HELMHOLTZ AI Artificial Intelligence Cooperation Unit

HELMHOLTZ LOKI Pandemics

# A Comparative Analysis of Classical and Approximate Bayesian Inference Techniques in Multiparameter Compartmental Epidemiological Models for COVID-19 Pandemic

**Alina Bazarova[1,2], Johann Fredrik Jadebeck[3,4], Katharina Nöh[3], Wolfgang Wiechert[4], Martin Joachim Kühn[5,6], Stefan Kesselheim[1,2]**

[1]Forschungszentrum Jülich, Jülich Supercomputing Centre, Jülich, Germany [2]Helmholtz AI, Germany
[3]Forschungszentrum Jülich, Institute of Bio- and Geosciences, Jülich, Germany
[4]RWTH Aachen University, Computational Systems Biology, Aachen, Germany
[5]German Aerospace Center, Institute of Software Technology, Department High-Performance Computing, Cologne, Germany
[6]University of Bonn, Life and Medical Sciences Institute, Germany

## Problem Setting

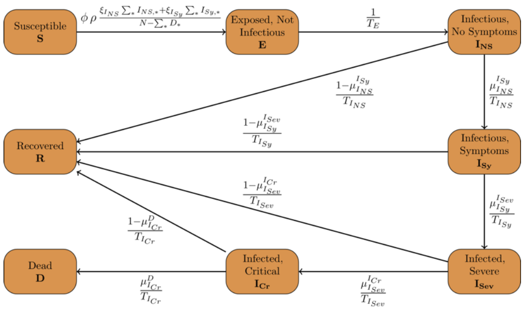
**Goal:** Bridging the gap between policy makers and science for future pandemics

- Requirement: Automatic Bayesian parameter estimation
- Research question: Can we use approximate Bayes to speed up estimation?

**Setting:**

- Compare Markov chain Monte Carlo & Simulation Based inference
- Same (mechanistic) epidemiological model
- Different parameters across counties
- data from COVID-19 pandemic

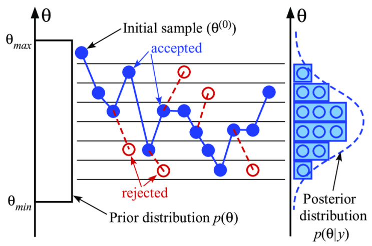
## Compartmental epidemiological model SECIR



Takes into account parameters such as: **contact patterns**, transmission probability **on contact**, probability of becoming **critical** if **severe**, probability of **dying** if **critical** etc.

20 parameters overall. Use Memilio [1] for simulation
Student-t likelihood for new-cases each day [2]
Lognormal priors

## MCMC Methodology



MCMC visualization from [3]

- Run 4 Markov chains using PyMC [4]
- DE-Metropolis-Z, because it does not require gradient
- Each Markov chain must converge to the posterior
- Check using rhat statistic < 1.1

## Simulation Based inference methodology

**Workflow:**
**Bayesian approach combined with AI**
simulations

Results and parameter values as input to AI model (**normalising flows**)

Approximate posterior distribution (NPE)

Sample from approxmate posterior given the observed data

Obtain **distributional parameter estimations**

**Important:** once the neural network is trained, **no need to re-train it for each German county!**

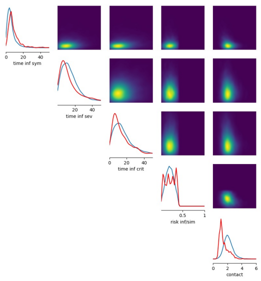
## Technical details of SBI

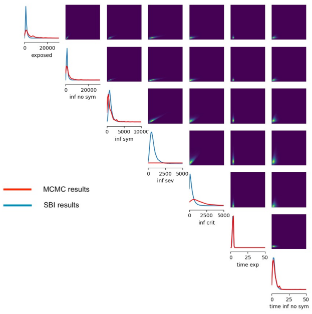
Simulations
CPU multiple nodes
**ray backend**

Normalizing flow training GPU

Implemented as patches to python sbi package [5,6]

## Results





**SBI**

- Running time < 5 min, 48 cores for up to $10^6$ simulations
- Requires reparametrization
- More narrow priors
- highly parallelizable

**MCMC**

- Running time 15 min on 4 cores (rhat < 1.1)
- Relaxed priors
- guaranteed to sample posterior
- limited parallelization

## Outlook: Inferring hotspots using SBI

**Regionally coupled graph models**

- 1 model per county (400) and cross-county travel of infected according to commuter patterns
- Compute time 1 forward simulation 16 seconds
- Minimize inference time on Supercomputer using SBI

## REFERENCES

[1] Koslow, W. et al. (2022). Appropriate relaxation of non-pharmaceutical interventions minimizes the risk of a resurgence in SARS-CoV-2 infections in spite of the Delta variant. PLoS Computational Biology, 18(5), e1010054.
[2] Dehning, J. et al. (2023). Impact of the Euro 2020 championship on the spread of COVID-19. Nature Communications, 14(1), 122. doi:10.1038/s41467-022-35512-x
[3] Tomic, S. et al. (2019). Distributed localization with complemented RSS and AOA measurements: Theory and methods. Applied Sciences (Basel, Switzerland), 10(1), 272. doi:10.3390/app10010272
[4] Abril-Pla, O. et al. (2023). PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ. Computer Science, 9, e1516. doi:10.7717/peerj-cs.1516
[5] Tejero-Cantero, A. et al. (2020). sbi: A toolkit for simulation-based inference. Journal of Open Source Software, 5(52), 2505. doi:10.21105/joss.02505
[6] Ray clusters Overview https://docs.ray.io/en/latest/cluster/getting-started.html

DLR

JÜLICH Forschungszentrum