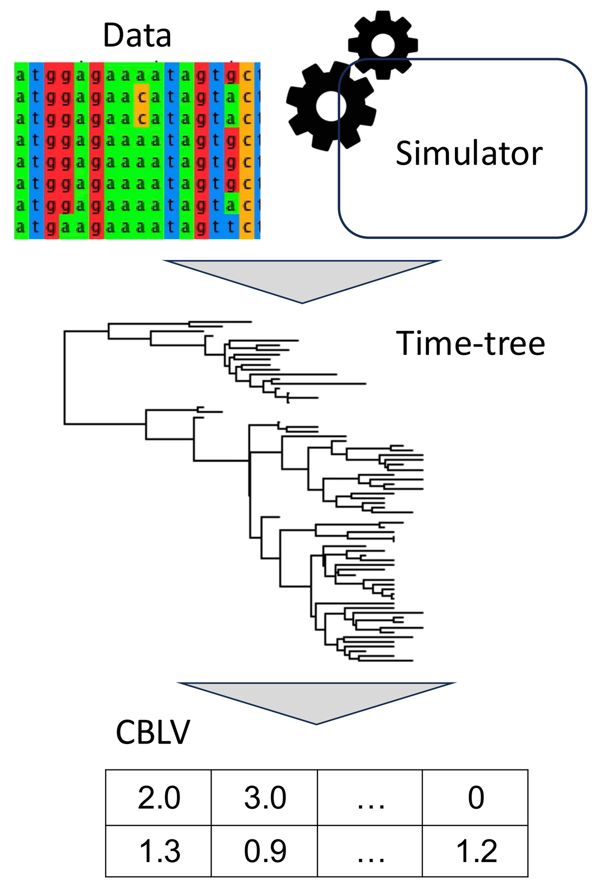
Sierra Leone data



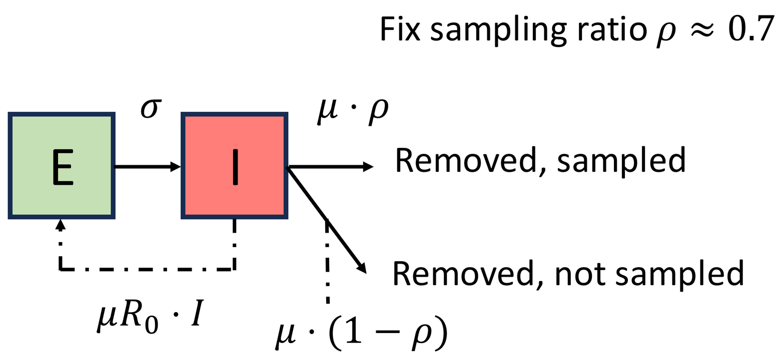
NPE retrieves the correct posterior.

Phylodynamics model

Data: 72 early genetic sequences from the 2014 Ebola virus outbreak in Sierra Leone. Model is fitted to estimated phylogenies[g,h].



