Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

STOCHASTIC MODELING AND BAYESIAN INFERENCE OF NATIONAL SCALE EPIDEMICS IN THE SWEDISH CATTLE NETWORK

Robin Eriksson
Stefan Engblom

Division of Scientific Computing, Department of Information Technology
Uppsala University, Lägerhyddsvägen 2
Uppsala, SE-752 37, SWEDEN

ABSTRACT

In this work, we study the spread of a *verotoxigenic E. coli* in the Swedish cattle population and parameterize a disease-spread model by combining the high-performance simulator SimInf with actual agent transport and bacterial testing data. We perform Bayesian inference by using Approximate Bayesian Computations (ABC) and Synthetic Likelihood Markov chain Monte Carlo (SLMCMC), and we obtain posterior parameter densities with desirable averages.

1 INTRODUCTION

Numerous infectious diseases have a zoonotic origin, i.e., infected animals transfer the infection to humans. An example of a zoonotic pathogen is the *verotoxigenic E. coli* O157:H7 (VTEC), which is present in the Swedish cattle population. The primary transmission route for the pathogen between herds is the movement of livestock, a standard action amongst the network participants. Transport data of bovine animals are now widely available, and it is, therefore, possible to develop realistic large-scale disease spread models that incorporate the transport network to better understand the transmission of zoonotic infections in the foodborne population. We implement our computational model with detailed information about the underlying network in SimInf (Widgren et al. 2016), a high-performance R package built with C compiled code, designed for stochastic disease spread simulations on networks. With the simulator and actual spatiotemporal bacterial sampling data as observations, we explore the posterior parameter density of the Swedish VTEC epidemic. We provide multiple samples from the model parameter posterior, thus showing the feasibility of Bayesian inference on national scale epidemics.

2 EPIDEMIOLOGICAL MODELING

The full dataset we utilize contains a total of 18,649,921 detailed transport records; it holds information about births, deaths, and transfers between nodes in the period 2005 to 2013 (Nöremark et al. 2011). The Swedish cattle network we study includes a total of 37,221 unique identifiers, each corresponding to a geographical area where one keeps animals, e.g., a farm building or pasture, and they are all distributed across the entire country. In our system simulation, agents move between nodes on a connected graph and enter or exit the system according to scheduled events. In each node, we apply the compartment-based model named SISE to determine the state of the housed individuals. The model consists of two state compartments: susceptible (S) and infected (I), and one environmental compartment (E). The environmental compartment E represents the infectious pressure from free-living pathogens in the environment. The infection is indirect between individuals, i.e., the infected individuals shed bacteria into the environment, from which susceptible individuals contract it. On the model implementation, there are three combined applications to the hybrid simulation. First, the state transitions are stochastic, and we realize them as continuous-time discrete state Markov chains. Second, we implement the environmental

compartment as continuous time ordinary differential equation, and we time-evolve them using a forward stepping solver. Third, we manage the connection between nodes, the data-driven transports of agents, in a discrete-event module. Finally, we integrate all parts with a consistent mathematical framework in SimInf (Bauer 2016).

3 POSTERIOR PARAMETER EXPLORATION

With the epidemiological model and actual observed data in mind, we consider a presupposed truth as a time-dependent stochastic process $X(t) = X(t,\theta)$, parameterized by some parameter θ. Assuming a set of observations $(x_i) = (x_i, t_i) \sim X(t_i)$, and the task to estimate the unknown parameter θ, conventionally, one would like to perform a Bayesian posterior parameter exploration. However, in our present case, the likelihood function is intractable, making Bayesian inference challenging. Thus, we consider two likelihood-free methods: ABC (Beaumont et al. 2002) and SLMCMC (Wood 2010). Both methods compare model generated data with observations as a surrogate for the likelihood. To evaluate and build confidence in our approach, we carry out an "inverse crime"-scheme. We start with an initial system that when compared to the target is simple and has a known ground truth. After a successful inference attempt, we increase the system complexity, e.g., network size and the number of parameters. This procedure continues until we reach the full target setup. In one instance, we carry out the inference with observational data and obtain the result seen in Figure 1. From it, we conclude that the SLMCMC method yields a tighter, more desirable, posterior adjacent to the actual parameter value when compared to ABC and consequently is the suggested method for any further inference attempts. The outcome of this work in full confirms that Bayesian inference is achievable on national scale epidemics, and will allow the evaluatory groups to perform more realistic and informative epidemic reviews than compared to what is possible using single point-values.

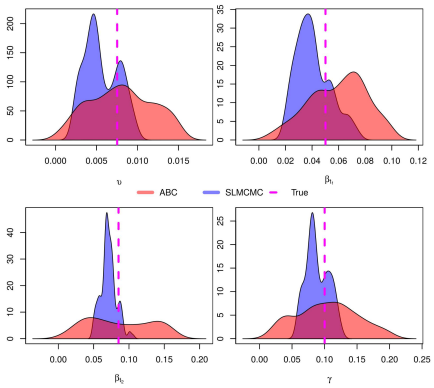

Figure 1: Parameter posterior density achieved using ABC and SLMCMC on a subset of the full system.

REFERENCES

Bauer, P., S. Engblom, and S. Widgren. 2016. “Fast Event-Based Epidemiological Simulations on National Scales”. *International J. of High Performance Computing Applications* 30(4):438–453.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. “Approximate Bayesian Computation in Population Genetics”. *Genetics* 162(4):2025–2035.

Nöremark, M., N. Håkansson, S. S. Lewerin, A. Lindberg, and A. Jonsson. 2011. “Network Analysis of Cattle and Pig Movements in Sweden: Measures Relevant for Disease Control and Risk Based Surveillance”. *Preventive Veterinary Medicine* 99(2):78–90.

Widgren, S., P. Bauer, R. Eriksson, and S. Engblom. 2016. “SimInf: An R Package for Data-Driven Stochastic Disease Spread Simulations”. *arXiv preprint arXiv:1605.01421v3*.

Wood, S. N. 2010. “Statistical Inference for Noisy Nonlinear Ecological Dynamic Systems”. *Nature* 466(7310):1102–1104.