CHAPTER 2

Modeling the Spread of Infectious Diseases: A Review

Dongmei Chen
Department of Geography, Faculty of Arts and Science, Queen's University, Kingston, ON, Canada

2.1 INTRODUCTION

Monitoring, analyzing, and predicting the impact of infectious diseases on the well-being of a society is the cornerstone of identifying effective ways to prevent, control, and manage disease spreads. It is a common perception that every infectious disease is transmitted through space and time from one individual to another in its own special spreading network in the environment. The use of models in public health decision-making has become increasingly important in the study of the spread of disease, designing interventions to control and prevent further outbreaks, and limiting their devastating effects on a population (McKenzie 2004; Day et al. 2006; Moghadas et al. 2008; Grassly and Fraser 2008; Arino et al. 2011).

Modern epidemiological analysis and modeling theory began in the late nineteenth and early twentieth centuries. By plotting cholera epidemic cases on a map, Snow (1849) hypothesized that contaminated water was the predominant contributor to the cholera transmission in London in 1849. Arthur Ransome, who first described the cyclic behavior of measles, developed a discrete-time epidemic model for cholera transmission in 1906 (Roberts and Heesterbeek 2003). This early spatial and temporal epidemic research, combined with the progress of contemporary biology studies, led to some important discoveries regarding disease transmission. Aside from efforts made to develop vaccinations and medicines, it is expected that human infectious disease research will relieve the threat of infection by revealing the temporal and spatial dynamics of spreads. To meet this expectation, a variety of analysis and

Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases, First Edition.
Edited by Dongmei Chen, Bernard Moulin, and Jianhong Wu.

modeling techniques have been developed with the assumption that there exists a fundamental spatial structure of disease spreading based on which the human and physical geographical world is formed (Lawson 2005; Riley 2007; Keeling and Ross 2008; Waller 2007; Auchincloss et al. 2012).

Many of the early disease models were devoted to mathematical modeling on a population level, assuming various kinds of homogeneity. The classic method of mathematical modeling considered a host population to be divided into distinct units, and each individual interacted with other individuals in his or her immediate neighbourhood. The simplest of these population-based models is the SIR model, initially described by Kermack and McKendrick (1927) for a closed population. However, possible spatial–temporal spread and effects of a disease outbreak in different communities usually play more important roles in determining public health interventions (Auchincloss et al. 2012). Traditional mathematical models that represent the dynamics of infectious diseases use a nonspatial and population-based lens to view disease spread and assume homogeneity in disease transmission. While useful in estimating the size of the affected population, these models do not explicitly address the causal factors in the development of epidemics.

The development of computer technology and increasing availability of disease-related spatial data have made different modeling approaches possible as they have the power to support modeling of large numbers of objects easily and examine disease spread through time and space (Moore and Carpenter 1999; Riley 2007; Yang et al. 2007; Bian and Liebner 2007; Grassley and Fraser 2008). Technological advances and the desire to design realistic models have led to the emergence of more advanced mathematical, individual-based statistical and simulation models. The explicit consideration of the causal factors in disease transmission, such as the behavior of individuals, individualized interactions, and the patterns of interactions, signifies that individual-based models have the flexibility to model the observed heterogeneity in disease transmission and are better able to provide insight into population health. These models also have the advantage of integrating data on the location of hosts and their typical movement patterns with a quantitative description of the infection process and the disease's natural history in order to investigate observed patterns and evaluate alternative intervention options.

Au: The citations "Yang 2007," "Kaplan 2012," "Balan et al. 2009," "Eubank et al. 2005," "Ajeli et al. 2010," and "Admatzky 2010" have been edited as "Yang et al. 2007," "Kaplan 2011," "Balcan et al. 2009," "Eubank 2005," "Ajelli et al. 2010," and "Adamatzky 2010," respectively, and we hope these changes are OK.

To date, various models have been developed and applied to modeling the spatial dynamics of infectious diseases, including mathematical models, statistical models, and spatial simulation models based on population levels and individual levels. This chapter gives a brief overview of the various models and their key considerations. The focus of this review is to obtain an insight into their basic technical composition. The advantages and limitations of these models, as well as issues existing on disease data are also discussed.

2.2 MATHEMATICAL MODELLING

Mathematical modeling uses mathematical concepts and language to describe the process of disease spread and propagation. Mathematical models have been widely

used to quantitatively represent and predict the dynamics of disease infection on the population level. Mathematical models have evolved from extremely simple models, such as SIR models, to complicated compartmental mathematical models (Diekmann and Heesterbeek 2000; Lloyd and Valeika 2007; Brauer 2008; Hejblum et al. 2011).

2.2.1 Classical Mathematical Models

The classic method of mathematical modeling considers a host population to be divided into distinct units. The simplest form of these population-based mathematical models is the SIR model, which was initially described by Kermack and McKendrick in 1927 for a closed population. Originally, this model was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665–1666, Bombay 1906) and cholera (London 1865).

This type of model is based on our intuitive understanding of how epidemics of simple communicable diseases occur in the real world, and comprises three categories of individuals: those who are susceptible to disease (S), those who are infectious and can spread the disease to susceptibles (I), and those who have recovered from previous infection and can no longer spread or catch the disease (R). Differential equations are used to illustrate the dynamics of each of these subpopulations through the course of an epidemic (Anderson and May 1991). Disease transmission occurs by the stochastic infection of a susceptible by a neighboring infective, and spread takes place when infected individuals mix among susceptibles. Thus, at any given time, some individuals from the susceptible segment become infected, while some of those who are infected join the recovered segment. It is assumed that these changes are continuous and can be described by the following differential equations:

$$\frac{dS}{dt} = \beta \frac{SI}{N}, \tag{2.1}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - gI \tag{2.2}$$

$$\frac{dR}{dt} = gI \tag{2.3}$$

where S, I, and R denote the susceptible, infectious, and recovered individuals, respectively; N is the total population (S+I+R); β is the infection coefficient, the product of the average number of contacts (C) within a given time period and the probability of infection (p) for a contact between susceptible and infectious individuals; and g is the recovery rate. In infectious disease epidemics, the severity and the initial rate of the increase have a positive correlation with the value of the Basic Reproduction Number (R_0). When the SIR model is applied, R_0 can be calculated using the transmissibility β multiplied by the duration of the infectious period (Coburn et al. 2009). For a given epidemic, the number of individuals within the infectious segment rises after an epidemic begins and falls after the epidemic peaks, and may be quantified using

the SIR model by adjusting β and g (Kermack and McKendrick 1927; Anderson and May 1991; Diekman and Heesterbeek 2000).

A SIR model specified by Equations 2.1–2.3 is a base compartmental model. When SIR is incorporated with an exposed (or latent) compartment (explicitly containing those infected but not yet infectious), the model is a SEIR model; and in the cases in which susceptibility returns after recovery, the model is called an SIS model (Hethcote 2009). More complex mathematical models have been developed from these initial models by taking additional variables into account, such as births, deaths, and migration into and out of the population, or by monitoring the spread of multiple epidemics simultaneously (Anderson and May 1991; Brauerner 2008).

SIR models are commonly seen in their deterministic versions. However, it is apparent that most epidemic development is stochastic (Tuckwell et al. 2007). Bailey (1975) gives a thorough discussion of the properties of deterministic and stochastic versions of the SIR model. While additional parameters have been added to the general framework of SIR models throughout the history of modern epidemiology, the basic framework has remained intact.

Despite the strengths of this modeling approach, there are limitations that have received increasing attention. Koopman and Lynch (1999), for example, have criticized the SIR model for failing to produce realistic and useful results, particularly for complex disease systems. One significant simplification is that populations are viewed as continuous entities, and individuals are not considered. The SIR model also imposes further simplifications with respect to contact patterns, as it is not designed to capture details of individual connection patterns and networks, and contact is assumed to be an instantaneous event (Koopman and Lynch 1999). In addition, the assumption of population-wide homogeneous parameters limits the ability of these models to assess and characterize (a) how diseases spread, and (b) whether the decline of an epidemic is primarily a result of intervention control measures or heterogeneity in infection (Dye and Gay 2003). Several recent reports have indicated that it is the heterogeneity in transmission that usually leads to the sustaining or decline of an epidemic, rather than intervention measures (Arita et al. 2003; Galvani 2004; Meyers et al. 2003; Dye and Gay 2003).

