Title
Modeling Epidemics: A Primer and Numerus Software Implementation

Authors
Getz, Wayne M
Salter, Richard
Muellerklein, Oliver
et al.

Publication Date
2017

DOI
10.1101/191601

Peer reviewed

Epidemics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

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

Review

Modeling epidemics: A primer and Numerus Model Builder implementation

Wayne M. Getz[a,b,c,*], Richard Salter[c,d], Oliver Muellerklein[a], Hyun S. Yoon[a], Krti Tallam[a]

[a] Dept. ESPM, UC Berkeley, CA 94720-3114, USA
[b] School of Mathematical Sciences, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa
[c] Numerus, 850 Iron Point Rd., Folsom, CA 95630, USA
[d] Computer Science Dept., Oberlin College, Oberlin, OH 44074, USA

ARTICLE INFO

Keywords:
SIR
SEIR models
Stochastic simulation
Dynamic networks
Compartmental models

ABSTRACT

Epidemiological models are dominated by compartmental models, of which SIR formulations are the most commonly used. These formulations can be continuous or discrete (in either the state-variable values or time), deterministic or stochastic, or spatially homogeneous or heterogeneous, the latter often embracing a network formulation. Here we review the continuous and discrete deterministic and discrete stochastic formulations of the SIR dynamical systems models, and we outline how they can be easily and rapidly constructed using Numerus Model Builder, a graphically-driven coding platform. We also demonstrate how to extend these models to a metapopulation setting using NMB network and mapping tools.

1. Introduction

In a comprehensive review of the mathematics of infectious disease, Hethcote (2000) traces the development of systems of differential equations used over the past 100 years to study disease processes. Once the purview of mathematicians, physicist and engineers, dynamical system formulations of epidemic processes are increasingly being used by epidemiologists, ecologists and social scientists to study the potential for disease pandemics to threaten the lives of humans, domesticated animals and plants, and all organisms across the globe.

Underpinning all dynamical systems models of epidemiological outbreaks and endemic disease are formulations based on the concept of an SEIR progression (Fig. 1), whereby susceptible individuals in disease class S enter disease class E on exposure to a pathogen (i.e., infected but not yet infectious themselves). Individuals in class E then transfer, after a period of latency, into the class of infectious individuals, I, only to transfer to a recovered or removed class R. Once mortality is included in the model (both natural and disease-induced) the removed designator R becomes ambiguous because removed individuals now include both recovered and dead. Thus we prefer to use the designator V for recovered (i.e., "V" for naturally vaccinated individuals that have "recovered with immunity") and D for dead. This SEIVD notation proves useful once SEIR process are elaborated to include birth, recruitment, immigration, death and emigration processes.

Beyond various demographic processes, particularly migration, as discussed further in our metapopulation formulation, elaborations to the SEIVD formulation variously include age-class (Castillo-Chavez et al., 1989) and sex (Leclerc et al., 2009) structure, spatial structure (Keeling, 1999; Balcan et al., 2010), and the genetic structure of hosts and pathogens (Koelle et al., 2006; Gilchrist and Sasaki, 2002). Additionally, stochastic formulations of epidemic models (Allen, 2017; Britton, 2010) are becoming increasingly important. These are needed to explore the inherent stochastic aspects of epidemics, such as probabilities of fadeouts (versus breakouts) in the early stages of epidemics, and risk analyses that are necessary for managing epidemics when the ultimate size and length of outbreaks are uncertain.

Hethcote (2000) provided a comprehensive review of the fundamental deterministic dynamics of SEIR models and their elaboration to include an M class (that is, individuals born with maternally provided immunity that wears off over time), and elementary characterizations of birth and death process. His results have been considerably extended to elaborated SEIR models that include more complex characterizations of births and of deaths, and some age-structure, with a strong focus in the mathematics literature on the existence and stability of outbreak and endemic equilibria (Li and Muldowney, 1995). Analyses of stochastic SEIR models and elaborations, which from hereon we refer to as SEIVD models, remain more challenging, though some analytical results do exist. Most stochastic analyses, however, are computationally intensive and the results numerical rather than analytical. In addition, unlike deterministic SEIVD models which are readily fitted to data, it remains a challenge to fit stochastic models to data, with new approaches involving concepts well beyond first courses in calculus or linear algebra.

Given that a susceptible/infected class structure underpins all

* Corresponding author at: Dept. ESPM, UC Berkeley, CA 94720-3114, USA.
E-mail address: wgetz@berkeley.edu (W.M. Getz).

https://doi.org/10.1016/j.epidem.2018.06.001
Received 20 September 2017; Received in revised form 30 May 2018; Accepted 1 June 2018

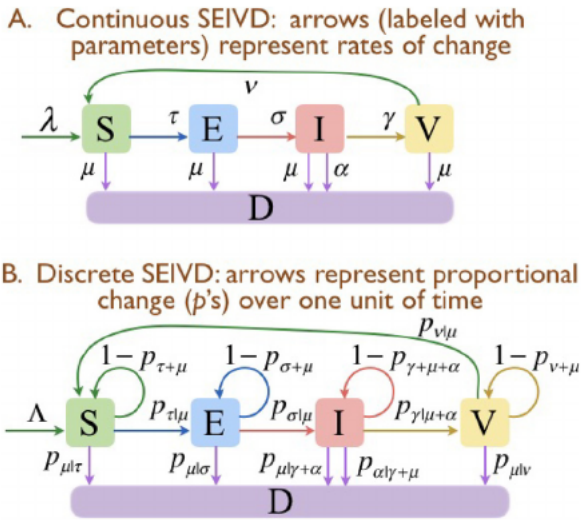

Fig. 1. Flow diagrams for the basic SEIVD continuous (A) and discrete (B) time models with transition rates τ, σ, γ and ν from disease classes S to E, E to I, I to V and V back to S, respectively. In the continuous-time formulation, λ is the rate at which new individuals are recruited to the susceptible population (births or emigration), μ and α are natural and disease-induced mortality rates respectively. In the discrete-time formulation, a competing rates approach is used to derive transition proportions p, with identifying subscripts, as described in the text. We note that Λ is the recruitment rate $\lambda(t)$ integrated over one discrete time unit to obtain the number of individuals recruited each step (see Eq. (11)). We also note that τ is the well-known "force of infection," which should not be confused with the recruitment rate λ, since λ is often used in many presentations to represent the force of infection.

epidemiological models, whether deterministic, stochastic, or even agent-based, a succinct, pedagogical review of such models is useful. In particular a clear exposition of SEIVD models for the non-mathematician—by which we mean, scientists who have some understanding of calculus, but do not have formal training in dynamical systems theory, or the numerical techniques to competently build and implement computational models. In addition, the literature lacks expository articles that elucidate for biologists and social scientists the relationship among continuous and discrete SEIVD models (but see Getz and Lloyd-Smith, 2006), their stochastic elaborations in systems and agent-based (i.e., individual-based) computational settings (Ajelli et al., 2010; Burke et al., 2006), as well as extensions to metapopulation settings (Lloyd and Jansen, 2004). These are lacunae, apart from agent-based models, that we hope to fill with this paper, while at the same time providing those scientists who are looking for fast, reliable ways to obtain and modify code needed to address their epidemiological models with a means to do so in the context of the Numerus Model Builder software development platform.

At the end of this paper, we provide a list of links to online material that are available to readers to facilitate use of the models described in this primer. This includes a link to the Numerus Model Builder (NMB) website where the reader can download a free version of the software that will run the models made available with this primer. The website also links to an instruction wiki and a number of training videos that the user can view to start using NMB. These include: (1) Introduction to the Numerus Model Builder Interface, (2) Simple Population Model, (3) Logistic Population, and (4) Discrete Density-Dependent Growth. In addition, we provide links to the six epidemiological models used in this primer and to nine videos that discuss aspects of constructing and implementing these models using NMB.

Finally, given the centrality of dynamic epidemic models to containment of outbreaks, policy formulation and response logistics, modeling tools are needed that can be used by healthcare professionals not trained in computational methods to carry out containment policy, and response analyses. Thus, our strong focus is on how to apply NMB to building basic epidemiological models and to demonstrate its use in the context of epidemics that are spatially structured, such as the recent outbreak of Ebola in West Africa (Kramer et al., 2016; Getz et al., 2015).