Model



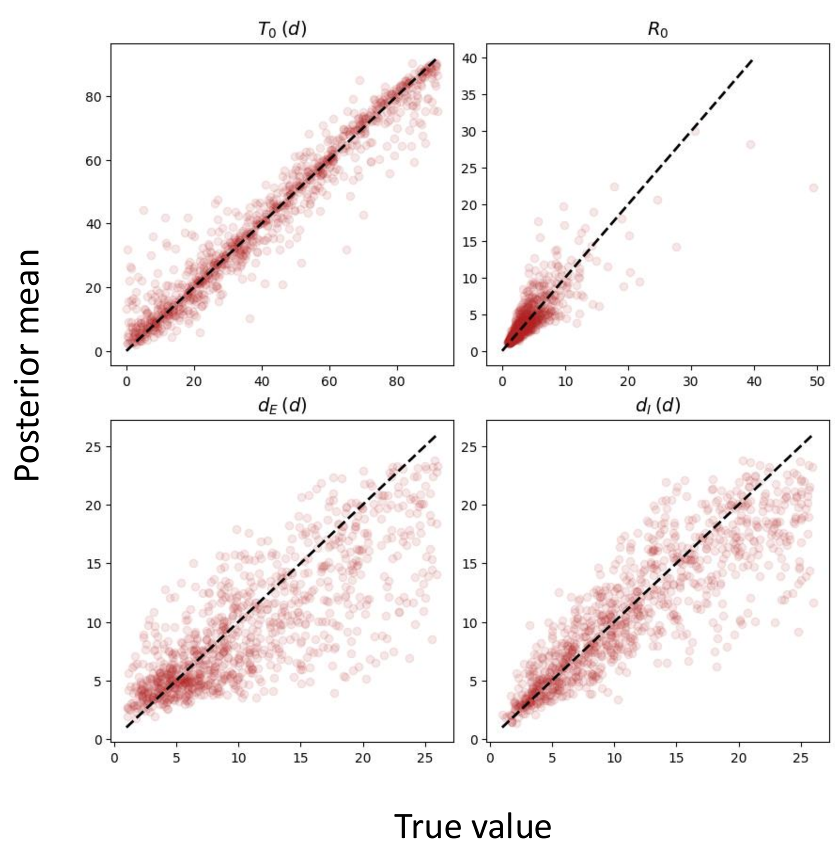
Observed and simulated trees are formatted into Compact Bijective Ladderized Vectors (CBLV)[i].

Inferred parameters:

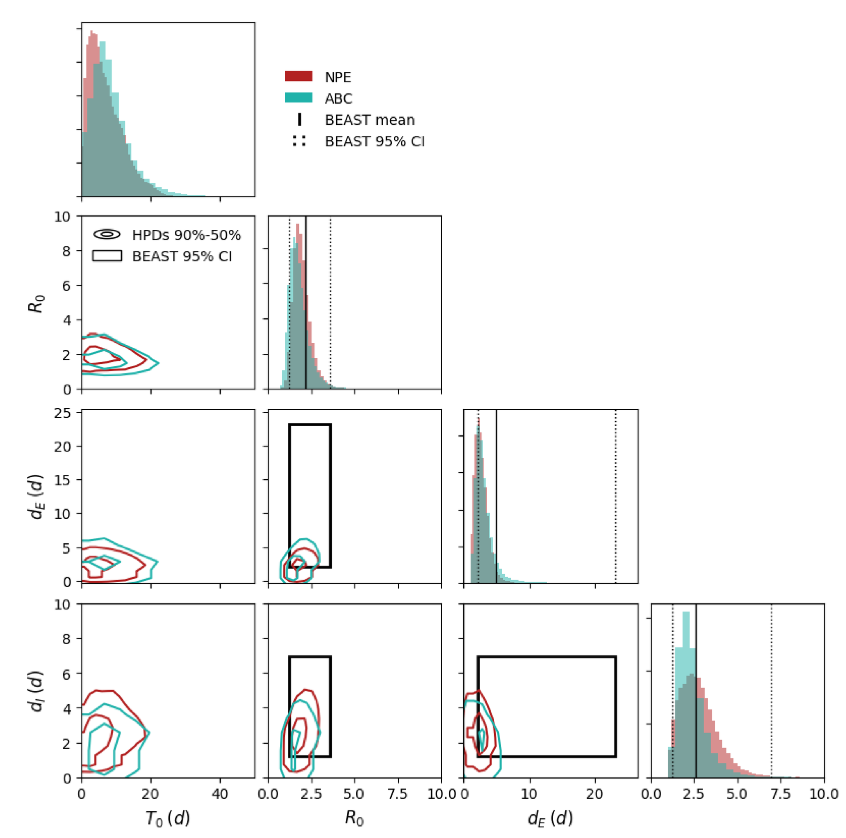
- R_0: basic reproduction number
- T_0: time to first reported cases
- $d_E = \sigma^{-1}$: incubation period
- $d_I = \mu^{-1}$: infectious period

Results

Simulated data



Sierra Leone data



NPE yields posterior estimates that agree with ABC ones without calculating tree summary statistics.