2.2.2 Spatiotemporal Mathematical Disease Modeling

The spatial version of the aforementioned mathematical models assumes that the spread of disease is a spatial process (Ferguson et al. 2001; Rhodes and Anderson 1997; Riley 2007; Gilberto et al. 2011). One approach to representing the geographic distribution of hosts or vectors and their movement in space is the use of spatiotemporal compartmental models, which consider the border edges of units where population is located. A simple nearest-neighbour mixing model adjusted from Equation 2.1 can be used to define the rate at which infectious people within an area (or patch) j cause susceptible people at the same area j to become infectious:

$$\frac{dS_j}{dt} = \beta \frac{S_j I_j}{N_j} \tag{2.4}$$

The rate of infection between the population in the area j is

$$S_j \frac{\beta}{\sum_{i-1}^{K} M_{ji} N_i} \sum_{i=1}^{K} M_{ji} I_i \tag{2.5}$$

where M_{ji} is the mixing rate between j and its neighbouring area i (note that $m_{jj} = 1$); I_i is the number of infectious individuals in i; K is the total number of areas (patches); N_i is the total population in i; β is the transmission coefficient.

The spatial component that connects different areas (patches) is represented by the set of M_{ji} in the spatiotemporal compartmental models. M_{ji} can be constructed based on the common border edges, transportation edges (see Chapter 5), mixing edges, or migration edges (see Chapter 7 for examples).

Another spatial approach treats the dispersion of disease as travelling waves of infection across a landscape. In this model, disease spreading or propagation in space is relevant to the so-called traveling waves; a disease invades when the susceptible receives infection by contacting the infected at the travel front and leaving behind the recovered and immune (Zhang 2009; Wang et al. 2012). The utility of these models has been primarily in predicting the spread of human infectious disease between human communities (such as from one city to another, and from city to rural region), and of vector-borne diseases across natural landscapes. Having been developed from the population-based modeling approach, which views a community as a homogeneous entity with transmission of disease occurring between these large units, the traveling-wave model treats a smaller portion of the population as a homogeneous unit, for example, a neighborhood. While this model allows for greater heterogeneity in space, it still remains highly connected to the traditional models and thus remains inappropriate for individual-based models.

The wave-front model has been used successfully in several diverse applications. For example, Murray et al. (1986) developed a model of a rabies epidemic in the event of a United Kingdom outbreak, which demonstrated the propagation of an epizootic front through the susceptible fox population. This model was built on the observation that the spread of fox rabies westward from Poland after World War II showed a front-like progression in which propagation was faster in regions of higher fox density (Kallen et al. 1985; Anderson et al. 1981). This deterministic model was used to predict the wave speed of the disease and estimate the width of an intervention zone necessary to prevent spread. The rate of disease spread was of primary interest, while the determinants and possible deterrents of disease spread were secondary (Moore and Carpenter 1999). A second early example is the model of Noble (1974) for the spread of plague in human populations. This model incorporated epidemiologically derived parameters to track the rate of progress of the disease, and results were consistent with the historical record of the epidemic's progress after its introduction into southern Europe in 1347. More examples involve applications of those models to vector-borne diseases, including West Nile virus, malaria, measles, avian flu, dengue fever, and Lyme disease (Okubo 1998;

Gourley 2000; Grenfell et al. 2001; Ruan and Xiao 2004; Lewis et al. 2006; Gourley et al. 2007; Zhang 2009; Wang and Wu 2010). In these examples, structured population models and their associated differential equations were used to describe the interaction of different subpopulations comprising susceptible, infected, recovered vectors.

Other spatial–temporal mathematical models involve using complicated differential equations with time delay or lags to capture the rich variety of dynamics observed in disease transmission (such as Cooke et al. 1999; Culshaw and Ruan 2000; Nelson and Perelson 2002; Forde 2005). A complete review of different mathematical models can be found in Grassly and Fraser (2008), Brauer (2008), Diekmann et al. (2010), and Hejblum et al. (2011).

2.3 STATISTICAL MODELING

Statistical modeling can be loosely defined as "fitting equations to data" (Scott 2010). In disease analysis, statistical modeling usually formalizes relationships among variables that may influence the spread of disease, describes how one or more variables are related to each other, and tests whether some statements or assumptions we had on disease spreading process are true. Statistical modeling is often done through either exploratory data analysis (EDA) in order to obtain the main characteristics of disease data, or confirmatory data analysis (CDA) in order to test a statistical hypothesis. Many traditional statistical analysis and modeling approaches are provided in statistical textbooks (such as Selvin 2004; Kaplan 2011; Freeman 2009). In the following the focus is put on spatial aspects of statistical modeling for infectious diseases.

Spatial statistical models involve the statistical analysis and modeling of disease observations with their locations and their potential impacting factors in space and time domains. Often these observations do not follow a Gaussian distribution and are not independent of the development of statistical methods (Waller 2007). Spatial statistics may surmount mathematical models in depicting regional risk factors, and are able to take into account both spatial and temporal residual variations in the analysis. Thus, the actual practice of statistics has moved beyond the conventional statistical methods by incorporating spatial effects, for example, autocorrelation (spatial dependency or clustering) or dealing with spatial data problems (Lawson 2005).

Spatial statistical analysis and modeling are often used to test three broad classes of hypotheses: disease mapping, disease clustering, and ecological analysis (Lawson 2005). Disease mapping concerns the use of models to describe the overall disease distribution on the map. Cressie (1993) broadly divided spatial data into three categories: (1) geostatistical data that are primarily parameterized with continuous values and chosen locations; (2) aggregated lattice data based on either regular or irregular lattice; and (3) point process data containing observations with responses for the random spatial process. Aggregated lattice count data is produced by regionalized point processes and primarily treated as the basic disease data formation. Various mapping techniques have been developed and applied for each category. For example, spatial smooth mapping techniques have been used to clean the noise, reduce the abrupt risk variation in point process disease data, and estimate disease risk by computing the

value at a location as the average of its nearby locations (Lawson 2005). Various spatial interpolation methods can be used to construct risk trend surface based on sample disease data collected from limited locations for continuous data and/or aggregated lattice count data (Berke 2004; Zhong et al. 2005).

The second class, disease clustering, is used to reveal or detect unusual concentrations or nonrandomness of disease events in space and time (Wakefield et al. 2000; Wang 2006). Disease clustering can be tested globally or locally. For global analysis, disease clustering assesses whether there is a general clustering in the disease dataset. It is often tested as a form of global spatial autocorrelation. In contrast, local clustering analysis is aimed at detecting the locations of clusters on a map. Many statistical measures and modeling methods have been widely developed and used to test and model global and local clustering analysis of disease outbreaks (see Lawson 2005 for details).

The last class, ecological analysis, is used to analyze the relationship between the spatial distribution of disease incidence and measured risk factors that can potentially impact the disease occurrence on an aggregated spatial level. Spatial regression analysis is often used in this type of analysis by counting spatial autocorrelations existing in variables and residuals through a spatial weighting matrix. There are several types of spatial regression models. Consider the classical regression model in Equation 2.6:

$$Y = X\beta + \varepsilon \tag{2.6}$$

where Y is the disease incidence risk vector observed for regions, X is the matrix of a set of explanatory risk factors (variables), and ε is random error vector with a typical Gaussian distribution. In a classical ordinary least squares (OLS) regression, ε should be independent. However, when spatial autocorrelation exists in disease data, the residual error plot will no longer be independent; instead, it may display a clustering effect. In order to count spatial autocorrelation or clustering effect, a *spatial error model* uses a spatial contiguity matrix to incorporate the spatial configuration in a regression model as Equation 2.7:

$$\begin{aligned} Y &= X\beta + \mu \\ \mu &= pW + \varepsilon \end{aligned} \tag{2.7}$$

where W is the contiguity matrix, and β and p are parameters to be estimated in the model. The parameter matrix p indicates the extent to which the variation of Y can be explained by the neighboring values.

Another spatial regression model, called *spatial autoregressive model* (SAR), uses the spatial weight matrix as shown in Equation 2.8:

$$Y = pWY + X\beta + \varepsilon \tag{2.8}$$

This model is similar to the lagged dependent variable model for time series regressions.