2. Homogeneous SEIVD formulations

2.1. Continuous deterministic models

SEIVD infectious disease models are based on dividing an otherwise homogeneous population into the following disease classes: susceptible (S), exposed (E; infected but not yet infectious), infectious (I), recovered with immunity (V; which may wane over time) and dead (D; from both natural and disease-induced mortality). Throughout, we use the roman fonts S, E, I, V and D to name the classes themselves and the italic fonts S, E, I, V and D to refer to the variables representing the number of individuals in these corresponding classes. The assumption of homogeneity implies that age and sex structure are ignored. We incorporate population spatial structure—as would be found in countries comprising of a network of cities, towns, and villages—into a metapopulation framework (Fulford et al., 2002; Lloyd and Jansen, 2004), if we assume that a set of homogeneous subpopulations can be organized into a network of subpopulations, among which individuals move in a fashion that reflects appropriate movement rates (e.g., propensity to move as a function of age and sex (Getz et al., 2017)) and geographical factors (e.g., distances, geographical barriers, desirability of possible destinations).

If the time scales of the epidemic and movement processes among subpopulations, including disease-induced mortality, are much faster than the time scale of the background population demography (births, recruitment, natural mortality and population level migration) then we can ignore the demography; otherwise we cannot. For example, in the case of influenza, epidemiological and local movement processes involve noticeable changes at the scale of weeks, while demographic changes in the underlying population itself (beyond epidemic disease-induced death rates) are obvious only at the scale of years. In this case, we can ignore natural births and deaths, and focus on epidemic processes alone.

In the context of an epidemic occurring in a single homogeneous population of size N at the start of the epidemic (i.e., at time $t = 0$), denote the per-capita susceptible (S) disease transmission rate (i.e., force of infection) by $\tau(I, N)$, which we assume depends on both the number of infectious individuals $I(t)$ and the total number of individuals $N(t)$ in the population. The form of this dependence is discussed in more detail later. In addition, we denote per-capita rates of progression from exposed (E) to infectious (I) and onto to removed with immunity (V) using the symbols σ and γ respectively (Fig. 1). For generality, as depicted in Fig. 1A, we include a population net recruitment function $\lambda(t)$, where all these recruits are assumed to be susceptible. Later we generalize this in the context of a metapopulation structure and allow other disease classes to migrate. We also include per-capita disease-induced and natural mortality rates α and μ, respectively, that accumulate in disease class D, as well as allow for the occurrence of a per-capita immunity-waning rate ν (Fig. 1). Some or all of this latter group of parameters may be zero, and only become non-zero as the scope of the analysis undertaken is enlarged. Also, as a starting point, all parameters are assumed to be constant, except for the transmission function τ, which at its most fundamental has the relatively simple "frequency-dependent" structure (Getz and Pickering, 1983)

$$\tau(I, N) = \frac{\beta I}{N}, \quad \text{where } N = S + E + I + V \tag{1}$$

with the disease transmission parameter β itself constant over time. The even simpler "density-dependent" form $\tau(I, N) = \beta I$ should be used with caution, because at high population densities it is unduly unrealistic; although at low population density it can be quite useful to generalize

$\tau(I, N)$ as being approximately density-dependent (e.g. see McCallum et al., 2001 and the Discussion section below).

With the above notation, the basic continuous time differential equation formulation of an SEIVD epidemic process in a homogeneous population takes the form:

$$\begin{aligned}\frac{dS}{dt} &= \lambda(t) + \nu V - \big(\tau(I, N) + \mu\big)S \\ \frac{dE}{dt} &= \tau(I, N)S - (\sigma + \mu)E \\ \frac{dI}{dt} &= \sigma E - (\gamma + \alpha + \mu)I \\ \frac{dV}{dt} &= \gamma I - (\nu + \mu)V\end{aligned} \tag{2}$$

To complete the description, we need to include the relationship

$$N(t) = S(t) + E(t) + I(t) + V(t) \tag{3}$$

We may also want to evaluate the number of deaths $M(t)$ and $D(t)$, due respectively to natural and disease-induced causes that have accumulated over the period $[t-1, t)$ using the equations

$$\begin{aligned}M(t) &= \int_{t-1}^{t} \mu N(z)dz \\ D(t) &= \int_{t-1}^{t} \alpha I(z)dz\end{aligned} \tag{4}$$

The properties of Eq. (2) have been extensively studied over the past three to four decades (Li and Muldowney, 1995; Hethcote, 2000; Li and Wang, 2002; Getz and Lloyd-Smith, 2006), with the most important results pertaining to both pathogen-invasion (i.e., disease-outbreak) and pathogen-persistence (i.e., endemicity) conditions. This is informally derived here for the case $\tau = \beta I/N$, $\lambda(t) = \mu N$ (i.e., individual birth and death rates are the same) and $\nu = 0$ from the following considerations. Each infectious individual infects susceptible individuals at a rate $\beta S/N$ ($\approx\beta$ when $S \approx N$) over an infectious period that lasts on average for a time $1/(\gamma + \alpha + \mu)$. However, only a proportion $\sigma/(\sigma + \mu)$ of infected individuals become infectious, due to natural and disease-induced mortality rates while in state E. Thus the basic reproductive number R_0, which is the number of susceptible individuals that each infectious individual is expected to infect at the onset of an epidemic, is given by

$$R_0 = \frac{\beta\sigma}{(\sigma + \mu)(\gamma + \mu + \alpha)} \tag{5}$$

This derivation for the SEIVD model, in the context of density-dependent transmission, can be found in Li and Wang (2002): it uses the the so-called "next generation matrix" method (Heffernan et al., 2005; Van den Driessche and Watmough, 2002; Diekmann and Heesterbeek, 2000) to compute this result. Because an outbreak cannot occur unless $R_0 > 1$, Eq. (5) in turn implies that under outbreak conditions, the following inequality applies:

$$\underline{\text{Outbreak threshold}}: \quad \beta > \frac{(\sigma + \mu)(\gamma + \mu + \alpha)}{\sigma} \tag{6}$$

2.2. Numerus SEIVD continuous-time implementation

As an introduction to using Numerus Model Builder to code dynamical systems models, we begin with the very simple population growth model (known as the logistic model)

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad N(0) = N_0 \tag{7}$$

This equation can be thought of as a special case of Eqs. (1) and (2) when $S(0) = N_0$, $E(0) = I(0) = R(0) = 0$, and $\lambda(t) = rN(t)(1 - N(t)/K)$: under these conditions, $E(t)$ and $I(t)$ remain zero and $S(t) \equiv N(t)$ for all $t \geq 0$.

In Video 1 at the supporting website, the reader can find a complete construction of the logistic Eq. (7) using Numerus Model Builder, with solutions generated for various values of r when $K = 1$. We note that there is no loss of generality in setting $K = 1$ because it can be seen to be a scaling constant that varies with the measurement units selected for N. (Verify this by making the transformation $x = N/K$ in Eq. (7).)

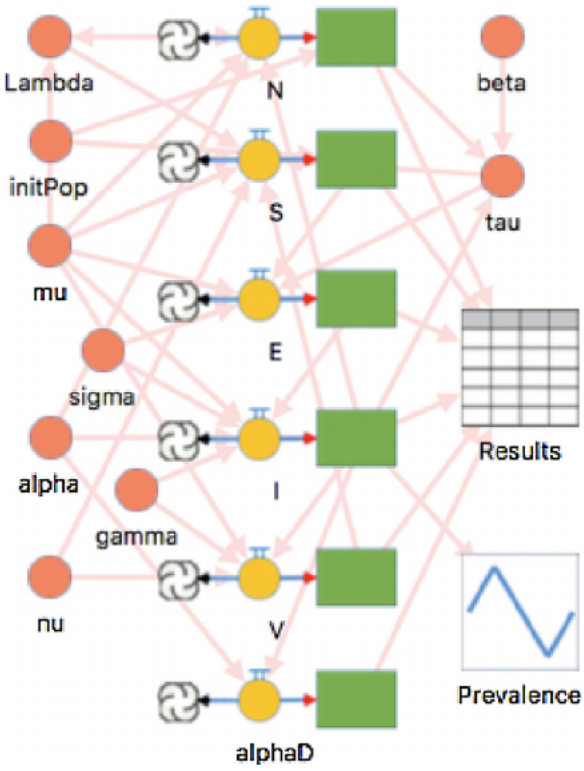

Fig. 2. A Numerus Model Builder representation of Eqs. (2)–(4), showing the dynamic variables as green boxes with orange circle icons that represent the differential equations of the system, and pink circles to represent input parameters or terms such as $\lambda = \mu N$ or $\tau = \beta I/N$. The grid represents an output table, and the blue in-square zig-zag a graphing tool. See Video 2 at the supporting website for more details.