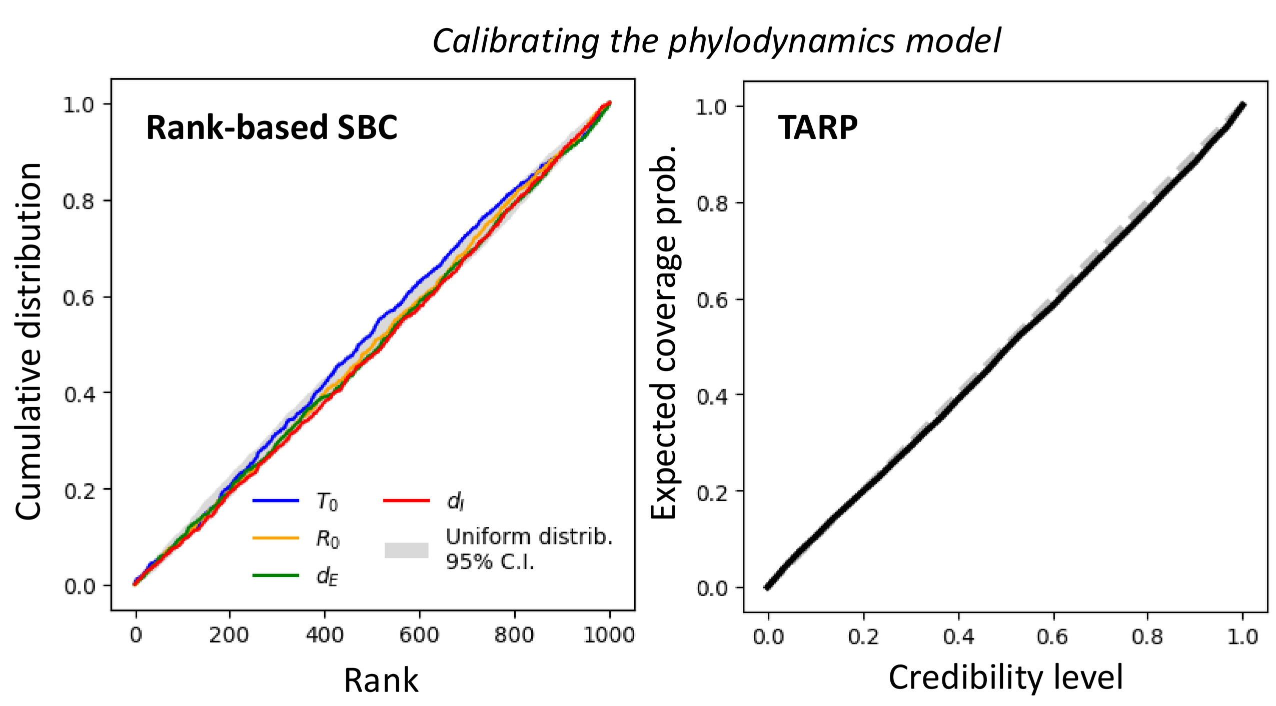
Amortized inference & posterior calibration

Amortization refers to the NPE estimator being instantly applicable to new data without further training.

Posterior calibration methods

- *Rank-based simulation-based calibration[j] (SBC)*: checks global accuracy of marginal distributions.
- *Test of Accuracy with Random Points[k] (TARP)*: checks global accuracy of entire posterior distribution.
- Local Classifier 2-Samples Test[l]: checks accuracy of entire posterior distribution on particular data.

Calibrating the phylodynamics model



References

[a]Cranmer et al, *PNAS*, 2020
[b]Sisson et al, *Chapman & Hall*, 2018
[c]Papamakarios et al, *NEURIPS*, 2016
[d]Lueckmann et al, *NEURIPS*, 2017
[e]Althaus, *PLoS Curr*, 2014
[f]Gire et al, *Science*, 2014
[g]Stadler et al, *PLoS Curr*, 2014
[h]Saulnier et al, *PLoS Comp Biol*, 2017
[i]Voznica et al, *Nat Comms*, 2022
[j]Talts et al, *ArXiv*, 2020
[k]Lemos et al, *ICML*, 2023
[l]Linhart et al, *NEURIPS*, 2023