More complicated models can be constructed by mixing Equations 2.7 and 2.8 to build mixed SAR models (LeSage 1997; Beale et al. 2010). For example, generalized linear mixed models (GLMM) and spatial hierarchy models can incorporate both individual level and district level spatial effects (Breslow and Clayton 1993; Banerjee et al. 2004). Geographically weighted regression or conditional autoregressive (CAR) can also be used to build models in order to analyze the relations of disease risk factors and disease incidence risk (Lin and Wen 2011). Spatial regression differs from disease mapping in that the aim is to estimate the association between risk and covariates, rather than to provide area-specific relative risk estimates (Wakefield 2007).However, spatial regression analysis used for ecological analysis is based on data measured at aggregated units rather than individual subjects (Lawson 2005). This leads the ecological analysis and models to suffer from the issue of ecological fallacy in the interpretation of statistical results (Morgenstern 1982; Schwartz 1994).

Individual-level statistical models (ILMs) have been developed to model the transmission between disease states on the level of the individual instead of population groups at aggregated units (Deardon et al. 2010). The general form of ILMs considers the transition from a susceptible state to an infected state based on a time-dependent Poisson process for an individual, and then adds spatial structure of potential risk factors to explain the underlying disease dynamics. For more details and examples of ILMs, please see Chapter 11 of this book.

In the past decade, there have been enormous advances in the use of Bayesian statistical methodology for analyzing, mapping, and modeling epidemiologic data (Bernardinelli et al. 1995; Dunson, 2001; Xia et al. 2004; Browne and Draper 2006; Alkema et al. 2007; Forrester et al. 2007; McV Messam et al. 2008; Lawson 2008; Jandarov et al. 2013). The key principle of Bayesian statistical models is that the uncertainty about the parameters for the model is expressed through probability statements and distributions. The fundamental assumption of Bayesian methodology is that the values of parameters can be derived from distributions, which leads naturally to the use of models in which parameters arise within hierarchies (Lawson 2008). Bayesian methods benefit from incorporating space and time dependency in the modeling (Robertson et al. 2010). The Bayesian approach has great value in regional geostatistical and point process data analysis, especially when it is applied with hierarchical structures (Waller 2005; Lawson 2008; Lavine 1999).

Generally, statistical models are flexible regarding the format of the input data and the parameters selection, which makes them an ideal tool for observing spatial and temporal influence in preliminary exploration, especially when the dataset is small. However, spatial statistical modeling is challenged by its demand on data. As mentioned earlier, statistical modeling is based on data. It often requires a great effort to preprocess the data in large quantity and variety. Missing data, underreporting, uncertainty, and zero-counted data often occur in disease data. Developing methods to deal with disease data problems is a challenge in spatial statistical modeling. For advanced methods like Bayesian statistical modeling, computation complexity and difficulty are often the first obstacle to the application of Bayesian statistical model and advanced individual-based spatial statistical methods.

2.4 GRAVITY MODELS

Gravity models have been used in various fields to describe certain interactions (or attractions) based on Newton's Law of Gravity, which states that attraction between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. The gravity model is used in spatial epidemiology studies by treating traffic volumes as the migrant host populations invading neighborhood regions. In a simple form, the potential spreading risk of community i from community j can be described by Equation 2.~~10~~:

Au: Please provide equation 2.9 or renumbered equation

$$R_{ij} = \beta \frac{P_i P_j}{d_{ij}^2} \qquad (2.~~10~~)$$

where β is the infection rate, P_i and P_j are the population size of the community i and community j, and d_{ij} is the distance between i and j.

Li et al. (2011) validated the performance of the gravity model in predicting the global spread of H1N1 influenza by formulating the global transmission between major cities and Mexico. The variables associated with the confirmed cases, such as population sizes, per capita gross domestic production (GDP), and the distance between the countries/states and Mexico were taken into account in the gravity model. Li et al. concluded that the modified gravity model is valid for estimating the global transmission trend. The gravity model for infectious disease can be applied in combination with the SIR model to explain the spatial influence at different levels in some particular applications. Viboud et al. (2006) applied a gravity model to summarize between-state workflow movements on a set of stochastic SIR models on the U.S. state level to estimate the influenza epidemic dynamics. Xia et al. (2004) employed a gravity model as a spatial extension of the TSIR model that assesses transmissions between different communities.

The major criticism of the gravity model and its calibration are its lack of theoretical foundations related to human behavior and purely inductive, curve-fitting exercises (Ewing 1974). Filippo et al. (2012) pointed out that the limitations of the gravity model are on fitting the gravity equation formula with multiple parameters and systematic predictive discrepancies. Fitting the gravity model also requires the appropriate sample distribution and quantity. Calibrating a gravity model requires a study to determine the form of the sampling distribution (Kirby 1974). Pearson et al. (1974) stated that the best fit of the trip length distribution (TLD) of urban travel is the Gamma distribution, among other similar distributions. Celik (2010) confirmed that a sample size of approximately1000 for each trip purpose would produce approximately the same parameter estimate as fitting the gravity model with larger sample sizes. Thus, considering the gravity model for estimating disease transmission may have difficulties caused by the high cost and the unreliability involved in the sampling survey. Also, using the gravity model may lead to neglect of large, random contagious events, such as a super disease carrier traveling to the community.

2.5 NETWORK-BASED MODELS

Network-based models are built on the assumption that the spread of human disease follows its specified contact or spreading paths such as transportation or social contact networks (Kretzschmar and Morris 1996; Ghani et al. 1997; May and Lloyd 2001; Newman 2002; Eubank et al. 2004; Keeling 2005; Parham and Ferguson 2006). Network models are used to generalize complex contact networks in order to estimate the infectious disease transmission possibilities. The network model presents individual contacts in a graph as shown in Figure 2.1, constructed with vertices and links that are a pair of sets, usually noted as $G = (V, E)$. The elements of V are the vertices (or nodes), and the elements of E are the edges (or lines) of the graph G. The vertex set of a graph is referred to as $V(G)$, and its edge set as $E(G)$. The properties of nodes and links, and the topology of the graph, can be assigned multiple parameters in order to describe epidemics over space and time. The number of connections of an individual represents the number of links of a node and is useful for describing the topology of a network (Watts and Strogatz, 1998; Albert et al. 2000; Newman 2002). When coupled with infection rate, the number of connections of an individual helps define how a disease spreads throughout a network, and is one of the most important parameters for population-based models (Anderson and May 1991; Keeling 1999; Newman 2002).

Network models can be population- or individual-based, depending on the networks used and the data availability. For example, in a study of modeling 2009 H1N1 pandemic outbreaks, the airline network from Mexico to other cities was used (Figure 2.2) and a global connectivity matrix was established based on the airline network. Metapopulation-based mathematical models were used to simulate the potential H1N1 risk in different cities on the flight network based on flight travel volumes (Khan et al. 2009; Arino et al. 2011). In metapopulation models, the world is divided into geographical regions defining a subpopulation network where connections among subpopulations represent the individual fluxes due to the transportation and mobility infrastructure (Balcan et al. 2009).

Individual-based social contact networks have been used to simulate epidemics at urban scales (Eubank et al. 2004; Eubank 2005; Yang et al. 2007; Carley et al. 2006;

