Spatial and Spatio-temporal Epidemiology 6 (2013) 59–70

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Spatial approximations of network-based individual level infectious disease models

Nadia Bifolchi *, Rob Deardon, Zeny Feng

Department of Mathematics & Statistics, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada

ARTICLE INFO

Article history:
Received 5 January 2012
Revised 13 May 2013
Accepted 12 July 2013
Available online 22 July 2013

Keywords:
Individual-level models
Markov chain Monte Carlo
Epidemic modeling
Contact network
Spatial approximation

ABSTRACT

Often, when modeling infectious disease spread, the complex network through which the disease propagates is approximated by simplified spatial information. Here, we simulate epidemic spread through various contact networks and fit spatial-based models in a Bayesian framework using Markov chain Monte Carlo methods. These spatial models are individual-level models which account for the spatio-temporal dynamics of infectious disease. The focus here is on choosing a spatial model which best predicts the true probabilities of infection, as well as determining under which conditions such spatial models fail. Spatial models tend to predict infection probability reasonably well when disease spread is propagated through contact networks in which contacts are only within a certain distance of each other. If contacts exist over long distances, the spatial models tend to perform worse when compared to the network model.

1. Introduction

Having the ability to produce accurate mathematical models of infectious disease spread can help provide researchers and government officials with the knowledge needed for making policy decisions directed toward containment of disease spread. Quick and accurate disease models may answer critical questions that can potentially save lives and protect economies. For example, severe acute respiratory syndrome (SARS) in 2003 had a drastic affect on tourism, food and travel, costing China 8.5 and Canada 4.3 billion US dollars (Beutels et al., 2009). Another example is given by Meltzer et al. (1999) who estimated the economic impact of a future influenza pandemic in the United States at 71.3 to 166.5 billion US dollars.

Generally, infectious diseases propagate via complex individual-level interactions, or contacts, between infected and susceptible individuals in the population. Combining all individual contact information into a contact network enables researchers to analyze disease spread through the population. There is a substantial amount of literature on network based epidemiology in diseases such as foot-and-mouth disease and avian influenza, e.g. (Cauchemez et al., 2011; Dubé, 2009; Jewell et al., 2009; Marchbanks et al., 2011; Streftaris and Gibson, 2004; Zhen et al., 2011).

However, network data that we may wish to use to model the spread of various diseases is often difficult to obtain. Collection of such data is expensive and there are issues regarding recall and privacy encroachment. A connection, or a contact between two individuals, is usually deemed to be any contact between individuals by which the disease can spread from an infected individual to a susceptible individual. The connections themselves can be hard to describe, as researchers must quantify the type of relationship or contact needed for infection to transfer (Keeling and Eames, 2005). The networks may be social in nature, spatial proximity based, or demographic (Kolaczyk et al., 2009). For example, a network may be defined

Abbreviations: ILM, individual-level model; MCMC, Markov chain Monte Carlo; SARS, severe acute respiratory syndrome; SIR, susceptible-infected-removed.

* Corresponding author.
E-mail addresses: nbifolch@uoguelph.ca (N. Bifolchi), rdeardon@uoguelph.ca (R. Deardon), zfeng@uoguelph.ca (Z. Feng).

1877-5845/$ - see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.sste.2013.07.001

by sexual activity between two individuals in the case of a sexually transmitted disease. Alternatively, if trying to model the spread of the Norwalk virus in people, we might use knowledge about which individuals live together in the same house, attend the same schools, or work together, etc. If modeling a livestock disease at the level of individual farms, say, we may want data on the trade networks, supply networks, and even social networks of farmers and farm workers.

Due to the complexity of, and difficulties in obtaining accurate information about, such networks, simplifications are sometimes made in the model being used. For example, we may use a spatial network, rather than a more desirable trade network. Examples of such spatial simplification can be found in a number of models of the UK 2001 foot-and-mouth disease epidemic (Chis Ster and Ferguson, 2007; Chis Ster et al., 2009; Deardon et al., 2010; Keeling et al., 2001). Plant epidemiologists often make such simplifications as their subjects are generally stationary thus allowing infective pressures to decrease exponentially with distance (exponential decay), with a power of distance (geometric decay) or with a nearest neighbor effect (Beutels et al., 2009; Filipe and Maule, 2004). A piecewise function, similar in concept to the nearest neighbor effect only with a given probability of infection from a long distance source, has also been used in modeling wildlife infectious diseases in which a physical barrier (such as a river) reduces mixing within the population (Smith et al., 2005). The simplest assumption to make for any infectious disease model is to assume homogeneous mixing within the population thereby, with no other covariate information, assuming equal infective pressure on all individuals within the population. Bansal et al., 2007 provides insight into the ability of homogeneous-mixing compartmental model's ability to predict the characteristics of network-based epidemics.

