STOCHASTIC METAPOPULATION EPIDEMIC MODELS: A NUMERUS MODEL BUILDER™ IMPLEMENTATION

Wayne M Getz

Department Environmental Science, Policy and Management
University of California
Berkeley, CA 94720-3112, USA
wgetz@berkeley.edu

Richard Salter

Numerus, Inc. 850 Iron Point Rd
Folsom, CA 95630, USA
Richard.Salter@numerusinc.com

ABSTRACT

Here we formulate a discrete-time, stochastic SEIR (susceptible, exposed, infectious, removed) epidemic model, embedded in a metapopulation network and code it using the NUMERUS MODEL BUILDER™ Platform. Beyond being the first such formulation our approach includes three important innovations: i) the inclusion of a core-periphery concept for subpopulation centers; ii) the concept of an effective population at risk; iii) the elaboration of the transmission rate processes in terms of an intrinsic contact rate modified by a source density component to produce an effective contact rate, with the standard probability of transmission per effective contact added on. We discuss our model in the context of simulations that demonstrate the extent to which an assumption of spatial homogeneity in a Sierra-Leone-type Ebola outbreak can greatly mislead public health officials managing epidemics.

Keywords: SIR, SEIR, disease dynamics, effective-population-at-risk, Ebola, Sierra Leone

1 INTRODUCTION

Systems of ordinary differential equations in the disease-class variables S (susceptibles), E (exposed), I (infected) and R (removed) have been the mainstay of modeling the dynamics of disease outbreaks for the past 100 years (Smith et al. 2012, Hethcote 2000). In most cases, these so-called SEIR models assume homogeneity with respect to both disease class and spatial structure. These models have been greatly elaborated to take into account age (Castillo-Chavez et al. 1989), sex (Leclerc et al. 2009), the genetic structure of hosts and pathogens (Koelle et al. 2006, Gilchrist and Sasaki 2002), and the fact that the time spent by groups of individuals in the different disease classes follows a lognormal or gamma rather than exponential distribution (Getz and Dougherty 2018). Attention has also been paid to the incorporation of spatial structure (Lloyd and Jansen 2004, Watts et al. 2005, Kramer et al. 2016, Keeling 1999, Balcan et al. 2010), but more can be done in the context of embedding full SEIR models and their demographic and epidemiological elaborations in a metapopulation setting. This embedding process can be greatly facilitated through use

of the NUMERUS MODEL BUILDER™(NMB) graphically-driven, computer coding platform, because of its unique hierarchical coding architecture.

The purpose of this paper is three-fold: one, we present for the first time details of an SEIR discrete-time stochastic population model embedded in a metapopulation framework; two, we discuss coding implementation using the NMB™ network, mapping, and hierarchical chip generation tools; three, we discuss the suitability of the formulation for modeling real-time outbreaks such as the recent Ebola Viral Disease (EVD) outbreak in West Africa (Kramer et al. 2016, Getz et al. 2015). Additional novelties in our approach include implementation of a core-periphery concept for each subpopulation, because of its utility in modeling epidemics spreading through a combination of rural and urban centers. They also include the concept of an effective population at risk, as well as elaboration of the transmission rate processes in terms of an intrinsic contact rate modified by a source density component to produce an *effective contact rate* on an *effective population at risk*, with the standard probability of transmission per effective contact added on. These transmission details facilitate fine tuning transmission rates in different subpopulations at various densities, when both frequency and density-dependent (a.k.a mass action) processes are at play, as is the case for most diseases (McCallum et al. 2001). Finally, our illustrative simulations provide new insights into the 2014 Ebola Viral Disease (EVD) outbreak in Sierra Leone, particularly with regard to the role that population spatial structure plays in determining the spatial pattern of these kinds of outbreaks.

2 MODEL CONSTRUCTION

SEIR epidemic models are typically formulated as systems of ordinary differential equations (ODEs) in the variables S (susceptible), E (infected but not yet infectious), I (infectious) and R (removed) (Hethcote 2000). Discrete-time equivalents of these models are particularly useful when fitting models to periodically reported data (e.g. daily, weekly or monthly rates) or evaluating interventions (treatment regimens, isolation of patients) that are themselves discrete (e.g. daily rates) (Getz and Lloyd-Smith 2006, Getz et al. 2006). Further, these discrete formulations are more easily extended than continuous time formulations to stochastic settings, where the latter require implementation of event driven formulations (Gillespie 1976, Allen 2017, Britton 2010, Getz and Lloyd-Smith 2006, Getz and Dougherty 2018).