Using Numerus Model Builder to code up Eqs. (2)–(4), we obtain the model illustrated in Fig. 2. We then run the model for the case $\lambda(t) = \mu N$ and $\tau = \beta I/N$ to explore the effect of β on solutions to Eq. (2) as it increases from a value below the outbreak threshold (i.e. $R_0 < 1$) to one above the outbreak threshold (i.e. $R_0 > 1$), as embodied in Inequality (6). In particular, for the parameters listed in the caption to Fig. 3, it follows from Eq. (5) that the threshold occurring at $R_0 = 1$ implies that β satisfies:

$$1 = \frac{0.3\beta}{(0.3 + 0.01)(0.3 + 0.01 + 0.05)} \Rightarrow \beta = 0.372$$

The impact of β increasing from 0.35 to 0.40 is illustrated in Fig. 3A. In this panel, we see that after an initial drop related to the fact that individuals entering E must first transition to I before they can begin to infect individuals in S, the solution is declining for $\beta = 0.37$, but ultimately growing for $\beta = 0.38$. In addition, as illustrated in Fig. 3B for the case $\nu > 0$, the effect of recycling individuals from the R class back into the S class results a slightly higher peak-epidemic level. More importantly, the rebound associated with the case $\nu = 0$ has been greatly reduced and the prevalence remains higher through the waning phase of the epidemic.

2.3. Discrete deterministic models

Discrete-time models, as represented by systems of difference equations, are computationally more efficient than continuous-time differential equation models, such as Eq. (2), because discrete time models do not require numerically intensive integration. Further,

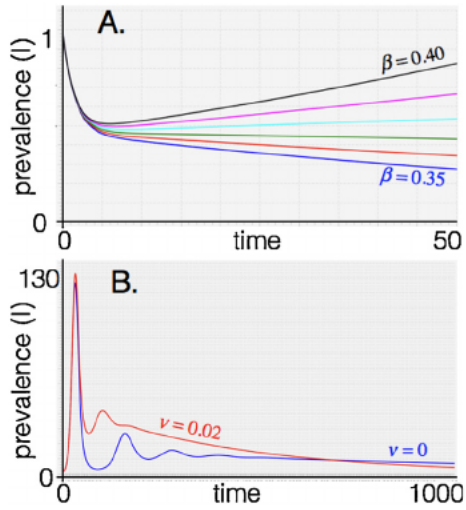

Fig. 3. The prevalence $I(t)$ is plotted over time (A – t [0, 50], B – t [0, 1000]) using the Numerus Model Builder implementation, depicted in Fig. 2, to generate numerical solutions to Eq. (2) under initial conditions $S(0) = 999$, $E(0) = 0$, $I(0) = 1$, and $V = 0$ for the case $\mu = 0.01$, $\alpha = 0.05$, $\sigma = \gamma = 0.3$, under the assumption that $\lambda(t) = \mu N(t)$. In addition: in (A) $\nu = 0$ and β varies from 0.35 to 0.40 (in steps of 0.01); and in (B) $\beta = 1$ and $\nu = 0.02$ (red) and $\nu = 0$ (blue). See Video 3 at the supporting website for details on making sets of batch runs using Numerus Model Builder.

discrete models synchronize directly with periodically collected data. These data may be: daily or weekly incidence rates in fast moving epidemics, such as influenza, SARS, or Ebola; or monthly or annual rates in slower moving epidemics such as HIV or TB. In long running epidemics that have a seasonal component, such as TB (Fares, 2011), the effects of seasonality can only be estimated from a model if incidence rates are reported monthly or, at least, quarterly.

Discrete models present a sequencing conundrum regarding mutually exclusive events (e.g., one cannot die from both disease and natural causes in the same time step). For example, consider outflow from the infectious class over time interval $(t, t+1]$, as modeled in the third equation in Eq. (2). If there are $I(t)$ individuals in class I at time t then, assuming no inflow, the total number of individuals still in class I at time $t+1$ is obtained by integrating the equation

$$\frac{\mathrm{dI}}{\mathrm{dt}} = -(\gamma + \alpha + \mu)I, \qquad I(t)\,\text{specified}$$

over $(t, t+1]$, i.e., over one unit of time, to obtain

$$I(t+1) = I(t)e^{-(\gamma+\alpha+\mu)} \tag{8}$$

Hence the proportion of individuals that leave the infectious class over time $(t, t+1]$ due to recovery at rate γ, dying from disease at a rate α, and dying from natural causes at a rate μ is

$$p_{\gamma+\alpha+\mu} = 1 - \frac{I(t+1)}{I(t)} = (1 - e^{-(\gamma+\alpha+\mu)}) \tag{9}$$

The way to allocate the proportions $p_{\gamma|\alpha+\mu}$, $p_{\alpha|\gamma+\mu}$, and $p_{\mu|\gamma+\alpha}$ of individuals leaving class I into those that respectively recover, die from disease, and die from natural causes, using the "competing rates" formulation, is (Andersen et al., 2012) (the symbol $:=$ indicates that these are definitions)

$$\begin{aligned} p_{\gamma|\alpha+\mu} &:= \frac{\gamma(1-e^{-(\gamma+\alpha+\mu)})}{\gamma+\alpha+\mu} \\ p_{\alpha|\gamma+\mu} &:= \frac{\alpha(1-e^{-(\gamma+\alpha+\mu)})}{\gamma+\alpha+\mu} \end{aligned} \tag{10}$$

From this, and Eq. (9), it easily follows that

$$p_{\gamma|\alpha+\mu} + p_{\alpha|\gamma+\mu} + p_{\mu|\gamma+\alpha} = p_{\gamma+\alpha+\mu}$$

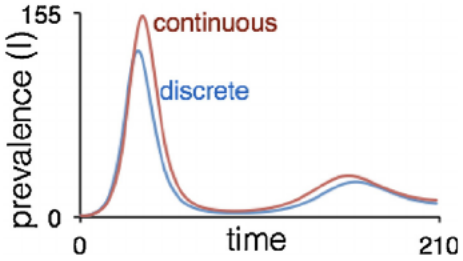

Fig. 4. The prevalence $I(t)$ is plotted over time t [0, 210] for the continuous (red) and discrete (blue) Numerus Model Builder coding of SEIVD models represented by Eqs. (2) and (12) respectively for the case $\tau(t) = \beta S(t)/N(t)$ and $\lambda(t) = \mu N(t)$ using the parameter values $\beta = 1$, $\alpha = 0.05$, $\sigma = \gamma = 0.3$, and $\nu = 0$. In the discrete model we note that $\Lambda_t = M(t)$ and $\tau(t)$ is assumed constant over $[t, t+1)$. The numerical solutions depicted here correspond to initial conditions $S(0) = 999$, $E(0) = 0$, $I(0) = 1$, and $V = 0$. See Video 4 at the supporting website for additional details on building the discrete model using Numerus Model Builder.

Using this notation in the context of the appropriate rates for each equation and making the assumption that

$$\Lambda_t = \int_t^{t+1} \lambda(s)\mathrm{ds} \quad \text{and} \quad \tau_t = \tau(I(t), N(t)) \tag{11}$$

are constants that apply over time interval $(t, t+1]$, a discrete equivalent of Eq. (2) takes the form

$$\begin{aligned} S(t+1) &= \Lambda_t + p_{\nu|\mu} V(t) + S(t)(1 - p_{\tau_t+\mu}) \\ E(t+1) &= p_{\tau_t|\mu} S(t) + E(t)(1 - p_{\sigma+\mu}) \\ I(t+1) &= p_{\sigma|\mu} E(t) + I(t)(1 - p_{\gamma+\alpha+\mu}) \end{aligned} \tag{12}$$

We note that solutions to this discrete system will differ from solutions to the continuous system, even for constant Λ_t, because $\tau(t)$ does not remain constant over the interval $(t, t+1]$ in the continuous model (Fig. 4). However, there is no *a priori* reason to favor a continuous over a discrete time formulation because what is regarded as an event at one time scale is actually a process at another time scale (e.g., see Getz and Schreiber, 1999 in the context of consumer–resource processes). Further discrete-time formulations are particularly convenient when the data come in discrete packets (e.g., disease class transitions or the occurrence of deaths are recorded daily or weekly, etc.) rather than a collection of events marked according to the flow of continuous time. Finally, stochastic approaches are needed to capture the full richness of epidemiological dynamics (Allen, 2017; Britton, 2010; Getz et al., 2015) and, as discussed below, it is far easier to make discrete-time models stochastic than continuous-time models.