Deardon et al., 2010 define a class of individual-level models (ILMs) that can be used to model the spread of disease when its spread depends on various individual-level risk factors. Spatio-temporal aspects of the infectious disease can easily be incorporated into such ILMs, enabling researchers to incorporate spatial proximity to infectious individuals in the model. Similarly, network information can be included in such models. The statistical process of fitting the model to observed data is one key aspect of analyzing epidemic data. ILMs, and similar models, can be fit to data within a Bayesian statistical framework using Markov chain Monte Carlo (MCMC).

The purpose of this paper is to examine the effect of using spatial information as a proxy to more complex network information when fitting ILMs to epidemic data. Our intention with this paper is to present generic insights into the cost of using a spatial model when the underlying population is connected by a spatially-based network. This is carried out via two simulation studies. These studies involve simulating epidemics, propagated through networks of varying complexity, and comparing the results obtained when both network-based, and spatial-based, ILMs are fit to the simulated data.

The paper is laid out as follows. The general ILM framework and specific ILMs used in the paper will be outlined in Section 2. Epidemic study and model assessment criteria will also be discussed. Section 3 presents the results of the simulation studies via the use of our chosen model assessment criteria. Conclusions that are made from the results as well as a list of possible future work will be given in Section 4.

2. Methodology

2.1. General model framework

The general framework of individual-level models (ILMs) for infectious disease is presented in Deardon et al. (2010). Here, we briefly review this framework in the context of a susceptible-infectious-removed (SIR) compartmental class of models.

In a discrete time SIR model each individual i can be in one of three states at any time point: $i \in S$ implies that the individual is susceptible to the disease; $i \in I$ implies that the individual is infected and is infectious; $i \in R$ implies that the individual is removed from the population and no longer able to be infected or infect other individuals (e.g. by recovering and gaining immunity to the disease or dying). An individual i in one of these states at time t is denoted to be in the set $S(t), I(t)$, or $R(t)$, respectively. The epidemic history comprises $S(t), I(t), R(t)$ for $t = 1, \ldots, t_{max}$ where, t_{max} is the time at which the last infectious individual enters the removed state. Individuals within the epidemic may only move from $S \rightarrow I$ and $I \rightarrow R$. Individuals are defined as discrete points in space and time with the probability of a susceptible individual i becoming infected with the disease at time t equal to

$$P_{it} = 1 - \exp\left[\{-\xi(i)\sum_{j \in I(t)} \rho(j)\kappa(i,j)\} - \varepsilon(i,t)\right], \tag{1}$$

where $\xi(i)$ is a function representing potential risk factors associated with susceptible individual i contracting the disease; $\rho(j)$ is a function representing potential risk factors associated with infectious individual j transmitting the disease; $\kappa(i,j)$ is an infection kernel representing potential risk factors involving both infected and susceptible individuals j and i, respectively; $\varepsilon(i,t)$ is a function that accounts for some random behavior within the epidemic that cannot be explained by the other terms in the model (e.g. infection of a susceptible individual by an infectious individual from outside the observed population). For the purpose of this paper $\varepsilon(i,t)$ is set to zero.

We define the epidemic history as $\{S(t), I(t), R(t)\}_{t=0}^{t_{max}}$. Given the complete epidemic history the likelihood can be computed as:

$$l(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{t=1}^{t_{max}} \left[\prod_{i \in I(t+1) \setminus I(t)} P_{it}\right] \left[\prod_{i \in S(t+1)} 1 - P_{it}\right], \tag{2}$$

where, $\mathbf{y}$ is the observed epidemic data; $\boldsymbol{\theta}$ is a vector of parameters;

$I(t+1) \setminus I(t)$ is the set of newly infected individuals at time $t+1$; and $S(t+1)$ is the set of susceptible individuals at time $t+1$.