2.1 Basic Stochastic Model

In the presenting our discrete-time stochastic SEIR model, we break R into V (recovered with temporary immunity) and D (Getz and Lloyd-Smith 2006, Getz et al. 2017) (Figure 1). Also, we use the roman fonts X=S, E, I, V and D to name classes and the italic fonts $X = S, E, I, V$ and D to represent the values of the corresponding variables. In terms of model functions and parameters, we use γ^{X} to denote flows (i.e. transitions) from compartment X to the next compartment, in the sequence, as depicted in Figure 1. It is useful to note that from a distributional point of view, individuals pass through these compartments in an exponential fashion, which implies that the reciprocal of these rates are the means of this distribution X: i.e. viewed from a stochastic point of view, the process is Poisson (Wearing et al. 2005)). Thus $1/\gamma^{\mathrm{E}}$ is the mean duration of the so-called latent period, $1/\gamma^{\mathrm{I}}$ is the mean duration of the infectious period and $1/\gamma^{\mathrm{R}}$ is the mean duration of the immunity period on recovery from the infection. The parameters μ and α in Figure 1 are the natural and disease-induced per capita death rates.

These transition and mortality rate parameters

At the end of each iteration interval all individuals in each compartment parameters q^{X} are the per-capita rates at which individuals flow out of each of the superscripted disease classes during each interval of time. These individuals are then assumed to re-assort themselves and move back into the different subpopulations at the next beat of the discrete time click, where these rates of inflow H^{X} depend both on the outflow rates q^{X}

and variables, as formulated in the next section. The variable Λ represents the total recruitment at each beat of the clock. This recruitment is often ignored in cases where epidemic processes occur at much faster time scales than demographic recruitment processes. This is the case for fast diseases such as isolated measles outbreaks or influenza, but not slow disease such as tuberculosis or HIV (Getz et al. 2006).

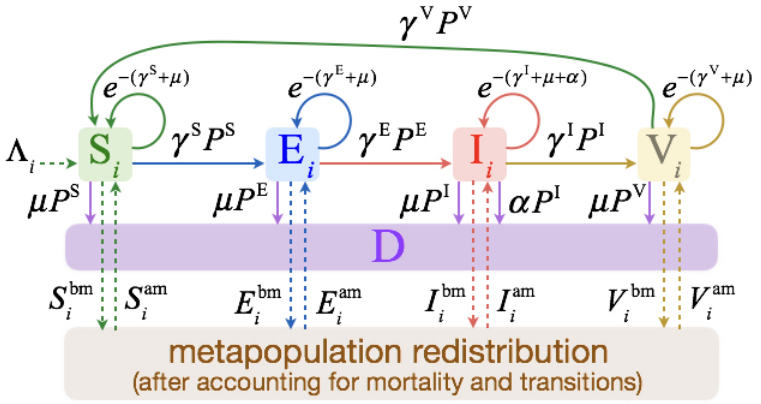

Figure 1: SEIVD compartment model for a focal subpopulation. Transition probabilities are based on a competing rates formulation, as detailed in Equations 1–6. The rates are the per-capita transition rates γ^{X} from disease class X=S, E, I, or V and the per-capita natural and disease-induced mortality rates μ and α (note: the latter applies only to infectious class I). Metapopulation redistribution is assume to occur at the end of each time interval where the superscripts "bm" and "am" refer to before and after "mixing" respectively. The recruitment rates Λ_i are assumed to be subpopulation specific and likely scale with the total population size $N_i = S_i + E_i + I_i + V_i$.

The per-capita-S transmission rate, γ^{S}, is a function rather than a constant parameter, because it depends on an the proportion I/N of infected individuals in the local subpopulation (this is the frequency dependent component of γ^{S}). It also depends on an intrinsic contact rate $\kappa(N)$, that itself may depend on the local density N of the population: individuals in high-density environments likely encounter others (on buses, in the streets, and so on) during their daily routines at higher rates than individuals in low density environments (this is the density dependent component of γ^{S}). Finally γ^{S} is scaled by the probability p_β (in continuous models it is conventional to use the parameter β to scale the transmission rate) of pathogen transmission per effective contact. In many epidemics, the effective contact rate $\kappa(N,t)$ declines over time because of precautions taken to reduce transmission. Thus we introduce a time-dependent effective contact rate parameter such that the per-capita transmission rate takes the form

$$\gamma^{S}(I(t),N(t),t) = \frac{p_\beta \kappa(N(t),t) I(t)}{N(t)}, \quad \text{where} \quad N(t) = S(t) + E(t) + I(t) + V(t) \tag{1}$$

and (cf. Getz et al. 2017)

$$\kappa(N,t) = \frac{\kappa_0(N)}{1 + \left(\frac{t}{t_c}\right)^{\varepsilon}} \tag{2}$$

for parameters $t_c > 0$ and $\varepsilon > 1$ (i.e., a reversed S-shaped relation with abruptness parameter ε analogous to the onset of density-dependence discussed in (Getz 1996)).