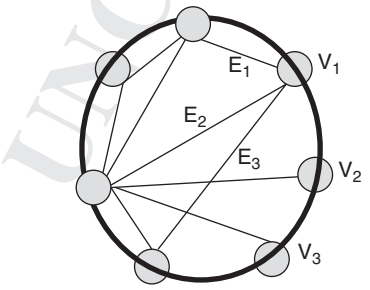
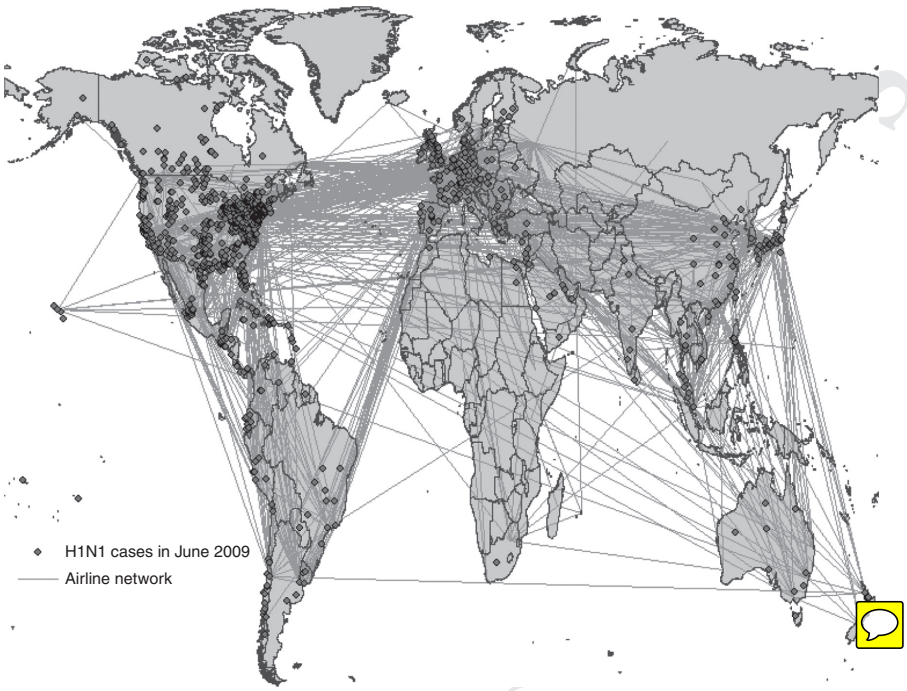


Figure 2.1 An illustration of network graph G. $\{V_1, V_2, V_3 \ldots\}$ represent the nodes in the network, $\{E_1, E_2, E_3, \ldots\}$ are links, representing the spreading paths



Au: The date in the caption for Figure 2.2 has been set as "June 2009" as in the respective artwork. We hope this is O

~~Figure 2.2~~ ~~The illustration of the airline network and H1N1 cases up to June 2009~~

Bian and Leibner 2007; Mao and Bian 2010b). In the theoretical network model created by Bian and Liebner (2007), the length of the latent and infectious period is used as properties of nodes $V(G)$ (individuals). The disease infection rate is used as the property of $E(G)$. Three parameters are used to characterize the connections, namely, the topology of the network. The three parameters refer to (1) number of connections of an individual, (2) degree of interconnection between family members and coworkers, and (3) ratio between workplace connections and family connections. A two-level (community level and urban level) and two-population (home-grouped population and work-grouped population) framework is thus constructed to simulate the epidemic dynamics. In the end, the vulnerability of the communities to disease is evaluated based on the deterministic estimation of parameters from multiple data sources.

Recent mathematical studies have provided extensive insight into network topology, including how it is characterized and how it affects the performance of a network (Watts and Strogatz, 1998; Albert et al. 2000; Keeling 1999). Furthermore, epidemiological work has offered much anecdotal evidence supporting a network structure for describing the spread of diseases (Arita et al. 2003; Francesconi et al. 2003; Meyers et al. 2003; Hsueh et al. 2004; Lau et al. 2004).

A significant challenge to network modeling is the collection of data. Eubank et al. (2004) and Eubank (2005) employed census data, land use data, activity (or time use), and transportation networks in a three-step modeling procedure. A very similar data structure was presented by Yang et al. (2007). A more complex dataset with personal

Au: Note that we have edited the citation "Miao and Bian 2010" as "Mao and Bian 2010." Please check and confirm if the citation is "2010a" or "2010b."

survey information was introduced by Bian and Leibner (2007) in order to estimate the multiple parameters that describe the links. Mao and Bian (2010b) employed similar data structure in larger data size. In addition, without tuning, a network model might not be the best fit for a particular status of disease transmission. An interior probability model is commonly used to tune the host interaction parameters such as infection rate (Bian and Liebner 2007; Mao and Bian 2010). With interior statistical inference absent from the disease host analysis, the network model eventually ends up as deterministic. An alternative strategy is to statistically infer the network models using available host and disease data, as suggested in Welch (2011).

2.6 COMPUTATIONAL SIMULATION APPROACHES

In recent years computational approaches for simulating infectious disease spread in spatially structured environments have been used increasingly in public health prediction and management (Ajelli et al. 2010). The development of simulation methods is closely related to the development of computing technology and programming methods, as well as geolocated demographic, social, and disease data.

2.6.1 Cellular Automation Simulation

Cellular automata (CA) is one type of discrete spatial dispersion (or adjacency) model and has been a commonly used approach for the simulation of dynamic phenomena across space and time (von Neumann 1966; Gutowitz 1991; Adamatzky 2010). A CA consists of a regular discrete lattice (or grid) of cells, each of which is in one of the defined finite states. For each cell, a set of neighborhood cells is defined. An initial state (time $t = 0$) is selected by assigning a state for each cell. Each cell is updated (advancing t by 1) by a state transition function synchronously in discrete time steps, according to some fixed rule (generally, a mathematical function) that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood. A CA is specified by a regular discrete lattice (or grid) of cells, their boundary conditions, a finite set of states, a finite set of cells that defines the interaction neighborhood, and a state transition rule that determines the dynamics of the state of the cells.

CA is premised on the notion that disease transmission is an intrinsically spatial process, in direct contrast with the purely mathematical models, which are not spatially oriented. In CA models, a grid is established, with each cell representing an individual, and disease is transmitted locally from an infected cell to adjacent susceptible cells based on translation rules (usually SIR or extended models) (Beauchemin et al. 2005). Each individual can be at a different stage of infection (latent, infectious, recovered, incubation, symptomatic). CA models incorporate disease translation rules that decide the next state of the current cells, based on the status of the cell and its neighbors as well as additional constraints for future disease transmission. The state of each cell evolves through discrete time steps, with transition rules applied to all cells repeatedly at different time steps.

Au: Please check if the sentence "However, this simple. . . " conveys its intended meani

The key element in CA disease modeling is the design of the impacting neighborhood and driving factors for disease spreading. The simplest CA model uses the direct eight adjacent neighbors of a center cell in the simulation. However, this simple neighborhood definition cannot meet the needs of many disease-spreading applications. Multiple regular or irregular neighborhood rings can be used to differentiate the influence of cells at different distances on the spread of disease. The disease transition rules are usually defined through mathematical or statistical models.

The advantage of CA modeling is its ability to integrate environmental factors (such as land use and soil) and demographic distribution (such as population density and age structure) into the model (Fuentes and Kuperman 1999). It can be easily used to simulate and visualize the spreading and impact of infectious diseases on the population. Furthermore, the population can be divided into subgroups, which enables the simulation of different impacts of the disease on each individual. CA for simulating infectious diseases has been used to discover the behavior of the disease or to work out contingency plans for different diseases (Sloot et al. 2002; Beauchemin et al. 2005; Doran and Laffan 2005; Xiao et al. 2006; Pfeifer et al. 2008).

While these models benefit from enhanced insight into the epidemiological phenomena being studied as compared with the nonspatial models, they remain focused at the population level. Assumptions that disease spreading only occurs in neighboring cells limit the applicability of this model. For example, human daily mobility cannot be modeled in CA models (Pfeifer et al. 2008). CA models are most appropriate for simulating disease transmission among immobile objects such as disease spreading among plants.

2.6.2 Field Simulation Modeling

To address some of the limitations of the SIR model, Venkatachalam and Mikler (2006) have proposed the Global Stochastic Field Simulation (GSFS) model, which takes into account heterogeneous populations, demographic constraints, contact structure, and disease dynamics in order to model the spread of disease. This approach assigns a geographic region to a grid representation and overlays a field encompassing the spatial distribution of population and interaction distributions. It is assumed that each location contains a population of n individuals with associated demographics based on census data. An individual at a specific location is characterized by a state and likelihood of exposure and of contracting the disease of interest. The three possible states (S, I, R) represent an individual's clinical disease stage as previously defined in the SIR model.

To best model the spatial spread of a disease over a geographic region, it is necessary to understand the dynamics of the underlying population and demographics. Publically available datasets such as census data, which outline the composition and behavior of the population, may be utilized to describe the sociodemographic variables, race/ethnicity, age, and sex at different levels of geographic aggregation for the population of interest. Multiple sources of background data may be used to develop a complete picture of the region, and GIS facilitates the integration of this information. This structure is used as an overlay to a global stochastic field that

incorporates the associated census information to define its corresponding interactions among individuals and places.