To complete this discrete model, we have the discrete analogues of the continuous live and dead population variables:

$$\begin{aligned} N(t) &= S(t) + E(t) + I(t) + V(t) \\ D(t) &= p_{\alpha|\gamma+\mu} I(t) \\ M(t) &= M^S(t) + M^E(t) + M^I(t) + M^V(t) \end{aligned} \tag{13}$$

where

$$\begin{aligned} M^S(t) &= p_{\mu|\tau_t} S(t) \\ M^E(t) &= p_{\mu|\sigma} E(t) \\ M^I(t) &= p_{\mu|\gamma+\alpha} I(t) \\ M^V(t) &= p_{\mu|\nu} V(t) \end{aligned} \tag{14}$$

2.4. Discrete stochastic SEIVD models

Before presenting a stochastic formulation of Eqs. (12)–(14), it is worth noting that one can simulate continuous systems models, such as Eqs. (1)–(4), as a stochastic process of randomly occurring events using Gillespie's algorithm (Gillespie, 1976, 1977) and its refinements

(Gibson and Bruck, 2000; Press et al., 2007; Vestergaard and Génois, 2015). This general, event-oriented approach, however, involves considerably more computations invoking numerical integration schemes than working directly with discrete models. Further, as we stressed earlier on, a continuous-time model is theoretically no more privileged than its analogous discretized formulation that has its iteration interval synchronized with the frequency at which the data are collected.

In developing a stochastic formulation we use the notation

$$\widetilde{X} := \text{BINOMIAL}[n, p]$$

to denote that $\widetilde{X}$ is one drawing of a binomial variable representing the number of times one of two outcomes occurs in n independent trials, when the probability of this outcome occurring in a single trial is p (i.e., a Bernoulli process with probability p). More generally, we use the notation $(\widetilde{X}_1, \ldots, \widetilde{X}_r)$ to denote one instance or one particular drawing of $(x_1, \ldots, x_r) \sim \text{MULTINOMIAL}[n; p_1, \ldots, p_r]$, where $\widetilde{X}_i$ is the number of times one of r possible outcomes occurs over n trials, each have probability p_i $(i = 1, \ldots, r)$ of occurring in any one trial.

With this notation, we can write down equations for the stochastic equivalent of the discrete deterministic model represented by Eq. (12). We use the additional notation $\widetilde{U}$ with appropriate designator subscripts to denote the number of individuals transferring between disease classes:

$$\begin{aligned}
\left(\widetilde{S}(t), \widetilde{U}^{\mathrm{S}}(t), \widetilde{M}^{\mathrm{S}}(t)\right) &:= \text{MULTINOMIAL}[S(t); 1 - p_{\tilde{\eta}+\mu}, p_{\tilde{\eta}|\mu}, p_{\mu|\tau_t}] \\
\left(\widetilde{E}(t), \widetilde{U}^{\mathrm{E}}(t), \widetilde{M}^{\mathrm{E}}(t)\right) &:= \text{MULTINOMIAL}[E(t); 1 - p_{\sigma+\mu}, p_{\sigma|\mu}, p_{\mu|\sigma}] \\
\left(\widetilde{I}(t), \widetilde{U}^{\mathrm{I}}(t), \widetilde{M}^{\mathrm{I}}(t), \widetilde{D}(t)\right) &:= \text{MULTINOMIAL}[I(t); 1 - p_{\gamma+\alpha+\mu}, p_{\mu|\alpha+\mu}, p_{\gamma|\alpha+\mu}, p_{\alpha|\gamma+\mu}] \\
\left(\widetilde{V}(t), \widetilde{U}^{\mathrm{V}}(t), \widetilde{M}^{\mathrm{V}}(t)\right) &:= \text{MULTINOMIAL}[V(t); 1 - p_{\nu+\mu}, p_{\mu|\nu}, p_{\nu|\mu}]
\end{aligned} \tag{15}$$

The stochastic version of our discrete SEIVD model is thus represented by the following equations:

$$\begin{aligned}
S(t+1) &= \widetilde{S}(t) + \widetilde{\Lambda}_t + \widetilde{U}^{\mathrm{V}}(t) \\
E(t+1) &= \widetilde{E}(t) + \widetilde{U}^{\mathrm{S}}(t) \\
I(t+1) &= \widetilde{I}(t) + \widetilde{U}^{\mathrm{E}}(t) \\
V(t+1) &= \widetilde{V}(t) + \widetilde{U}^{\mathrm{I}}(t)
\end{aligned} \tag{16}$$

where $\widetilde{\Lambda}_t$ are generated from an appropriate discrete distribution, such as a Poisson distribution with expected value Λ_t determined from local population birth rates or other relevant recruitment processes. We note that the term $\widetilde{U}^{\mathrm{S}}(t)$ is the proportion derived for the transmission-driven transition τ_t, and hence typically depends on the variable number of infectious individuals in the population at the start of each time step, while the other transitions $\widetilde{U}^{\mathrm{X}}(t)$, X = E, I and R, are proportions derived from typically constant flow rates σ, γ and ν.

To complete this stochastic discrete model represented by Eqs. (15) and (16), we have

$$\widetilde{M}(t) = \widetilde{M}^{\mathrm{S}}(t) + \widetilde{M}^{\mathrm{E}}(t) + \widetilde{M}^{\mathrm{I}}(t) + \widetilde{M}^{\mathrm{V}}(t) \tag{17}$$

as well as $N(t)$ in Eq. (13), which is needed to calculate $\tau_t = \tau(I(t), N(t))$ using an appropriate expression, such as given in Eq. (1) (Fig. 5).

2.5. Weakness of the SEIVD formulation

Real epidemics are far more complicated than the idealized epidemics encapsulated in the above SEIVD models. Among assumptions in SEIVD models used to keep them relatively simple are:

1 *The assumption of host homogeneity.* This assumption is tenuous at best: the host's age (Klepac and Caswell, 2011), sex (Bellan et al., 2013), genetic makeup (particularly MHC locus genes) (Hill et al., 1991), physiological state (Lyte, 2004), and history of exposure to the current and related pathogens (the latter due to cross-immunity issues (Kurts et al., 2010)) all play a role in affecting the vulnerability of the host to infection, the length of time the host is infectious, and the risk of the host dying from disease. For a variety of reasons that include behavior, physiology, and genetics, some individuals are also much more infectious than others. These individuals are sometimes referred to as superspreaders and it is well-known and that in some epidemics fewer than 20% of infected individuals may be responsible for more than 80% of transmission events (Lloyd-Smith et al., 2005b).

2 *The assumption of a well-mixed population.* This is related to the assumption that hosts contact one another at random. Contact is never random. At best, contact can be assumed to be locally random. This implies that the probability individuals contact one another over some future period is inversely related to their current distance from one another. One way around this assumption is to extend SEIVD models to a metapopulation setting in which subpopulations are regarded as well-mixed and rates of exchange of individuals among subpopulations is some inverse function of the distance among the centers of these subpopulations, as discussed in Section 3 below.

3 *The assumption that the transmission rate per susceptible individual is a relatively simple function of host population and infectious class densities (or numbers).* For example, Eq. (1) assumes that total transmission has a frequency dependent form. A more general case that assumes transmission is essentially (i) density dependent when population density N is small relative to some function-location parameter L, and (ii) frequency dependent when the population is much larger than L, takes the form

$$\tau(I, N) = \frac{\beta I}{N + L}$$

More complicated functions have been proposed (McCallum et al., 2001; Getz and Lloyd-Smith, 2006), including a negative binomial expression that accounts for susceptible host aggregation (Merl et al., 2009) or the phenomenon of superspreaders (Lloyd-Smith et al., 2005b), however, they are also limited in adhering to the following assumption.