In formulating a discrete-time equivalent of an SEIVD model with the various transition, movement and mortality rates listed above, we need to convert these instantaneous rates either to proportions (deterministic models) or to probabilities (stochastic models). The simplest way to do this is to allocate, for example, the proportions/probabilities $p_{\gamma|\alpha+\mu}$, $p_{\alpha|\gamma+\mu}$ and $p_{\mu|\gamma+\alpha}$ of individuals that transition out of class I during

tick of the clock into those that respectively recover, die from disease, die from natural causes, and leave the subpopulation in proportion to the rates γ, α, and μ themselves. For X=S, E, I or V, this results in the so-called "competing rates" transition probabilities given by (Andersen et al. 2012, Getz et al. 2017, Getz and Dougherty 2018):

$$p_{\gamma^{\mathrm{X}}|\alpha^{\mathrm{X}}+\mu} = \gamma^{\mathrm{X}} P^{\mathrm{X}}, \qquad p_{\alpha^{\mathrm{X}}|\gamma^{\mathrm{X}}+\mu} = \alpha^{\mathrm{X}} P^{\mathrm{X}}, \qquad p_{\mu|\gamma^{\mathrm{X}}+\alpha^{\mathrm{X}}} = \mu P^{\mathrm{X}} \tag{3}$$

where

$$P^{\mathrm{X}} = \frac{\left(1 - e^{-(\gamma^{\mathrm{X}}+\alpha^{\mathrm{X}}+\mu)}\right)}{\gamma^{\mathrm{X}}+\alpha^{\mathrm{X}}+\mu}, \quad \text{X=S, E, I or V} \tag{4}$$

Note, for notational convenience we have generalized the disease-induced mortality rate parameter, which we then restrict to

$$\alpha^{\mathrm{X}} = 0 \ \text{ for X=S, E, or V;} \ \ \text{else} \ \ \alpha^{\mathrm{I}} = \alpha \tag{5}$$

to obtain the formulation depicted in Figure 1. We also note that it can easily be shown that

$$p_{\gamma|\alpha^{\mathrm{X}}+\mu} + p_{\alpha^{\mathrm{X}}|\gamma+\mu} + p_{\mu|\gamma+\alpha^{\mathrm{X}}} = p_{\gamma+\alpha^{\mathrm{X}}+\mu}$$

or, equivalently (with relevance to Figure 1),

$$1 - \left(\gamma^{\mathrm{X}} + \alpha^{\mathrm{X}}\right) P^{\mathrm{X}} = e^{-(\gamma^{\mathrm{X}}+\alpha^{\mathrm{X}}+\mu)} \tag{6}$$

This verifies that all transitions depicted in Figure 1, when D and the population external to the focal subpopulation are included, sum to 1.

In developing a stochastic formulation we use the notation $(\hat{X}_1, \cdots, \hat{X}_r)$ to denote one drawing of $(x_1, \cdots, x_r)$ from a *multinomial distribution*: i.e. $(\hat{X}_1, \cdots, \hat{X}_r) \sim \text{MULTINOMIAL}\left[n; p_1, \cdots, p_r\right]$, where $\hat{X}_i$ is the number of times one of r possible outcomes occurs over n trials, each have probability p_i $(i = 1, ..., r)$ of occurring in any one trial. With this notation, and dropping arguments in t to keep the equations compact, we obtain the model for X=S, E, I, and V (i.e. 4 equations):

$$\left(\hat{X}, \hat{\Gamma}^{\mathrm{X}}, \hat{A}^{\mathrm{X}}, \hat{D}^{\mathrm{X}}\right) := \text{MULTINOMIAL}\left[X; 1 - (\gamma^{\mathrm{X}} + \alpha^{\mathrm{X}} + \mu)P^{\mathrm{X}}, \gamma^{\mathrm{X}} P^{\mathrm{X}}, \alpha^{\mathrm{X}} P^{\mathrm{X}}, \mu P^{\mathrm{X}}\right] \tag{7}$$

Note similar equations are presented in Getz et al. 2017, but in expanded and less general form.

Once these drawings have been made at time t, say, we can use the values obtained to move the stochastic simulation one step forward using the equations (cf. (Getz et al. 2017))

$$\begin{aligned} S(t+1) &= \hat{S}(t) + \hat{\Lambda}_t + \hat{\Gamma}^{\mathrm{V}}(t) \\ E(t+1) &= \hat{E}(t) + \hat{\Gamma}^{\mathrm{S}}(t) \\ I(t+1) &= \hat{I}(t) + \hat{\Gamma}^{\mathrm{E}}(t) \\ V(t+1) &= \hat{V}(t) + \hat{\Gamma}^{\mathrm{I}}(t) \end{aligned} \tag{8}$$

where the recruitment rates $\hat{\Lambda}(t+1)$ can be generated each time period using an appropriate stochastic process. The simplest such process is a one parameter Poisson, where the parameter value my be subpopulation size dependent if recruitment is influenced by local population per capita birth rates. We can also keep track of individuals that have died from causes other than the disease and those that have died from the disease itself through the equations

$$\begin{aligned} D(t+1) &= D(t) + \hat{D}^{\mathrm{S}}(t) + \hat{D}^{\mathrm{E}}(t) + \hat{D}^{\mathrm{I}}(t) + \hat{D}^{\mathrm{V}}(t) \\ A(t+1) &= A(t) + \hat{A}^{\mathrm{S}}(t) + \hat{A}^{\mathrm{E}}(t) + \hat{A}^{\mathrm{I}}(t) + \hat{A}^{\mathrm{V}}(t) \end{aligned}$$

2.2 Metapopulation Embedding

Our primary interest here is to embed our SEIVD model in a metapopulation setting. The first step to embedding Equations 1–8 in such a setting is to introduce a subscript i used to denote the i^{th} subpopulation in the metapopulation complex. We also use the notation X_i^{bm} to represent the individuals left in disease class X in subpopulation i after computation of the outflows due to mortality and transitions to the next disease class have been computed. In addition, we use the notation X_i^{am} to denote the number of individuals in disease class X after we have redistributed individuals by age class back into the different subpopulations once we have accounted for disease class transfers and the movement process among subpopulations.