The GSFS has been used to model the epidemic spread of influenza in Denton City, Texas, with much success. It has also been used to simulate what-if scenarios under different policies for infectious disease outbreaks (Armin et al. 2007). GSFS supports analysis of disease spread in heterogeneous environments and integrates geography, demography, environment, and migration patterns within its framework. While further research is necessary to assess the reliability and validity of this computational model with other applications, it shows great promise for surveillance, monitoring, prevention, and control of infectious diseases, and more effective and efficient utilization of public health resources.

2.6.3 Individual or Agent-Based Modeling

Discrete individual transmission models or agent-based models are based on the belief that individuals are different from one another, and this assumption should form the foundation of epidemiological studies (Koopman and Lynch 1999; Keeling 1999; see Part IV in this book for more details). These models were developed primarily to examine the spread of diseases on the basis of discrete individuals and the contact between them (Ghani et al. 1997; Kretzschmar and Morris 1996; Adams et al. 1998; Welch et al. 1998; Sattenspiel 2009). These models represent a fundamental shift in thinking, as individual network patterns and the ongoing contacts are now viewed as important determinants of population infection levels (Koopman and Lynch 1999; Roche et al. 2008). More specifically, these individual-based models are better able to incorporate these determinants and causal factors directly than the common population-based differential equation models.

Assumptions are made within the agent-based simulation framework to account for the heterogeneity in individuals, the interactions between them, and the heterogeneity of the disease transmission process, both spatially and temporally. For example, to simulate the spread on an urban scale, an agent-based model will assume that (a) individuals may differ with respect to characteristics such as age, race, occupation, as well as infection status; (b) individuals interact with a finite number of other individuals within a given period of time; (c) the number of interactions is not constant and varies by person; (d) individuals are spatially distributed; and (e) individuals are mobile. These assumptions are most appropriate for disease modeling in an urban environment on a daily basis (Bian and Liebner 2007).

Using the aforementioned heterogeneity assumptions as a foundation, the conceptual framework of agent-based models includes several principles such as the representation of unique individuals, their characteristics and behaviors, their relationships with other individuals and with the environment, and how these interactions change through time and space. Object-oriented modeling, with its strong means of representation and abstraction to represent human perceptions of reality, has been identified as an important and widely used approach to support individual-based models (Bian 2000; Maley and Caswell 1993; Silvert 1993). This approach assumes that the world comprises discrete objects and that each individual has unique attributes

and behaviors. Furthermore, individuals may be organized vertically in a hierarchy or horizontally as aggregates, depending on the intended representation using this framework. Agent-based modeling provides a clear structure and allows characteristics of an individual such as age, probability of infection, occupation, infection status, location, and connections to be represented as attributes of the individual, as well as between-group and within-group interactions (Hägerstrand 1970; Pred 1977; Kwan 1999; Miller 2005; Bian and Liebner 2007; Mao and Bian 2011). Consequently, an individual's infection probability, coupled with the contact and interaction among susceptible and infected individuals, will ultimately determine the outcome (Judson 1994; Keeling 1999; Van der Ploeg et al. 1998).

The individualized modeling approach has been shown to have significantly different outcomes from that of population-based approaches. An agent-based modeling approach has been used to simulate the spread of West Nile virus in the United States and Canada (Li et al. 2006; Bouden et al. 2008) by using agents to represent mosquitoes, avian hosts, mammalian hosts, and humans. Interactions among these agents are found within a specific geographic area, and a raster map is used to represent the area as a regular array of cells, which are linked to individual and environmental data such as habitat suitability values, weather conditions, and vegetation cover. Agent-based simulation, combined with social contact networks and deterministic models, have also been widely used to simulate influenza-type diseases (Eubank et al. 2004; Eubank 2005; Carley et al. 2006; Bian and Leibner 2007; Ajelli et al. 2010; Mao and Bian 2010a, 2011).

Most of the current agent-based simulation applications are not predictive models; rather, they are used as tools for researchers to assess and visualize the dynamic behaviors of the system as a result of the interactions among individual agents in the system. The agent-based simulation framework has the capacity to incorporate a larger scale of data heterogeneity. The underlying framework of this model may be applicable for modeling the transmission of many epidemic diseases, and offers researchers an opportunity to better understand the conditions under which an epidemic may occur. In agent-based simulation practices, parameter tuning for individual heterogeneity is a significant challenge due to the complicity of individual behaviors and/or data insufficiency (Chao et al. 2010).

2.7 DISCUSSIONS AND CONCLUSIONS

In the last several decades, there has been a major shift from population-based to individual-based modeling. In general, individual-based models are most appropriate for relatively small spatial extents and populations, for example, cities, communities, neighborhoods, or areas where mobility and heterogeneity are assumed, while population-based models are best suited for modeling pandemics over large homogeneous areas. Most mathematical models for the spread of disease use differential equations based on uniform mixing assumptions and a homogeneous contact process for all individuals within a space. Using actual census data, family age structure, individual infection rates, population-mobility data, and information on contact

networks among people have provided a more realistic representation of heterogeneity of disease propagation. The explicit accounting of spatial distribution and mobility of individuals, in particular, support the modeling of spatial heterogeneity, and the flexibility to incorporate additional parameters into the individual-based analytical framework to represent various conditions and scenarios is a significant advantage of this structure and approach.

Several unique challenges present with individual-based statistical or network models and simulation. First, the increased realism of individual-based representation demands greater accuracy and reliability in the estimation of many aspects of disease transmission with confirmation through clinical or field work. Disease records used in individual-based disease modeling studies come from various surveillance systems. Often the data quality may cause insufficiency in analysis. By reviewing contemporary data capturing methods, Papoz et al. (1996) conclude that the quality of the data obtained with these methods is often far below the standards in specific prevalence surveys and may differ between medical records. Diggle et al. (2003) state that common problems of surveillance data are either underreported or delayed. Furthermore, in practice the surveillance data are collected from many different sources, which means that the observed process is multivariate (Sonesson and Bock, 2003).

Second, to best estimate individual parameters, data is required at the individual level; however, for many variables estimation is drawn from data at the aggregate level such as census data or anecdotal reports. The potential for ecological fallacy in which an inference about an individual is based on aggregate data for a group is an important consideration when assessing the representativeness of the parameter values. Also, many environmental variables and their relationship to disease occurrences are scale dependent. Different results can be yielded with different spatial and temporal scales (Rohani et al. 2003; Robertson et al. 2010).

Third, individual counts and aggregated counts are fundamental data used in epidemiologic analysis (Selvin 2004). The individual-based records have one property in common—they have geographical locations and time intervals. In general, counts that are geographically close will display residual spatial dependence; "residual" here acknowledges that known confounders have been included in the analysis model (Wakefield 2007). Count data could be problematic in many aspects for modeling. The uncertainty contained in the count data sometimes is extreme, for example, high level missing values and many zero counts. Because of the scarcity of the disease data on a small temporal and geographical scale, little information is to be gained from limiting the analysis to raw disease counts (Knorr-Held and Richardson 2003). The overdispersion is also a common issue regarding the count data in epidemic modeling (Ridout et al. 1998).

Overall, disease analysis and modeling techniques are essential parts of understanding and controlling the spreading dynamics of infectious diseases. Recognizing the conditions under which an epidemic may occur and how a particular disease spreads is critical to designing and implementing appropriate and effective public health control measures. Different modeling methods suit different data and purposes. The simple and deterministic nature of classical mathematical models limit their application to modeling the spatial structure of an epidemic. Statistical-based

models are appropriate for revealing relationships among disease risks and potential risk factors. Social contact network models, embodied in agent-based simulation frameworks, may directly capture individual-level heterogeneities. Network models also incorporate multiple data sources into the simulation, which often results in a lack of statistic inference. Gravity model has the advantage of incorporating human traffic information into disease modeling in order to depict the infection, but the calibration demand for survey data is challenging. CA models are simple and easy to use for visualizing disease-spreading results. Agent-based simulation through dynamic models has an advantage over other models with its capability of incorporating individual behavior and mobility info, but it requires a heavy programmable computing effort. As the development of computing capability and programming algorithms increases, as does the wide use of location-based devices for collecting disease-related data and human/animal mobility info, various more robust disease models will undoubtedly aid in planning control measures, evaluating disease intervention strategies, and determining the optimal use of public health resources.