4 *The assumption that individuals exit disease states following exponential (continuous time) or geometric (discrete time) distributions.* We see this clearly in Eq. (8) in the context of infectious individuals in the continuous time model over a single time period, which leads to the geometric rate of decay when applied iteratively over several time periods. This rather severe assumption (which implies that the highest exit proportions occur closest to entry into the disease class—or put another way, the mode of the exit distribution is at the point of entry into the given disease class) can be obviated using a box-car model or distributed-delay approach. In these formulations, infected individuals pass through disease class E by passing through a sequence of disease subclasses $E_1, E_2, \ldots, E_r$, before passing into the disease subclass sequence $I_1, I_2, \ldots, I_k$, and then finally into disease class V (Clancy, 2014; Leclerc et al., 2014; Blythe and Anderson, 1988; Bame et al., 2008; Anderson and Watson, 1980). In this case, the exit distributions from E and I are no longer exponential, but are now Erlang (i.e., a subclass of the Gamma distribution). Further, the mode of the Erlang distribution becomes increasingly peaked and approaches the mean as the number of subclasses increases. This assumption can also be obviated by using a discrete time model that tracks the number of days each individual has been in a particular disease state, as seen for example in a model of the 2003 Asian outbreak of SARS (Lloyd-Smith et al., 2003).

5 *The assumption that the transmission rate parameter is time independent.* Apart from seasonal considerations, it is often assumed the

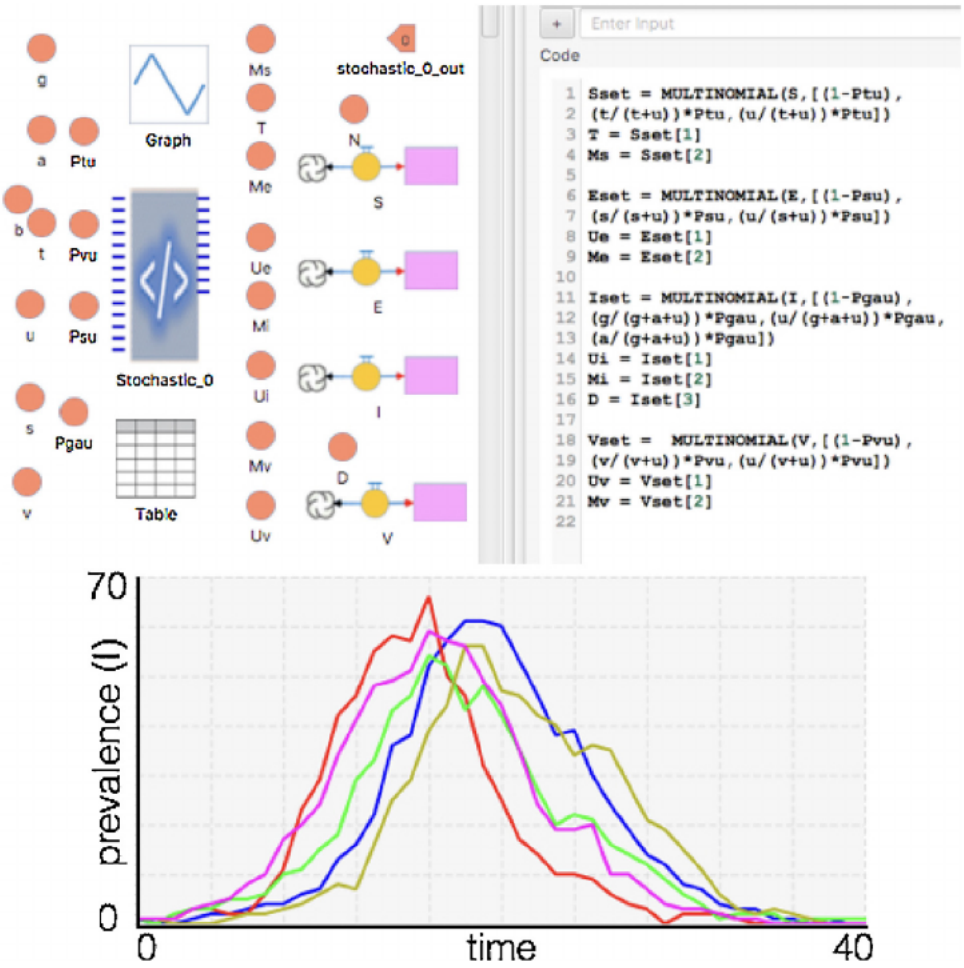

Fig. 5. The top left panel provides a Numerus Model Builder representation of the discrete stochastic model with the stochastic chip (Stochastic_0) at its center. This chip contains the code (right top panel) for generating the multinomial distributions seen in Eq. (15). The bottom panel illustrates plots of prevalence from five repeated runs of the stochastic model using the identical set of parameters in each run. See Video 5 at the supporting website for more details.

transmission parameter β in assumption 3 is constant. The primary reasons why epidemics subside, however, are that either the proportion of susceptible individuals in the population is reduced to the point where the epidemic can no longer be sustained (so-called threshold effect (Getz and Pickering, 1983; Lloyd-Smith et al., 2005a)) or the rate at which susceptible individuals contact infectious individual during the course of an epidemic, as in the recent Ebola outbreak in West Africa (Getz et al., 2015; Camacho et al., 2014), precipitously falls due to behavioral reasons as the epidemic proceeds. One approach is to assume that β has the exponential form $\beta(t) = \beta_0 e^{-\varepsilon t}$ (e.g. as in Althaus, 2014). This is a little extreme because we should not expect β not to start to decline precipitously at the start of the epidemic, but only part way into the epidemic, once public awareness of the full potential of the epidemic has become apparent. In this case, a mirror-image, s-shaped curve of the form

$$\beta(t) = \frac{\beta_0}{1 + \left(\frac{t}{t_c}\right)^{\varepsilon}} \tag{18}$$

for parameters $t_c > 0$ and $\varepsilon > 1$ is more appropriate (in a manner analogous to the onset of density-dependence, as discussed in Getz (1996)).

6 *The assumption that pathogen dose can be ignored.* The human immune system is extremely complex and takes a variable amount of time to gear up once invaded by a replicating army of pathogens, as the gear-up time depends on the condition of the host, host genetics, and prior host experience with the same and other pathogens. Small pathogen armies (i.e., low doses) are more easily contained by the hosts immune system—that is, before they can replicate to reach levels that may overwhelm and kill the host—than high doses or repeated exposure to lower doses over a short window of time. Such host-immune-system/pathogen dynamics can only be understood using models that are often more complicated than the SEIVD model itself (Perelson and Weisbuch, 1997; Bauer et al., 2009). Further, ignoring both single and repeated dose effects may severely compromise the reliability and transferability of SEIVD models fitted to one population and then applied to another population or even to the same population at a later date.

3. Metapopulation formulation

The first step in extending homogeneous SEIVD models to a metapopulation setting is to prepare the homogeneous models by embedding them in a background population through the addition of migration processes (Fig. 6A). For example, for the *ith* subpopulation we can add local per-capita emigration rates q_i^X to Eq. (2) for X = S, E, I, and R, to account for individuals that move in and out of the four different disease classes. If we do this for all subpopulations using a set of subpopulation emigration rates q_i^X, $i = 1, \ldots, m$, we can then also generate a set of immigration rates h_i^X that conserves total movement numbers within the metapopulation. In this case, we need to define a movement matrix with elements π_{ij}^X such that