With this notation, after computing the drawings

$$\left(\hat{X}_i(t),\hat{\Gamma}_i^{\text{X}}(t),\hat{A}_i^{\text{X}}(t),\hat{D}_i^{\text{X}}(t)\right) := \text{MULTINOMIAL}\left[X_i^{\text{am}}(t);1-(\gamma^{\text{X}}+\alpha^{\text{X}}+\mu)P^{\text{X}},\,\gamma^{\text{X}}P^{\text{X}},\,\alpha^{\text{X}}P^{\text{X}},\,\mu P^{\text{X}}\right] \quad (9)$$

the total number of individuals in disease class X (X=S, E, I, and V) in subpopulation i as time approaches the end of the right-open interval $[t,t+1)$—that is, before metapopulation redistribution takes place—is

$$\begin{aligned} S_i^{\text{bm}}(t+1) &= \hat{S}_i(t)+\hat{\Lambda}_i(t)+\hat{\Gamma}_i^{\text{V}}(t) \\ E_i^{\text{bm}}(t+1) &= \hat{E}_i(t)+\hat{\Gamma}_i^{\text{S}}(t) \\ I_i^{\text{bm}}(t+1) &= \hat{I}_i(t)+\hat{\Gamma}_i^{\text{E}}(t) \qquad i=1,...,m \\ V_i^{\text{bm}}(t+1) &= \hat{V}_i(t)+\hat{\Gamma}_i^{\text{I}}(t) \end{aligned} \quad (10)$$

To compute the redistribution of individuals among the m subpopulations we define a set of redistribution matrices Π^{X}, X=S, E, I, and V, consisting of elements π_{ij}^{X} that allocate the proportion of individuals that have come from class X in subpopulation j over the interval $[t,t+1)$ to class X in subpopulation i at time t. We assume that the movement elements $\pi_{ij}^{\text{X}}(t)$ are derived from two sets of parameters. First, a set of connectivity strengths $c_{ij}^{\text{X}}(t)$ that reflect that relative ease with which an individual in disease class X in subpopulation j can move to disease class X in subpopulation i over the time interval $[t,t+1)$. Second, a set of relative attractivity values $a_i(t)$ that are characteristics of the nodes $i=1,...,m$. We note that c_{ii} accounts for those individuals who actually stay where they are so that $1-c_{ii}/\sum_{i=1}^{m}c_{ij}$ is proportion of individuals that move. The attractivity values $a_i(t)$, formulated more fully below, are assumed to bias the movement of any individual leaving subpopulation j to move to subpopulation i with probabilities computed using the formula:

$$\pi_{ij}^{\text{X}}(t)=\frac{a_i(t)c_{ij}^{\text{X}}}{\sum_{r=1}^{m}a_r(t)c_{rj}^{\text{X}}} \quad (11)$$

Note, by construction, $\sum_{i=1}^{m}\pi_{ij}^{\text{X}}=1$ for $i=1,...,m$ and X=S, E, I, and V.

The attractivity factors $a_i(t)$ are likely to reflect several different factors, including subpopulation size, the proportion of infected or immune individuals in the subpopulation, and so on. As in (Getz et al. 2017), we assume the attractivity of a subpopulation i falls off on either side of its most desirable size N_i^* and also declines as the ratio of infectious individuals $I_i(t)/N_i(t)$ for the i^{th} population, $i=1,...,m$, increases. Thus, for a weighting parameter $k\in[0,1]$, we use the following form that trades off the relative importance of the most desirable size with the avoidance of infectious individuals—i.e.,

$$a_i(t)=k\left(1-\frac{I_i(t)}{N_i(t)}\right)+(1-k)\left(\frac{N_i}{eN_i^*}e^{-N_i(t)/N_i^*}\right) \quad (12)$$

We are now ready to compute the redistributed numbers using the multinomial distributions

$$\left(\hat{X}_{1j}^{\text{am}}(t+1),\cdots,\hat{X}_{mj}^{\text{am}}(t+1)\right)=\text{MULTINOMIAL}[\hat{X}_j^{\text{bm}}(t+1);\pi_{ij}^{\text{X}}(t),...,\pi_{mj}^{\text{X}}(t)] \quad (13)$$

and then finally compute the redistributed number of individuals in each disease class for each of the subpopulations:

$$X_i^{\text{am}}(t+1) = \sum_{j=1}^{m} \hat{X}_{ij}^{\text{am}}(t+1), \quad \text{X=S, E, I, and V, and } i = 1,\ldots,m \tag{14}$$

