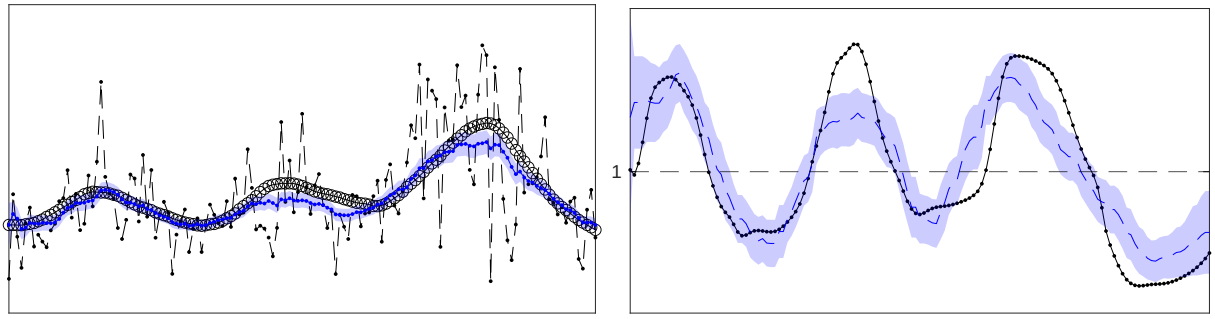
POSTDOC PROJECT IN COMPUTATIONAL SCIENCE: TRUSTWORTHY USE OF NEURAL NETWORKS IN BAYESIAN EPIDEMICS

STEFAN ENGBLOM



Left: noisy measurements of virus concentration in wastewater, *right:* reproduction number estimates with synthetic truth in black. Blue shade is the estimated uncertainty.

Effective summary

Public health agencies can be expected to be increasingly dependent on sophisticated sets of data in order to accurately plan for disease scenarios, to design mitigation- and suppression programs, and to assess risks. The quality of such strategies ultimately depend on the development of a range of tools to analyze this data.

In this project we are particularly interested in designing flexible data-driven computational models which are effective in producing not only forecasts, but also in accurate "nowcasts" in order to provide for situation awareness during decision making. Given the specifics of the application, there is a strong component relating to *trustworthiness* and *robustness*.

The proposed project consists of two parts as follows.

Neural Network Priors: The primary goal here is to construct a novel Neural Network (NN)-driven framework for generating Bayesian priors in epidemiological models; with more effective priors, we can substantially accelerate model calibration to incoming data. We will utilize *normalizing flows* to represent prior distributions, and the design will take place under a maximum entropy constraint with additional terms representing the stability of the identifiability problem.

Diagnostics for Simulation-Based Inference: In this part the focus is on developing a diagnostics framework to continuously assess the consistency of posterior models. The challenge lies in maintaining model reliability when handling diverse

data sources that introduce complex and evolving uncertainties. To address this, we will propose a diagnostics test that integrates parametric bootstrapping and Data Consistency Checks (DCC), comparing model simulations to observed data to detect discrepancies as they arise.

To sum up, the first part tries to automate prior generation using NNs to provide flexibility, while the overall Bayesian context provides for transparency. The idea with the second part is to compare observations with model-generated data in real time, allowing for an early detection of inconsistencies and a need for model adjustment.

Perspectives

Mathematical modeling of the spread of disease is a fairly mature field. However, the increasing volumes and the improved granularity of data puts different perspectives on modeling in epidemics. Computational models tend to become more complex, more detailed, and containing more information. But without sufficient data quality to drive the simulations, the accuracy may well be imaginary.

In this project we are interested to partially counter these issues by designing epidemiological models driven by data. Given the constraints of most Public Health related applications, black-box solutions are not acceptable. For any estimates obtained using computational models, there is a substantial need for independent statistical validation. Modern streams of epidemiological data offer greatly improved model resolution but also call for a greater modeling flexibility. In this project we aim to improve on this situation via flexible, yet transparent data-driven epidemiological models, and via enhanced control relying on data consistency checks.

Expectations

The work consists of research of the described situation for specific set-ups and examples, using real as well as synthetic data sets. Implementation and evaluation of the proposed methods as well as writing research papers in collaboration is also expected.

A suitable background includes one or more of

- Computational Science or Data Science
- Compute-intensive Statistics or Applied Mathematics

Proficiency in a relevant programming language, e.g., Matlab, Python, R, C/C++ is expected and practical or research experience with machine learning techniques in general and neural networks in particular will be highly beneficial. A PhD in a relevant field is required.

Application deadline February 28th 2025.
Welcome to contact me for informal enquiries and discussions.

(S. Engblom) Division of Scientific Computing, Department of Information Technology, Uppsala University, Sweden.

Email address: stefane@it.uu.se, https://stefanengblom.github.io