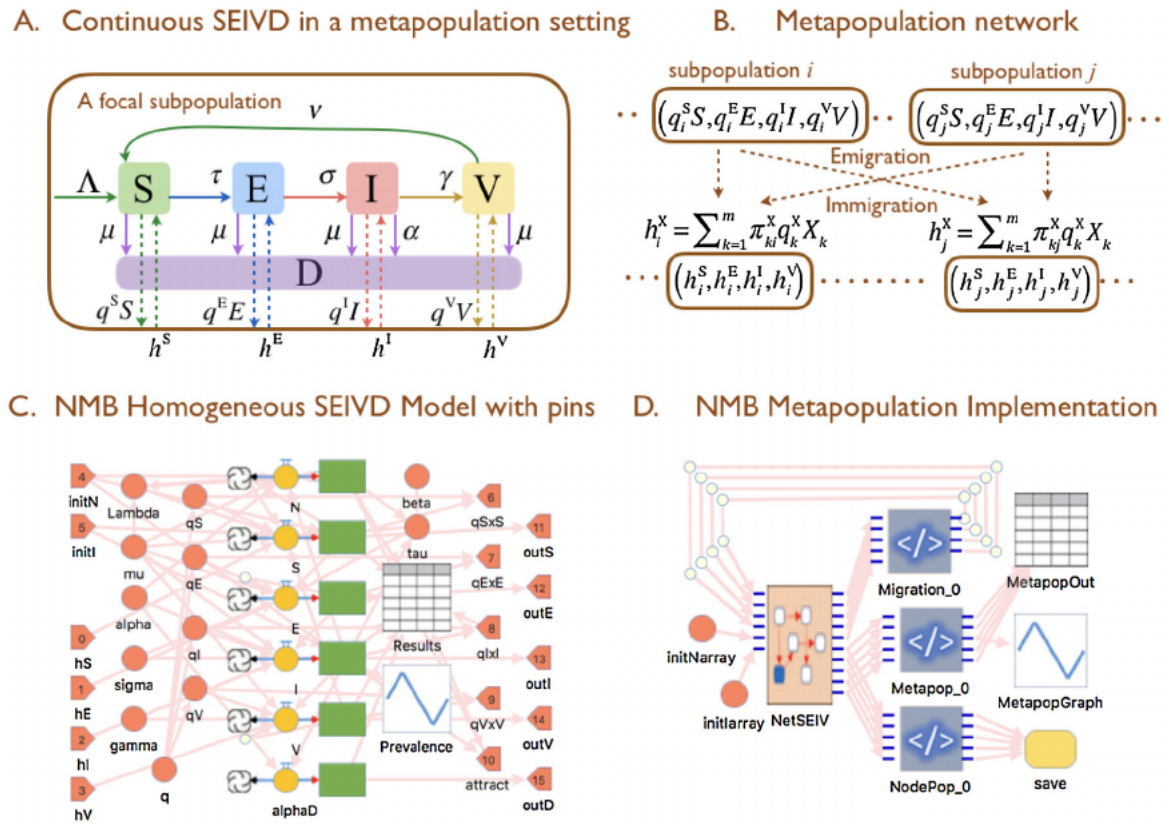

Fig. 6. (A) A homogeneous subpopulation with input and output flows of individuals to and from other subpopulations in the metapopulation (cf. Fig. 1A). (B) The flow network quantities and governing equation for the metapopulation as a whole. (C) The Numerus Model Builder (NMB) submodel of the subpopulation processes depicted in A, which is basically the NMB model illustrated in Fig. 2 with input and output pins added as described in Video 6 at the supporting website. (D) The NMB metapopulation model formulated one hierarchical level above the subpopulation model depicted in C, with the use of the NetSEIV, Migration and Metapop codechips explained in Video 7 at the supporting website. The NodePop codechip connected to the yellow "save" event container allows the one-time event of saving trajectories of all variables from all nodes at the end of the simulation.

$$h_i^{\mathrm{X}}(t) = \sum_{j=1}^{m} \pi_{\mathrm{ji}}^{\mathrm{X}} q_j^{\mathrm{X}} X_j(t), \tag{19}$$

where $\sum_{j=1}^{m} \pi_{\mathrm{ij}}^{\mathrm{X}} = 1$ for $i = 1, \ldots, m$ and X = S, E, I, and V. We can then add this migration process to Eq. (2), applied to the ith subpopulation in the metapopulation network, to obtain:

$$\begin{aligned}
\frac{\mathrm{dS}i}{\mathrm{dt}} &= h_i^{\mathrm{S}} + \lambda_i(t) + \nu_i R_i \\
&\quad - (\tau_i(I_i, N_i) + \mu_i + q_i^{\mathrm{S}}) S_i \\
\frac{\mathrm{dE}i}{\mathrm{dt}} &= h_i^{\mathrm{E}} + \tau_i(I_i, N_i) S_i - (\sigma_i + \mu_i + q_i^{\mathrm{E}}) E_i \\
\frac{\mathrm{dI}i}{\mathrm{dt}} &= h_i^{\mathrm{I}} + \sigma_i E_i - (\gamma_i + \alpha_i + \mu_i + q_i^{\mathrm{I}}) I_i \\
\frac{\mathrm{dV}i}{\mathrm{dt}} &= h_i^{\mathrm{V}} + \gamma_i I_i - (\nu_i + \mu_i + q_i^{\mathrm{V}}) V_i
\end{aligned} \tag{20}$$

We can further assume that the movement elements $\pi_{ji}(t)$ are derived from a set of connectivity strengths $\kappa_{ji}(t)$ that reflect the relative ease with which an individual in subpopulation j can move (in this case flow because time is continuous) to subpopulation i over the time interval $[t, t + 1)$ and a set of relative attractivity values $a_i(t)$ that are characteristics of the nodes i. These attractivity values $a_i(t)$ are assumed to bias the movement of any individual leaving subpopulation j to move to subpopulation i with probabilities computed using the formula:

$$\pi_{\mathrm{j\,i}}(t) = \frac{\kappa_{\mathrm{j\,i}}\, a_i(t)}{\sum_{r=1}^{m} \kappa_{\mathrm{jr}}\, a_r(t)} \tag{21}$$

We further note that we allow $\kappa_{j\,i} \neq \kappa_{i\,j}$ to hold in general, though if κ_{ji} are constructed using a symmetric distance matrix with elements $\epsilon_{j\,i}$ such that

$$\kappa_{\mathrm{j\,i}} = e^{-\delta \epsilon_{\mathrm{j\,i}}}$$

for some scaling constant $\delta > 0$, then the relation $\kappa_{i\,j} = \kappa_{j\,i}$ will hold. The attractivity factor $a_i(t)$ could reflect several different aspects of the subpopulations, including their size, proportion of infected or immune individuals in the subpopulation, and so on. We will assume that two factors play a central role in determining the relative attractivity of each subpopulation: a characteristic size parameter N_i^c and the ratio of infectious individuals $I_i(t)/N_i(t)$ for the ith population, $i = 1, \ldots, m$. For example, we might assume attractivity falls off linearly from 1 to 0 with the ratio I_i/N_i (i.e., use the factor $(1 - I_i/N_i)$). Similarly, we might assume that the attractivity falls off as $N(t)$ [0, ∞) varies on either side of N_i^c (e.g. a factor of the form $(1/e)(N/N_i^c)e^{-N/N_i}$ which ranges between 0 and 1 and back to 0 as N increases from 0 to N_i^c and then beyond to infinity). In this case we may define

$$a_i(t) = k\left(1 - \frac{I_i(t)}{N_i(t)}\right) + (1 - k)\left(\frac{N_i(t)}{\mathrm{e}N_i^c} e^{-N_i(t)/N_i^c}\right)$$

where k [0, 1] switches the emphasis from the population size factor to the prevalence factor as k increases in value from 0 to 1.