ACKNOWLEDGMENT

This research is supported by grants from the GEOIDE, National Centre of Excellence, and Natural Sciences and Engineering Research Council (NSERC) of Canada. The author would like to thank Frank Wen and Justin Hall for their help in collecting the necessary literature.

REFERENCES

Adamatzky A. (editor). (2010). *Game of Life Cellular Automata*. Springer. 621 p.

Adams A., Barth-Jones D., Chick S. E., and Koopman J. S. (1998). Simulations to evaluate HIV vaccine trial designs. *Simulation*, 71:228–241.

Ajelli M., Goncalves B., Balcan D., Colizza V., Hu H., Ramasco J. J., Merler S., and Vespignani A. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190. Available at http://www.biomedcentral.com/1471-2334/10/190 (accessed May 22, 2014).

Albert R., Jeong H., and Barabasi A. (2000). Error and attack tolerance of complex networks. *Nature*, 406:378–382.

Alkema L., Raftery A. E., and Clark S. J. (2007). Probabilistic projections of HIV prevalence using Bayesian melding. *Annals of Applied Statistics*, 1:229–248.

Anderson R. M. and May R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press.

Anderson R. M., Jackson A. C., May R. M., and Smith A. M. (1981). Population dynamics of fox rabies in Europe. *Nature*, 289:765–771.

Arino J., Bauch C., Brauer F., Driedger S. M., Greer A. L., Moghads S. M., Pizzi N. J., Sander B., Tuite A., van den Driessche P., Watmough J., and Wu J. (2011). Pandemic influenza: modeling and public health perspectives. *Mathematical Biosciences and Engineering*, 8(1):1–20.

Arino J., Hu W., Khan K., Kossowsky D., and Sanz L. (2012). Some methodological aspects involved in the study by the Bio.Diaspora Project of the spread of infectious diseases along the global air transportation network. *Canadian Applied Mathematics Quarterly*, 19(2):125–137.

Arita I., Kojima K., and Nakane M. (2003). Transmission of Severe Acute Respiratory Syndrome. *Emerging Infectious Diseases*, 9:1183–1184.

Armin M., Venkatachalam S., and Ramisetty-Mikler S. (2007). Decisions under uncertainty: a computational framework for quantification of policies addressing infectious disease epidemics. *Stochastic Environmental Research and Risk Assessment*, 21(5):533–542.

Auchincloss A. H., Gebreab S.Y., Mair C., and Diez Roux A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual Review of Public Health*, 33:107–122.

Balcan D., Colizza V., Gonçalves B., Hu H., Ramasco J. J., and Vespignani A. (2009). Multiscale mobility networks and the large scale spreading of infectious diseases. *Proceeding of the National Academy of Sciences of the United States of America*, 106:21484–21489.

Banerjee S., Carlin B. P., and Gelfand A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, 1st edition. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.

Bailey N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd edition. London, 413 p.

~~Barden D., Ridzon R., Parashar U., Teshale E. H., Williams J., Noviello S., Perz J. F., Mast E. E., Swerdlow D. L., and Hadler J. L. (2003). Bioterrorism-related inhalational anthrax in an elderly woman, Connecticut, 2001. *Emerging Infectious Diseases*, 9:681–688.~~

Beale C. M., Lennon J. J., Yearsley J. M., Brewer, M. J., and Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, 13:246–264.

Beauchemin C., Samuel J., and Tuszynski J. (2005). A simple cellular automaton model for influenza A viral infections. *Journal of Theoretical Biology*, 232(2):223–234.

Bernardinelli L., Clayton D., and Montomoli C. (1995). Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine*, 14:2411–2431.

Berke O. (2004). Exploratory disease mapping: kriging the spatial risk function from regional count data. *International Journal of Health Geographics*, 3(18):1–11

Bian L. (2000). Object-oriented representation for modeling mobile objects in an aquatic environment. *International Journal of Geographical Information Science*, 14:603–623.

Bian L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B: Planning and Design*, 31:381–395.

Bian L. and Liebner D. (2007). A network model for dispersion of communicable diseases. *Transactions in GIS*, 11(2):155–173.

Bouden M., Moulin B., and Gosselin P. (2008). The geo-simulation of West Nile Virus propagation: a multi-agent and climate sensitive tool for risk management in Public Health. *International Journal of Health Geographics*, 7:35.

Brauer F. (2008). Compartmental models in epidemiology. In: Brauer F., van den Driessche P., and Wu J. (editors) *Mathematical Epidemiology*, Vol. 1945: Lecture Notes in Mathematics. Mathematical Biosciences Subseries. Springer. pp. 19–79.

Breslow N. E. and Clayton D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*, 88:9–25.

Au: The references "Arino et al. (2012)," "Barden et al. (2003)," "Bian (2004)," "Holmes (1997)," "Wonham et al. (2004)" are not referred to in the text. Please cite them in the text or delete them from the Reference List.

Au: Please provide the publisher information for "Bailey 1975".

Browne W. J. and Draper D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1:473–514.

Carley K. M., Fridsma D. B., Casman E., Yahja A., Altman N., Li-Chiou C., Kaminsky B., and Nave D. (2006). BioWar: scalable agent-based model of bioattacks. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 36(2):252–265.

Celik H. M. (2010). Sample size needed for calibrating trip distribution and behavior of the gravity model. *Journal of Transport. Geography*, 18:83–190.

Chao D. L., Halloran M. E., Obenchain V. J, and Longini I. M. Jr. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computation Biology*, 6 (1):e1000656.

Coburn B. J., Wagner B. G., and Blower S. (2009). Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Medicine*, 7:30.

Cooke K. L., van den Driessche P., and Zou X. (1999). Interaction of maturation delay and nonlinear birth in population and epidemic models. *Journal of Mathematical Biology*, 39:332–352.

Cressie N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc., New York. 887 p.

Culshaw R. V. and Ruan S. (2000). A delay-differential equation model of HIV infection of CD4+ T-cells. *Mathematical Bioscience*, 165:27–39.

Day T., Park A., Madras N., Gumel A. B., and Wu J. (2006). When is quarantine a useful control strategy for emerging infectious diseases? *American Journal of Epidemiology*, 163:479–485.

Deardon R., Brooks S. P., Grenfell B. T., Keeling M. J., Tildesley M. J., Savill N. J., Shaw D. J., and Woolhouse M. E. J. (2010). Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, 20:239–261.

Diekmann O. and Heesterbeek J. A. P. (2000). *Mathematical Epidemiology of Infectious Diseases*. Chichester, UK: John Wiley & Sons, Ltd.

Diekmann O., Heesterbeek J. A. P., and Roberts M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of Royal Society Interface*, 7:873–885.

Diggle P. J., Knorr-Held L., Rowlingson B., Su T. L., Hawtin P., and Bryant T. (2003). Online monitoring of public health surveillance data. In: Brookmeyer R. and Stroup D. F. (editors) *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. Oxford: Oxford University Press. pp. 233–266.

Doran R. J. and Laffan S.W. (2005). Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible infected-recovered cellular automata model. *Preventive Veterinary Medicine*, 70:133–152.

Dunson D. B. (2001). Commentary: practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology*, 153(12):1222–1226.

Dye C. and Gay N. (2003). Modeling the SARS epidemic. *Science*, 300:1884–1885.

Eubank S. (2005). Network based models of infectious disease spread. *Japanese Journal of Infectious Diseases*, 58(6):9–13.

Eubank S. Guclu H., Kumar V. S., Marathe M. V., Srinivasan A., Toroczkai Z., and Wang N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184.

Ewing G. O. (1974). Gravity and linear regression models of spatial interaction: a cautionary note. *Economic Geography*, 50(1):83–88.

Ferguson N. M., Donnelly C. A., and Anderson R. M. (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, 292: 1155–1160.

Filippo S., González M. C., Maritan A., and. Barabási A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484:96–100.

Forde J. E. (2005). Delay differential equation models in mathematical biology. Dissertation. University of Michigan. pp. 94.

Forrester M. L., Pettitt A. N., and Gibson G. J. (2007). Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostatistics*, 8:383–401.

Francesconi P., Yoti Z., Declich S., Onek P. A., Fabiani M., Olango J., Andraghetti R., Rollin P. E., Opira C., Greco D., and Salmaso S. (2003). Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda. *Emerging Infectious Diseases*, 9:1420–1437.