(where we note the hat is dropped in the left-hand-side of this equation because this is a deterministic computation rather than a stochastic drawing. We also note that total number of individuals that have died over the course of the epidemic from natural causes is

$$D(t+1) = D(t) + \sum_{i=1}^{m} \left(\hat{D}_i^{\text{S}}(t) + \hat{D}_i^{\text{E}}(t) + \hat{D}_i^{\text{I}}(t) + \hat{D}_i^{\text{V}}(t)\right)$$

and those that those that have died from disease is

$$A(t+1) = A(t) + \sum_{i=1}^{m} \hat{A}_i^{\text{I}}(t)$$

2.3 NUMERUS MODEL BUILDER™ Implementation

A detail description of how the NUMERUS MODEL BUILDER™ (NMB) platform can be used to code SEIVD models in deterministic and stochastic discrete and continuous time settings is discuss elsewhere (Getz et al. 2017), including embedding a continuous-time, deterministic SEIVD model in a metapopulation setting. The first phase is to construct a basic discrete-time, stochastic SEIVD model using the visual-based toolbox and then add output and input pins to this model, as illustrated in Figure 2A. The next phase is to embed the SEIVD basic model in a metapopulation setting (i.e., using a NetworkNode aggregator chip, as depicted in Figure 2B). Procedures for constructing both the first level (basic SEIVD stochastic models) and the second level (embedded metapopulation setting) in the context of continuous-time deterministic model can be found in the following two videos posted to Youtube: https://www.youtube.com/watch?v=VAGtIvq_YCo and https://www.youtube.com/watch?v=k8IEfIQCM9o. Beyond these two NMB contruction levels, a third level was constructed to run the model multiple times and produce statistics that could be used to compare different metapopulation scenarios.

3 SIMULATIONS

3.1 Focal System

purposes of illustration we apply our model to a cursory analysis of some spatial aspects of the 2014 outbreak of Ebola Viral Disease (EVD) in West Africa (Kramer et al. 2016, Getz et al. 2015). Though this epidemic has received considerable attention from the modeling and simulation community, many questions relating to the spatial aspects remained unanswered, particularly since the movement patterns of individuals among impacted subpopulations within the countries affected by this outbreak (notably Sierra Leone, Liberia, and Guinea). A deterministic, discrete-time, spatially-aggregated version of the stochastic metapopulation model developed here has been fitted to the Sierra Leone weekly incidence data with good fits of the data provided by the parameter set (Getz et al. 2017): $N_0 = 450{,}000$, $E_0 = 10$, $\gamma = 0.4$, $\kappa_0 = 1.6$ (assuming $p_\beta \equiv 1$), $\varepsilon = 2.5$ and $t_c = 17$.

We note, in fitting this model, that the interpretation of N_0 should not be the total population but the *effective population at risk*. Additionally, the best fit corresponds to estimating an initial number of exposed individuals E_0 that is greater than 1, because epidemics are not usually recognized as having started when $E_0 = 1$

