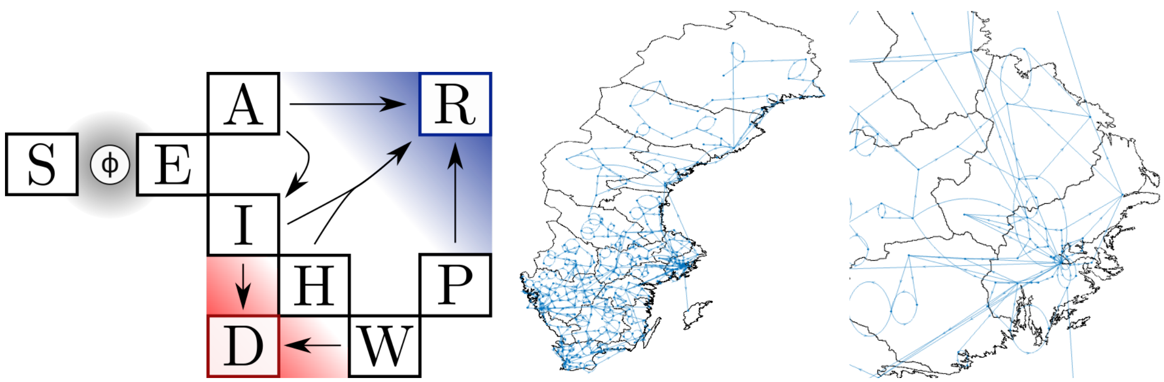
COMPUTATIONAL EPIDEMICS DRIVEN BY DATA

STEFAN ENGBLOM



Effective summary

Public health agencies, perhaps now more than ever, 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. In this project we are particularly interested in the following aspects.

A. Data-driven modeling: focused around a simulation software www.siminf.org and novel model formulations. The interesting aspects here include, amongst others, how do we design a computational model of predictive value which strikes a good trade-off with the available data and with modeling from first principles?

B. Bayesian inference: development of simulation-based posterior exploration algorithms, to support rigorous data-driven modeling within the context. In particular: how can computational tools in epidemics provide sound estimates of uncertainties?

Background

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 such issues by explicitly designing epidemiological models driven by data. The work will be centered around a software SimInf, www.siminf.org, and we will be developing novel ways of expressing the dynamics of disease spread such that data may be more efficiently incorporated. Issues with parameterization will be one focus point, and clear trade-offs in simulation uncertainty will be another.

This project

Data-driven modeling: With data of different qualities, being able to put a confidence value to different data streams becomes increasingly important. We are therefore interested in developing computational models which can effectively provide residuals relative to data-sets, and such that the fit/misfit can be judged. In an environment where risk is an important aspect, this is often measured in terms of probabilities and likelihoods.

Bayesian inference: A consistent way to probabilistically incorporate data is via Bayesian modeling. This formally allows pieces of information to be meaningfully incorporated as they are available. However, for this flexibility to also be realized, the computational complexity of the model can generally not be arbitrary, and the data-to-model flow must be designed with care. Ideally, Bayesian models output probabilistic answers which aligns very well with decision processes.

Candidates with a background in one or more of Scientific Computing, Automatic Control, Applied Mathematics, Data Science are more than welcome to contact me for further information.

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

Email address: stefane@it.uu.se, http://user.it.uu.se/~stefane