Freeman D. A. (2009). *Statistical Models: Theory and Practice*, revised edition. Cambridge University Press. 456 p.

Fuentes M. and Kuperman M. (1999). Cellular automata and epidemiological model with spatial dependence. *Physica A*, 272:471–486.

Galvani A. (2004). Emerging infections: what have we learned from SARS? *Emerging Infectious Diseases*, 10:1351–1352.

Ghani A. C., Swinton J., and Garnett G. P. (1997). The role of sexual partnership networks in the epidemiology of gonorrhea. *Sexually Transmitted Diseases*, 24:45–56.

Gilberto G.-P., Arenas A. J., Aranda D. F., and Segovia L. (2011). Modeling the epidemic waves of AH1N1/09 influenza around the world. *Spatial and Spatio-temporal Epidemiology*, 2(4): 219–226.

Gourley S. A. (2000). Travelling fronts in the diffusive Nicholson's blowflies equation with distributed time delays. *Mathematical and Computer Modelling*, 32:843–853.

Gourley S. A., Liu R., and Wu J. (2007). Some vector borne diseases with structured host populations: extinction and spatial spread. *SIAM Journal on Applied Mathematics* 67(2):408–433.

Grassly N. C. and Fraser C. (2008). Mathematical models of infectious disease transmission. *Nature Review Microbiology*, 6:477–487. DOI:10.1038/nrmicro1845

Grenfell B. T., Bjornstad O. N., and Kappey J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414:716–723.

Gutowitz H. (editor). (1991). *Cellular Automata: Theory and Experiment*. MIT Press. 483 p.

Hägerstrand T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24:7–21.

Hejblum G., Setbon M., Temime L., Lesieur S., Valleron A.-J. (2011). Modelers' perception of mathematical modeling in epidemiology: a web-based survey. *PLoS One*, 6(1):e16531. DOI:10.1371/journal.pone.0016531

Hethcote H. W. (2009). The basic epidemiology models: models, expressions for R_0, parameter estimation, and applications. In: Ma S. and Xia Y. (editors) *Mathematical Understanding of Infectious Disease Dynamics*, vol. 16. Lecture Notes Series. Institute for Mathematical Sciences, National University of Singapore. pp. 1–52.

Au: Please check if the edits made to "Hethcote 2009," "Ridout et al. 1998," "Sloot et al. 2002," and "Venkatachalam and Mikler 2006" are OK

Holmes E. E. (1997). Basic epidemiological concepts in a spatial context. In Tilman D. and Kareiva P. (editors) *Spatial Ecology*. Princeton, NJ: Princeton University Press. pp. 111–136.

Hsueh P., Chen P., Hsiao C., Yeh S., Cheng W., Wang J., Chiang B., Chang S., Chang F., Wong W., Kao C., and Yang P. (2004). Patient data, early SARS epidemic, Taiwan. *Emerging Infectious Diseases*, 10:489–493.

Jandarov R., Haran M., Bjørnstad O., and Grenfell B. (2013). Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. Available at http://arxiv.org/pdf/1110.6451v3.pdf (accessed May 22, 2014).

Judson O. P. (1994). The rise of the individual-based model in ecology. *Trends in Ecology & Evolution*, 9:9–14.

Kallen A., Arcuri P., and Murray J. D. (1985). A simple model for the spatial spread and control of rabies. *Journal of Theoretical Biology*, 116:377–393.

Kaplan D. T. (2011). *Statistical Modeling: A Fresh Approach*, 2nd edition. 381 p.

Keeling M. J. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London, Series B*, 266:859–867.

Keeling M. J. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67:1–8.

Keeling M. J. and Ross J. V. (2008). On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface* 5:171–181.

Kermack W. O. and McKendrick A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, 115(772):700–721.

Khan K. Arino J., Hu W., Raposo P., Sears J., Calderon F., Heidebrecht C., Macdonald M., Lieuw J., Chan A., and Gardam M. (2009). Spread of a novel in influenza A (H1N1) virus via global airline transportation, *New England Journal of Medicine*, 361:212–214.

Kirby H. (1974). Theoretical requirements for calibrating gravity models. *Transportation Research*, 8:97–104.

Knorr-Held L. and Richardson S. (2003). A hierarchical model for space–time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society. Series C. (Applied Statistics)*, 52(2):169–183.

Koopman J. S. and Lynch J. W. (1999). Individual causal models and population system models in epidemiology. *American Journal of Public Health*, 89(8):1170–1174.

Kretzschmar M. and Morris M. (1996). Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences*, 133:165–195.

Kwan M.-P. (1999). Gender and individual access to urban opportunities: a study using space–time measures. *Professional Geographer*, 51:210–227.

Lau J. T. F., Tsui H., Lau M., and Yang X. (2004). SARS transmission, risk factors, and prevention in Hong Kong. *Emerging Infectious Diseases*, 10:587–592.

Lavine M. L. (1999). What is Bayesian statistics and why everything else is wrong? *The Undergraduate Mathematics and its Applications Journal*, 20:165–174.

Lawson A. B. (2005). *Statistical Methods in Spatial Epidemiology*. John Wiley & Sons, Ltd. 277 p.

Lawson A. B. (2008*). Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Boca Raton, FL: CRC Press.

LeSage J. P. (1997). Regression analysis of spatial data. *The Journals of Regional Analysis & Policy*, 27(2):83–94.

Lewis M., Renclawowicz J., and van den Driessche P. (2006). Traveling waves and spread rates for a West Nile virus model. *Bulletin of Mathematical Biology*, 68:3–23.

Li Z., Hlohowsky J., Smith K., and Smith R. (2006). Agent-based model for simulation of West Nile virus transmission. *Proceedings of the Agent 2006 Conference on "Social Agents: Results and Prospects."* The University of Chicago. pp. 1–13.

Li X., Tian H., Lai D., and Zhang, Z. (2011). Validation of the gravity model in predicting the global spread of influenza. *International Journal of Environmental Research and Public Health*, 8:3134–3143.

Lin C. and Wen T. (2011). Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *International Journal of Environmental Research and Public Health*, 8:2798–2815.

Lloyd A. L. and Valeika S. (2007). Network models in epidemiology: an overview. In: Blasius B., Kurths J., and Stone L.(editors) *Complex Population Dynamics: Nonlinear Modeling in Ecology, Epidemiology and Genetics*. World Scientific.

Maley C. C. and Caswell H. (1993). Implementing *i*-state configuration models for population dynamics: an object-oriented programming approach. *Ecological Modeling*, 68:75–89.

Mao L. and Bian L. (2010a). Spatial–temporal transmission of influenza and its health risks in an urbanized area. *Computers, Environment and Urban Systems*, 34:204–215.

Mao L. and Bian L. (2010b). A dynamic social network with individual mobility for designing intervention strategies. *Transactions in GIS*, 14(4):533–545.

Mao L. and Bian L. (2011). Massive agent-based simulation for a dual-diffusion process of influenza and human preventive behavior. *International Journal of Geographical Information Science*, 25(9):1371–1388.

May R. M. and Lloyd A. L. (2001). Infection dynamics on scale-free networks. *Physical Review E: Statistical. Nonlinear, and Soft Matter Physics*, 64:066112.

McKenzie F. E. (2004). Smallpox models as policy tools. *Emerging Infectious Diseases*, 10:2044–2047.

McV Messam L. L., Branscum A. J., Collins M. T., and Gardner I. A. (2008). Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease. *Animal Health Research Reviews* 9:1–23.

Meyers L. A., Newman M. E. J., Martin M., and Schrag S. (2003). Applying network theory to epidemics: control measures for mycoplasma pneumoniae outbreaks. *Emerging Infectious Diseases*, 9:204–210.

Miller H. J. (2005). A measurement theory for time geography. *Geographical Analysis*, 37:17–45.

Moghadas S. M., Pizzi N., Wu J., and Yan P. (2008). Managing public health crises: the role of models in pandemic preparedness, *Influenza and Other Respiratory Viruses*, 3:75–79.

Moore D. A. and Carpenter T. E. (1999). Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews*, 21(2):143–161.

Morgenstern H. (1982). Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health*, 72(12):1336–1344.

Au: Please provide the page ranges for "Lloyd and Valeika 2007," "Roberts and Heesterbeek 2003," and "Waller 2005, 2007."