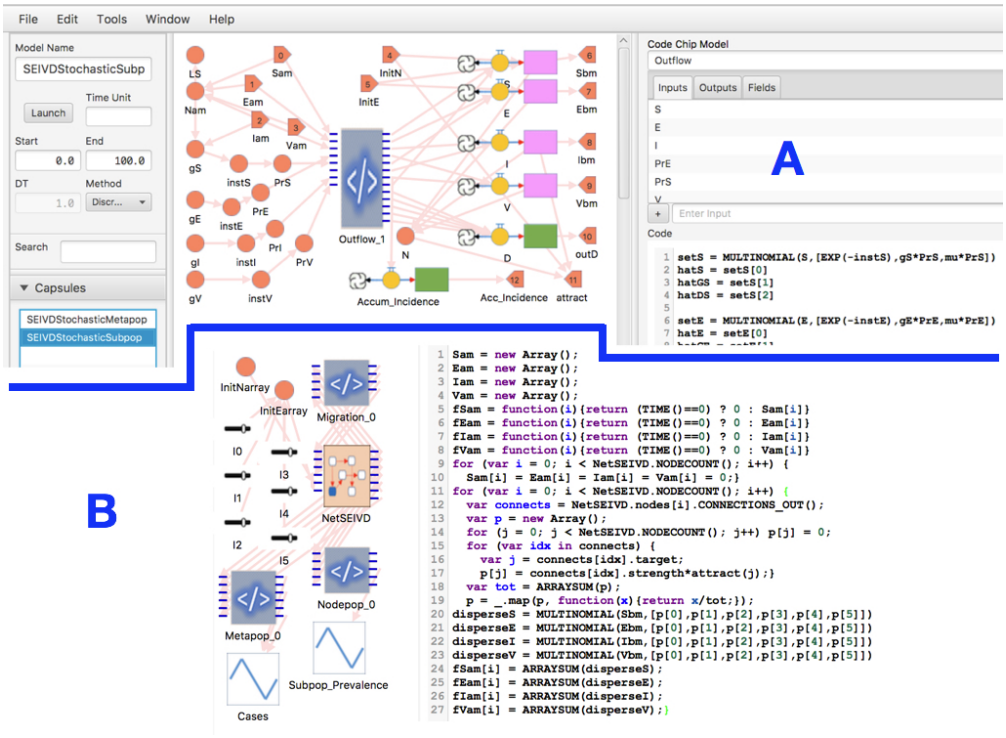

Figure 2: A NUMERUS MODEL BUILDER™ implementation of our stochastic SEIVD metapopulation model. Panel A (above the thick blue line divide) depicts the construction of a basic instance of our SEIVD model. Note the 6 input pins and 7 output pins (orange directional pentagons labeled with numbers). The right-hand partial window in Panel A shows the first 7 lines of the coded "Outflow_1" chip. Panel B depicts the next hierarchical level in which the basic SEIVD stochastic subpopulation model has been distributed through a *select and drop* action on to a NodeNetwork aggregator called a NetSEIVD chip which inherits the 6 input and 7 outpins constructed in at the lower level (Panel A). This chip is then "wired" to three other chips that perform subpopulation redistribution (Migration_0) and node aggregation functions (Metapop_0 and Node_pop). The code for the first of these is provided on the right side of Panel B.

(this is tantamount to sifting the epidemic to the left along the time axis so that $t = 0$ corresponds to some point after the actual start of the epidemic).

In extending the model to a metapopulation setting, we thus need to distribute this aggregated population at risk among the several subpopulations in the metapopulation. A refined spatial analysis of the distribution of the more then seven million people that live in Sierra Leone requires an data set well beyond what is currently available from a cursory search of the World Wide Web. A cursory search reveals that the population can be divided in to four regions, one of which is centered on the only major city in Sierra Leone (Freetown which has a million plus individuals) with the other three each containing at most two towns with between 0.1-0.2 million individuals (Figure 3). Without more refined spatial population data than this, as well as some information on how individuals move among the major population centers (i.e. towns with ten or more thousand individuals apiece) or from adjoining rural areas to these centers, a reliable analysis of the spatial aspects of an Ebola outbreak in Sierra Leone cannot be undertaken. In addition, we would also need some information on individuals moving from the neighboring states of Liberia and Guinea to account for across border viral transmission events.

Figure 3: Population centers in Sierra Leone. A crude distillation of data from www.citypopulation.de/SierraLeone-Cities and en.wikipedia.org/wiki/Districts_of_Sierra_Leone, and from links to these sites as well, suggests that the ≈ 7 million population of Sierra Leone can be roughly partitioned into a western sector (red oval: ≈ 1.5 million) centered on Freetown (≈ 1.1 million), a northern sector (blue oval: ≈ 2.5 million) centered on Makeni (≈ 0.1 million), an eastern sector (brown oval: ≈ 1.65 million) centered on Kenema (≈ 0.15 million), and a southern sector (green oval: ≈ 1.45 million) centered on Bo (≈ 0.2 million)

3.2 Comparative Study

To illustrate the potential complications of spatial structure on predicting the course of an epidemics, we compare predications made using a spatially aggregated (i.e. spatially homogeneous) SEIVD model, such as the model fitted to incidence data (i.e. using the parameters listed in the Focal System Section) from the 2014 EVD outbreak in Sierra Leone with a model that has spatial structure. Since we do not have an appropriate set of data to support a factual study of the problem, we compare the aggregated model predictions to a counter-factual disaggregation analyses in a way that illustrates critical issues that need to be considered when undertaking a factual analysis based on data that is adequate to support such an analysis (Getz et al. 2017).

For our counter-factual analysis, we posit a redistribution connection matrix $C = (c_{ij})$, $i, j = 1, ..., 6$ that reflects the metapopulation configuration illustrated in Figure 4. This matrix is

$$C = \begin{pmatrix} 1 & 1/4 & c & c & 0 & 0 \\ 1/4 & 1 & 0 & 0 & 0 & 0 \\ c & c & 1 & 1/4 & c & c \\ 0 & 0 & 1/4 & 1 & 0 & 0 \\ 0 & 0 & c & c & 1 & 1/4 \\ 0 & 0 & 0 & 0 & 1/4 & 1 \end{pmatrix} \tag{15}$$

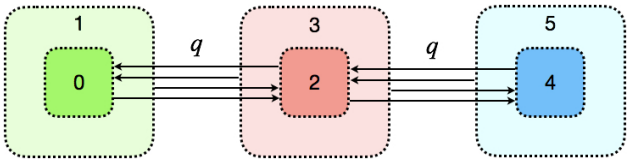

Figure 4: Metapopulation structure used in our simulations. Three core subpopulations (urban centers) indexed by $i = 0$, 2,and 4 are surrounded by peri-urban subpopulations indexed by $i = 1$, 3, and 5. Redistribution parameters are scaled so that: within each subpopulation $c_{ii} = 1$, i=0,...,5; between each urban and its surrounding peri-urban subpopulation the relevant $c_{ij} = 1/4$, and between an urban and its peri-urban are there is a redistribution with its neighbor only (i.e. centers 0 and 4 are not neighbors) scaled by a variable parameter c that ranges in our simulations between $[0, 1/4]$. For simplicity we assume the same asymmetric redistribution matrix C, given by Equation 15, applies to all four disease classes X=S, E, I and V.

where $c \in [0, 0.25]$ is a variable parameter that allows us to scale the movement between core centers from three isolated core-periphery complexes (0-1,2-3,4-5) to flows between the neighboring cores at the same rate as within a single periphery-to-core flow rate. We note that we have set up connections so that 2-3 has neighbors on either side while 0-1 and 4-5 have only the central 2-3 complex as their neighbor.