The inputs $h_i^{\mathrm{X}}(t)$ and per-capita flow rate outputs q_i^{X} (X = S, E, I and V) for the focal ith subpopulation, can either be 0, constants, or generated using probability distributions in stochastic versions of the model. The inputs will, of course, depend on the density or number of individuals available in the environment surrounding the focal subpopulation i, with population structure taken into account using network or nearest neighbor concepts. In the context of discrete deterministic or stochastic models, we need to account for the per-capita flow rate outputs q^{X} in our competing rates formulations to obtain the

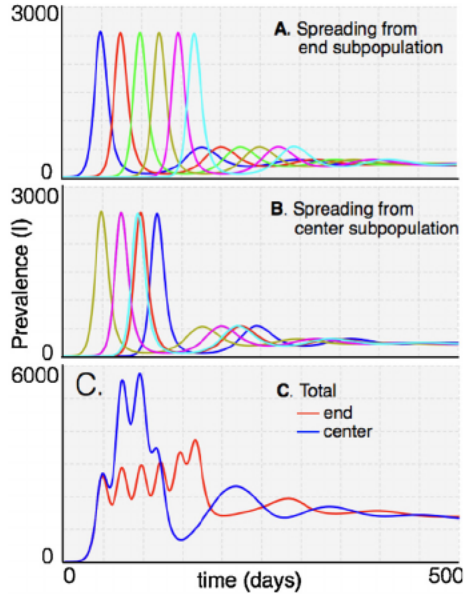

Fig. 7. Prevalence plots predicted by a continuous-time deterministic metapopulation model when the index case starts out in one of six possible locations (each with $S(0) = 20,000$ and other classes at 0, except for the location that has $I(0) = 1$ and $S(0) = 19,999$), where the locations numbered from 0 to 5 are strung out in a straight line in the order 0, 1, 2, 3, 4, 5, and individuals flow only to subpopulations that are their immediate neighbors. Subpopulation prevalence rates are plotted in (A) (index case in subpopulation 0) and (B) (index case in subpopulation 2) with the subpopulation containing the index case clearly leading the outbreaks in closest and the next closest neighboring populations (note that one of the curves is completely obscured by another). Total prevalence is plotted in (C) for outbreaks with two index cases that are either the end two (red plot) or center two (blue plot) subpopulations.

extended probability for the case of rates assumed to be constant over each interval of time (though the rates themselves can vary from one time interval to the next). In this case, we augment the proportions/probabilities in Eq. (10) to define the following terms for constructing the infectious class equation in the ith subpopulation

$$\begin{aligned} p_{\gamma_i|\alpha_i+\mu_i+q_i^{\mathrm{I}}} &\equiv \gamma_i P_i^{\mathrm{I}} \\ p_{\alpha_i|\gamma_i+\mu_i+q_i^{\mathrm{I}}} &\equiv \alpha_i P_i^{\mathrm{I}} \\ p_{\mu_i|\gamma_i+\alpha_i+q_i^{\mathrm{I}}} &\equiv \mu_i P_i^{\mathrm{I}} \\ p_{q_i^{\mathrm{I}}|\gamma_i+\alpha_i+\mu_i} &\equiv q_i^{\mathrm{I}} P_i^{\mathrm{I}} \end{aligned} \tag{22}$$

where

$$P_i^{\mathrm{I}} \equiv \frac{\left(1 - e^{-(\gamma_i+\alpha_i+\mu_i+q_i^{\mathrm{I}})}\right)}{\gamma_i + \alpha_i + \mu_i + q_i^{\mathrm{I}}} \tag{23}$$

and, as before, it follows that

$$\begin{aligned} p_{\gamma_i+\alpha_i+\mu_i+q_i^{\mathrm{I}}} &\equiv \left(1 - e^{-(\gamma_i+\alpha_i+\mu_i+q_i^{\mathrm{I}})}\right) \\ &= p_{\gamma_i|\alpha_i+\mu_i+q_i^{\mathrm{I}}} + p_{\alpha_i|\gamma_i+\mu_i+q_i^{\mathrm{I}}} \\ &\quad + p_{\mu_i|\gamma_i+\alpha_i+q_i^{\mathrm{I}}} + p_{q_i^{\mathrm{I}}|\gamma_i+\alpha_i+\mu_i} \end{aligned}$$

with similar expressions following for the susceptible, the exposed and the immune classes expressed in terms of P_i^{X}, X = S, E, I, and V following the patterns of Eqs. (22) and (23)

These expressions can be used to write down an extended version of the deterministic discrete model given by system of Eq. (12) or of the stochastic model given by system of Eqs. (15) and (16). By way of illustration, using $\widetilde{Q}^{\mathrm{X}}$ to represent the proportion of individuals leaving class X = S, E, I, and V due to immigration, Eq. (15) now become (dropping the argument in t and the subscript i)

$$\begin{aligned} \left(\widetilde{S}, \widetilde{U}^{\mathrm{S}}, \widetilde{M}^{\mathrm{S}}, \widetilde{Q}^{\mathrm{S}}\right) &:= \mathrm{MULTINOMIAL}[S; 1 - p_{\tau_i+\mu+q^{\mathrm{S}}}, p_{\tau_i|\mu+q^{\mathrm{S}}}, \\ &\quad p_{\mu|\tau_i+q^{\mathrm{S}}}, p_{q^{\mathrm{S}}|\tau_i+\mu}] \\ \left(\widetilde{E}, \widetilde{U}^{\mathrm{E}}, \widetilde{M}^{\mathrm{E}}, \widetilde{M}^{\mathrm{E}}\right) &:= \mathrm{MULTINOMIAL}[E; 1 - p_{\sigma+\mu+q^{\mathrm{E}}}, p_{\sigma|\mu+q^{\mathrm{E}}}, \\ &\quad p_{\mu|\sigma+q^{\mathrm{E}}}, p_{q^{\mathrm{E}}|\sigma+\mu}] \\ \left(\widetilde{I}, \widetilde{U}^{\mathrm{I}}, \widetilde{D}, \widetilde{M}^{\mathrm{I}}, \widetilde{Q}^{\mathrm{I}}\right) &:= \mathrm{MULTINOMIAL}[I; 1 - p_{\gamma+\alpha+\mu+\tilde{q}^{\mathrm{I}}}, \\ &\quad p_{\gamma|\alpha+\mu+\tilde{q}^{\mathrm{I}}}, p_{\alpha|\gamma+\mu+\tilde{q}^{\mathrm{I}}}, p_{\mu|\gamma+\alpha+\tilde{q}^{\mathrm{E}}}, p_{\tilde{q}^{\mathrm{I}}|\gamma+\alpha+\mu}] \\ \left(\widetilde{V}, \widetilde{U}^{\mathrm{V}}, \widetilde{M}^{\mathrm{V}}, \widetilde{Q}^{\mathrm{V}}\right) &:= \mathrm{MULTINOMIAL}[V; 1 - p_{\nu+\mu+\tilde{q}^{\mathrm{V}}}, p_{\nu|\mu+\tilde{q}^{\mathrm{V}}}, \\ &\quad p_{\mu|\nu+\tilde{q}^{\mathrm{V}}}, p_{\tilde{q}^{\mathrm{V}}|\nu+\mu}] \end{aligned} \tag{24}$$

Recall that the only source for individuals immigrating to a subpopulation during the interval $[t, t+1)$ are those emigrating from all of the other subpopulations. Given this, as in setting up Eqs. (19) and (21), we can now express the emigrants $H_i^{\mathrm{X}}(t)$ $(i = 1, \ldots, m)$ in terms of the immigrants $\widetilde{Q}_j^{\mathrm{X}}(t)$ $(j = 1, \ldots, m$, for each X = S, E, I, or R), and the parameters π_{ji}:

$$\widetilde{H}_j^{\mathrm{X}}(t) = \sum_{i=1}^{m} \tilde{h}_{\mathrm{ij}}^{\mathrm{X}}(t), \quad \mathrm{X} = \mathrm{S, E, I, and\ V}$$

where the individual $\tilde{h}_{\mathrm{ij}}^{\mathrm{X}}(t)$ are generated from the drawings

$$(\tilde{h}_{\mathrm{ij}}^{\mathrm{X}}(t), \ldots, \tilde{h}_{\mathrm{mj}}^{\mathrm{X}}(t)) = \mathrm{MULTINOMIAL}[\widetilde{Q}_j^{\mathrm{X}}(t); \pi_{j1}(t), \ldots, \pi_{\mathrm{jm}}(t)] \tag{25}$$

With this process completed, we then obtain the following extended version of Eq. (16)

$$\begin{aligned} S_i(t+1) &= \tilde{S}_i(t) + \widetilde{\Lambda}_i(t) + \widetilde{H}_i^{\mathrm{S}}(t) + \widetilde{U}_i^{\mathrm{V}}(t) \\ E_i(t+1) &= \tilde{E}_i(t) + \widetilde{H}_i^{\mathrm{E}}(t) + \widetilde{U}_i^{\mathrm{S}}(t) \\ I_i(t+1) &= \tilde{I}_i(t) + \widetilde{H}_i^{\mathrm{I}}(t) + \widetilde{U}_i^{\mathrm{E}}(t) \end{aligned} \tag{26}$$

The recruitment numbers $\widetilde{\Lambda}_i(t)$, generated during each interval $(t, t+1]$, are drawn from an appropriate discrete stochastic process. The simplest is a Poisson process with expected value Λ_b where the latter is determined by local population birth rates or other processes generating new individuals. Illustrative simulations of the metapopulation model depicted in Fig. 6 for the case of 6 locations strung out in a row, where individuals can only move between neighboring locations are provided in Fig. 7.