Murray J. D., Stanley A. E., and Brown D. L. (1986). On the spatial spread of rabies among foxes. *Proceedings of the Royal Society of London, Series B*, 229:111–150.

Nelson P. W. and Perelson A. S. (2002). Mathematical analysis of delay differential equation models of HIV-1 infection. *Mathematical Bioscience*, 179:73–94.

Newman M. E. J. (2002). The spread of epidemic disease on networks, *Physical Review E*, 66:016128.

Noble J. V. (1974). Geographic and temporal development of plagues. *Nature*, 250:726–728.

Okubo A. (1998). Diffusion-type models for avian range expansion. In: Quellet H. (editor) *Acta XIX Congressus Internationalis Ornithologici*. National Museum of Natural Sciences, University of Ottawa Press. pp. 1038–1049.

Papoz L., Balkau B., and Lellouch J. (1996). Case counting in epidemiology: limitations of methods based on multiple data sources. *International Journal of Epidemiology*, 25:474–478.

Parham P. E. and Ferguson N. M. (2006). Space and contact networks: capturing the locality of disease transmission. *Journal of the Royal Society Interface* 3:483–493.

Pearson D. F., Stover V. G., and Benson J. D. (1974). A Procedure for Estimation of Trip Length Frequency Distributions. Texas Transport Institute. Report No.: TTI-2-10-74-17-1.

Pfeifer B., Kugler K., Tejada M. M., Baumgartner C., Seger M., Osl M., Netzer M., Handler M., Dander A., Wurz M., Graber A., and Tilg B. (2008). A cellular automation framework for infectious disease spread simulation. *Open Medical Informatics Journal*, 2:70–81.

Pred A. (1977). The choreography of existence: comments on Hägerstrand's time geography and its usefulness. *Economic Geography*, 53:207–221.

Rhodes C. J. and Anderson R. M. (1997). Epidemic threshold and vaccination in a lattice model of disease spread. *Theoretical Population Biology*, 52:101–118.

Ridout M., Demetrio C. G. B., and Hinde J. (1998). Models for count data with many zeros. *Proceedings of the XIXth International Biometric Conference, Invited Papers*, December 14–18, 1998, Cape Town, South Africa. International Biometric Society. pp. 179–192.

Riley S. 2007. Large-scale spatial-transmission models of infectious disease. *Science* 316:1298–1301.

Roberts M. G. and Heesterbeek J. A. P. (2003). Mathematical models in epidemiology. In: Filar J. A. and Krawczyk J. B. (editors) *Mathematical Models in Encyclopedia of Life Support Systems (EOLSS)*. Oxford, UK: Eolss Publishers.

Robertson C., Nelson T. A., MacNab, Y. C., and Lawson A. B. (2010). Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1 (2–3):105–116.

Roche B., Guégan J.-F., and Bousquet F. (2008). Multi-agent systems in epidemiology: a first step for computational biology in the study of vector-borne disease transmission. *BMC Bioinformatics*, 9:435. DOI:10.1186/1471-2105-9-435

Rohani P., Green C. J., Mantilla-Beniers N. B. and Grenfell B. T. (2003). Ecological interference between fatal diseases. *Nature* 422:885–888.

Ruan S. and Xiao D. (2004). Stability of steady and existence of travelling waves in a vector-disease model. *Proceedings of the Royal Society of Edinburgh A*, 134: 991–1011.

Sattenspiel L. (2009). *The Geographic Spread of Infectious Diseases, Models and Applications*. Princeton University Press. 304 p.

Schwartz S. (1994). The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *American Journal of Public Health*, 84(5): 819–824.

Scott J. G. (2010). Statistical Modeling: A Gentle Introduction [Lecture notes]. Available at www.mccombs.utexas.edu/faculty/james.scott (accessed May 22, 2014).

Selvin S. (2004). *Statistical Analysis of Epidemiologic Data*, 3rd edition. Oxford: Oxford University Press. 487 p.

Silvert W. (1993). Object-oriented ecosystem modeling. *Ecological Modeling*, 68:91–118.

Sloot P, Chen F, and Boucher C. (2002). Cellular automata model of drug therapy for HIV infection. In: Bandini S., Chopard B., and Tomassini M. (editors) *Cellular Automata: Proceedings of the 5th International Conference on Cellular Automata for Research and Industry (ACRI)*, October 9–11, 2002, Geneva, Switzerland. Berlin, Heidelberg: Springer. pp. 282–293.

Snow J. (1849). *On the Mode of Communication of Cholera*. London: John Churchill. 31 p.

Sonesson C. and Bock D. (2003) A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A*, 166:5–21.

Tuckwell H. C. and Williams R. J. (2007). Some properties of a simple stochastic epidemic model of SIR type. *Mathematical Biosciences*, 208(1):76–97.

Van der Ploeg C. P. B., Van Vliet C., De Vlas S. J., Ndinya-Achola J. O., Fransen L., Van Oortmarssen G. J., and Habbema, J. D. F. (1998). STDSIM: a microsimulation model for decision support in STD control. *Interfaces*, 28:84–100.

Venkatachalam S. and Mikler A. R. (2006). *Modeling Infectious Diseases Using Global Stochastic Field Simulation*. 2006 IEEE International Conference on Granular Computing, May 10–12, 2006, Atlanta, GA. IEEE. pp. 750–753.

von Neumann J. (1966). *Theory of Self-Reproducing Automata*, compiled by A. W. Burks. University of Illinois Press. 87 p.

Viboud C., Bjørnstad O. N., Smith D. L., Simonsen L., Miller M. A., and Grenfell B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312 (5772): 447–451.

Wakefield J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8 (2):158–183.

Wakefield J., Kelsall J. E, and Morris, S. E. (2000). Clustering, cluster detection, and spatial variation in risk. In: Elliott P., Wakefield J., Best N., and Briggs D. J. (editors) *Spatial Epidemiology: Methods and Applications*. Oxford University Press. pp. 128–152.

Waller L. A. (2005). Bayesian thinking in spatial statistics. In: Dey D. K. and Rao C. R. (editors) *Bayesian Thinking: Modeling and Computation*. Vol. 25 (Supplement 1): Handbook of Statistics. Elsevier.

Waller L. A. (2007). Spatial Epidemiology. In: Biswas A., Datta S., Fine J. P., and Segal M. R. (editors) *Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics*. Hoboken, NJ: John Wiley & Sons, Inc.

Wang F. (2006). *Quantitative methods and applications in GIS*. Boca Raton, FL: CRC Press.

Wang Z.-C. and Wu J. (2010). Travelling waves of a diffusive Kermack–McKendrick epidemic model with non-local delayed transmission. *Proceedings of the Royal Society A*, 466:237–261. DOI:10.1098/rspa.2009.0377

Wang X.-S., Wang H., and Wu J. (2012). Travelling waves of diffusive predator–prey systems: disease outbreak propagation. *Discrete and Continuous Dynamical Systems*, 32(9):3303–3324.

Watts D. J. and Strogatz S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.

Welch G., Chick S. E., and Koopman J. S. (1998). Effect of concurrent partnerships and sex-act rate on gonorrhea prevalence. *Simulation*, 71:242–249.

Welch D., Bansal S., and Hunter, D. R. (2011). Statistical inference to advance network models in epidemiology. *Epidemics* 3(1):38–45.

Wonham M. J., de Camino-Beck T., and Lewis M. (2004). An epidemiological model for West-Nile virus: invasion analysis and control application. *Proceedings of the Royal Society of London, Series B*, 271:501–507.

Xia Y., Bjornstad O. N., and Grenfell B. T. (2004). Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281.

Xiao X., Shao S. H., and Chou C. K. (2006). A probability cellular automaton model for hepatitis B viral infections. *Biochemical and Biophysical Research Communications*, 342:605–615.

Yang Y., Atkinson P., and Ettema D. (2007). Individual space–time activity-based modelling of infectious disease transmission within a city. *Journal of the Royal Society Interface*, 5:759–772.

Zhang J. (2009). Existence of travelling waves in a modified vector-disease model. *Applied Mathematical Modeling*, 33(2):626–632.

Zhong S., Xue Y., Cao C., Cao W., Li X., Guo J., and Fang L. (2005). Explore disease mapping of hepatitis b using geostatistical analysis techniques. *Lecture Notes in Computer Science*, 3516: 464–471.