For simplicity we distribute the 450,000 individuals-at-risk that were obtained from our best-fit SEIVD model (Getz, Salter, Muellerklein, Yoon, and Tallam 2017) by placing 50,000 in each of three metapopulation core nodes (i.e., nodes 0, 2 and 4) and 100,000 in each of the three periphery nodes (i.e., 1, 3, and 5). We also rescaled κ_0 in Equation 2 so that $\kappa_0 = 2.3$ in core nodes was twice that of periphery nodes, and the expected peak prevalence in metapopulation matched that produced by the spatially aggregated model, when the epidemic in the metapopulation model was initiated in core node 0 (i.e., $E_0(0) = 10$ and $E_i(0) = 0$ for $i = 1, ..., 5$). We then compared outbreak patterns in the metapopulation for two values of c in matrix Equation 15. In particular when $c = 0.1$ the outbreak is significantly higher, but likely to stop earlier than is the case when $c = 0.25$.

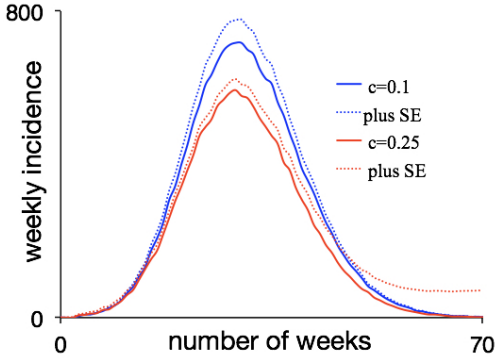

Figure 5: Comparison of expected incidence (solid lines from 20 replications) plus one standard error (dotted lines) for the cases $c = 0.1$ (blue profiles) and $c = 0.25$ (red profiles

4 CONCLUSION

Epidemics, which by definition involve only human hosts, are generally complex systems that need to account for considerable population-level spatial structure, as well as individual-level physiological and behavioral variation, when making predications about the course of disease outbreaks. On the other hand, zoonotic diseases, such as Ebola (Getz et al. 2015), Zika (Carlson et al. 2016), malaria, or anthrax (Turner et al. 2016) involve either reservoir species or animal vectors to promote or fuel outbreaks; and, hence

present additional levels of complexity that must be accounted for when making predictions. Not all of this complexity can be included in models, and the appropriate level of complexity to include depends both on an assessment of the most critical processes and the questions that need to be answered (Getz et al. 2017). Here we have demonstrated the importance of including spatial structure when modeling outbreaks of Ebola in a country such as Sierra Leone. However, it will not be possible to adequately account for this spatial structure unless good data are used on the distribution of towns and villages across the landscapes and the rates at which individuals move among these various population centers.

ACKNOWLEDGMENTS

RS work was supported in part by NSF/CPATH-2 CNS0939153 and WMG in part by the A. Starker Leopold Chair of Wildlife Ecology at UC Berkeley.

REFERENCES

Allen, L. J. 2017. "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis". *Infectious Disease Modelling*.

Andersen, P. K., R. B. Geskus, T. de Witte, and H. Putter. 2012. "Competing risks in epidemiology: possibilities and pitfalls". *International journal of epidemiology* vol. 41 (3), pp. 861–870.

Balcan, D., B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani. 2010. "Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model". *Journal of computational science* vol. 1 (3), pp. 132–145.

Britton, T. 2010. "Stochastic epidemic models: a survey". *Mathematical biosciences* vol. 225 (1), pp. 24–35.

Carlson, C. J., E. R. Dougherty, and W. Getz. 2016. "An ecological assessment of the pandemic threat of Zika virus". *PLoS neglected tropical diseases* vol. 10 (8), pp. e0004968.

Castillo-Chavez, C., H. W. Hethcote, V. Andreasen, S. A. Levin, and W. M. Liu. 1989. "Epidemiological models with age structure, proportionate mixing, and cross-immunity". *Journal of mathematical biology* vol. 27 (3), pp. 233–258.

Getz, W. M. 1996. "A hypothesis regarding the abruptness of density dependence and the growth rate of populations". *Ecology* vol. 77 (7), pp. 2014–2026.

Getz, W. M., and E. R. Dougherty. 2018. "Discrete stochastic analogs of Erlang epidemic models". *Journal of biological dynamics* vol. 12 (1), pp. 16–38.

Getz, W. M., J.-P. Gonzalez, R. Salter, J. Bangura, C. Carlson, M. Coomber, E. Dougherty, D. Kargbo, N. D. Wolfe, and N. Wauquier. 2015. "Tactics and strategies for managing Ebola outbreaks and the salience of immunization". *Computational and mathematical methods in medicine* vol. 2015.

Getz, W. M., and J. O. Lloyd-Smith. 2006. "Basic methods for modeling the invasion and spread of contagious diseases". In *Disease Evolution: Models, Concepts, and Data Analyses*, pp. 87–112.

Getz, W. M., J. O. Lloyd-Smith, P. C. Cross, S. Bar-David, P. L. Johnson, T. C. Porco, and M. S. Sánchez. 2006. "Modeling the invasion and spread of contagious diseases in heterogeneous populations". In *Disease Evolution: Models, Concepts, and Data Analyses*, pp. 113–144.

Getz, W. M., C. R. Marshall, C. J. Carlson, L. Giuggioli, S. J. Ryan, S. S. Romañach, C. Boettiger, S. D. Chamberlain, L. Larsen, P. D'Odorico et al. 2017. "Making ecological models adequate". *Ecology Letters*.

Getz, W. M., R. Salter, O. Muellerklein, H. S. Yoon, and K. Tallam. 2017. "Modeling Epidemics: A Primer and Numerus Software Implementation". *bioRxiv*.

Gilchrist, M. A., and A. Sasaki. 2002. "Modeling host–parasite coevolution: a nested approach based on mechanistic models". *Journal of Theoretical Biology* vol. 218 (3), pp. 289–308.

Gillespie, D. T. 1976. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *Journal of computational physics* vol. 22 (4), pp. 403–434.

Hethcote, H. W. 2000. "The mathematics of infectious diseases". *SIAM Review* vol. 42 (4), pp. 599–653.

Keeling, M. J. 1999. "The effects of local spatial structure on epidemiological invasions". *Proceedings of the Royal Society of London B: Biological Sciences* vol. 266 (1421), pp. 859–867.

Koelle, K., S. Cobey, B. Grenfell, and M. Pascual. 2006. "Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans". *Science* vol. 314 (5807), pp. 1898–1903.

Kramer, A. M., J. T. Pulliam, L. W. Alexander, A. W. Park, P. Rohani, and J. M. Drake. 2016. "Spatial spread of the West Africa Ebola epidemic". *Open Science* vol. 3 (8), pp. 160294.

Leclerc, P. M., A. P. Matthews, and M. L. Garenne. 2009. "Fitting the HIV epidemic in Zambia: a two-sex micro-simulation model". *PloS one* vol. 4 (5), pp. e5439.

Lloyd, A. L., and V. A. Jansen. 2004. "Spatiotemporal dynamics of epidemics: synchrony in metapopulation models". *Mathematical biosciences* vol. 188 (1), pp. 1–16.

McCallum, H., N. Barlow, and J. Hone. 2001. "How should pathogen transmission be modelled?". *Trends in ecology & evolution* vol. 16 (6), pp. 295–300.

Smith, D. L., K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie. 2012. "Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens". *PLoS pathog* vol. 8 (4), pp. e1002588.

Turner, W. C., K. L. Kausrud, W. Beyer, W. R. Easterday, Z. R. Barandongo, E. Blaschke, C. C. Cloete, J. Lazak, M. N. Van Ert, H. H. Ganz et al. 2016. "Lethal exposure: An integrated approach to pathogen transmission via environmental reservoirs". *Scientific reports* vol. 6.

Watts, D. J., R. Muhamad, D. C. Medina, and P. S. Dodds. 2005. "Multiscale, resurgent epidemics in a hierarchical metapopulation model". *Proceedings of the National Academy of Sciences of the United States of America* vol. 102 (32), pp. 11157–11162.

Wearing, H. J., P. Rohani, and M. J. Keeling. 2005. "Appropriate models for the management of infectious diseases". *PLoS medicine* vol. 2 (7), pp. e174.

AUTHOR BIOGRAPHIES

WAYNE M. GETZ is an Emeritus Starker Leopold Professor of Wildlife Ecology, as well as an Emeritus Chancellors Professor at the University of California, Berkeley. He is also Extraordinary Professor in the Mathematical Sciences at the University of KwaZulu-Natal. He received a PhD degree in Applied Mathematics from the University of the Witwatersrand, South Africa, in 1976, and a DSc in Zoology from the University of Cape Town, South Africa, in 1995. He is a Founding Partner of Numerus, Inc., which releases and maintains NUMERUS MODEL BUILDER™. Beyond his more than 300 research publications, he has coauthored a monograph on Population Harvesting and a textbook on Calculus for the Life Sciences. His email address is wgetz@berkeley.edu.

RICHARD SALTER is Emeritus Professor of Computer Science at Oberlin College. He received his PhD in Mathematics from Indiana University, USA in 1978. He is the author of more than 35 research and educational publications, and the designer of many software applications for industry and education. He is a Founding Partner of Numerus, Inc., which releases and maintains NUMERUS MODEL BUILDER™. His research has been supported by NSF and ONR. His email address is Richard.Salter@numerusinc.com.