4. Fitting models to data

Methods for fitting epidemiological and other types of dynamical systems models to data is a vast area of research in its own right. Here we only touch the surface of the topic and discuss how Numerus Model Builder can be used to address the issue under relatively straightforward and manageable situations (i.e., not too many equations and with a few parameters at most free to vary during the fitting procedure). A gentle introduction to the field of fitting population models to data is provided by Hilborn and Mangel (Hilborn and Mangel, 1997). The issue of model selection itself (Burnham and Anderson, 2003; Johnson and Omland, 2004)—i.e., fitting models with different numbers of parameters to data based on information theoretic concepts—is beyond the scope of our presentation.

Briefly, fitting dynamic models to a set of observations $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$, where the index i in Y_i refers to time $t = i$, $i = 1, \ldots, n$, typically involves generating a set of comparable values $= \{y_1(\hat{\ }), \ldots, y_n(\hat{\ })\}$ from a model that has a set of m parameters $= \{\theta_1, \ldots, \theta_m\}$, where $y_i(\hat{\ })$ is the value of some variable in the model at time $t = i$ when the parameter values are $\theta = \hat{\theta}$.

The two dominant approaches to fitting models to data are least-squares estimation (LSE), which is equivalent to maximum likelihood estimation (MLE) (Myung, 2003) when errors are normal or

asymptotically approaches MLE when sample sizes are very large. The latter is typically embedded in a Markov Chain Monte Carlo (MCMC) algorithm that constructs a probability distribution for using Bayes theorem (Roberts et al., 2004; Lele et al., 2007; Choi and Rempala, 2011; Ionides et al., 2006). MCMC requires the likelihood function to be known. This can be obviated, though, by assuming the distribution of model outcomes to be Poisson (as we do below), using likelihood-function-free methods (Marjoram et al., 2003), or using approximate Bayesian approaches (Beaumont, 2010).

LSE methods involve minimizing the sum-of-squares residuals (or error) measure

$$\mathcal{R}_{SS}(\hat{\ }) = \sum_{i=1}^{m} (y_i(\hat{\ }) - Y_i)^2 \tag{27}$$

On the other hand, MLE methods that assume model outcomes are Poisson, but with a different Poisson mean $y_i(\hat{\ })$ for each data point Y_i, $i = 1, \ldots, t$, involves maximizing the log-likelihood function

$$\ln \quad (Y|\hat{\ }) = \sum_{i=1}^{t} \ln\left(\frac{y_i(\hat{\ })^{Y_i} e^{-y_i(\hat{\ })}}{Y_i!}\right)$$

or minimizing its negative, which can be written for $\hat{y}_i = y_i(\hat{\ })$ as

$$-\ln \quad (Y|\hat{\ }) = \sum_{i=1}^{t} (\hat{y}_i + \ln(Y_i!) - Y_i \ln \hat{y}_i) \tag{28}$$

Here, for purposes of illustration, we fit our deterministic SEIV model to the Sierra Leone Ebola weekly incidence data (Backer and Wallinga, 2016) using both LSE and MLE approaches (Fig. 8: cf. fits obtained in Althaus, 2014). When fitting such data the appropriate initial conditions are generally uncertain because detection of the putative index case does not generally pin down the start of the epidemic: the actual index case may often go undetected and the number of individuals in class E at the time of the first case is also unknown. Thus, as

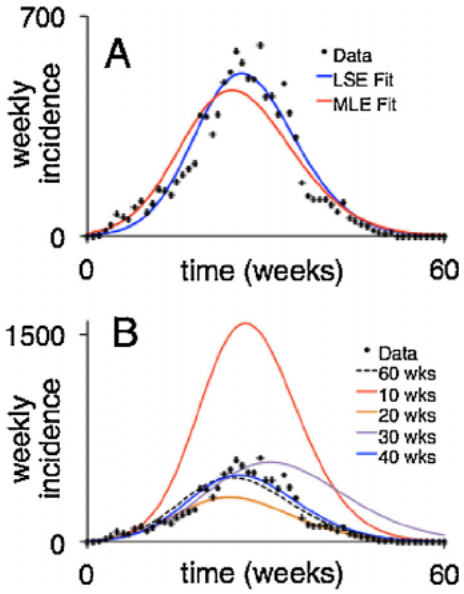

Fig. 8. The SEIVD discrete time model given by Eq. (12) with fixed parameters $\nu = 0$, $\mu = 0.001$, $\alpha = 0.05$ and $\Lambda_t = 0$ (cf. Fig. 4) has been fitted to Ebola data from the Sierra Leone 2014 outbreak in which more than 10,000 cases occurred during the course of an approximately one-year period (Backer and Wallinga, 2016). A. The blue and red curves are the best fit LSE (see Eq. (27)) and MLE (see Eq. (28)) obtained with the optimal parameter sets given in the text. B. The black dotted line is the MLE fit, as in Panel A, with the red, orange, purple and blue plots, simulations obtained after obtaining the best MLE fits to the first 10, 20, 30 and 40 weeks of incidence respectively. See Video 9 at the supporting website for more information on how to set up and run optimizations on this model using Numerus Model Builder.

part of the fitting procedure, we allow the initial values $E(0)$, $I(0)$ in the model to be fitted to the data. To keep the dimensions of the fitting problem down, however, we set

$$E(0) = I(0) = Z$$

and the search for the best fitting value of Z.

Another imponderable is the actual number of individuals $N(0)$ at risk at the start of the epidemic. Thus we also treat $N(0) = N_0$ to be an optimization parameter, though we set $V(0) = 0$ under the assumption that if some individuals in the population were immune to Ebola at the start of the epidemic, this would be reflected in a lower-valued estimate of N_0. Thus, paralleling others who have fitted models using a parameter that represents the proportion of the population at risk (Eisenberg et al., 2015), N_0 should be interpreted as the "initial population at risk" rather than actual population size. Also, in preliminary runs of our optimization using a Nelder–Mead algorithm, the difference between optimal values for σ and γ under variety of settings always lead to optimal values that differ by less than a few percent. Thus to further reduce the dimension of the optimization problem, we set $\sigma = \gamma$ during the optimization procedure. Further, we also fixed μ and γ (which in problems like this can be estimated outside of the incidence data) to 0.001 and 0.05 (which here are rates per week) respectively, as well as setting $\lambda = \nu = 0$. A more rigorous fitting of the Sierra Leone data would need these parameters to be properly estimated ahead of time, but our purpose here is to demonstrate different aspects of the fitting procedure, rather than undertaking an in depth analysis of the epidemic itself.

With the above parameter and model settings, the best-fitting remaining parameter values, denoted by asterisk, were

LSE:

$$(N_0^*,\ E_0^* = I_0^*,\ \sigma^* = \gamma^*) = (409387,\ 6.3,\ 0.43)$$

$$\beta^*(t) = \frac{1.70}{1 + \left(\frac{t}{18.1}\right)^{2.56}}$$

MLE:

$$(N_0^*,\ E_0^* = I_0^*,\ \sigma^* = \gamma^*) = (477435,\ 19.7,\ 0.37)$$

$$\beta^*(t) = \frac{1.49}{1 + \left(\frac{t}{16.8}\right)^{2.24}}$$

In Fig. 8A we see that the LSE and MLE provide similar fits, though both provided relatively poor fits to the first third of data: these data reflect a more linear than the exponential initial phase which is uncharacteristic of homogeneous SEIVD models, particular those that have no spatial structure. In general, we should not expect an SEIVD model to fit the data particularly well because many of the assumptions inherent in the SEIVD model, as discussed in the previous section, are violated to some often-unknown and likely-large degree.

Though the fits are similar—both, for example predict an initial β around 1.5 that drops to half that level in just over two weeks after the first cases have been detected—the LSE fit predicts an 14% smaller initial population at risk than the MLE fit. Thus the errors associated with this estimation can be quite large. These errors can be estimated using other methods such as Bayesian Markov Chain Monte Carlo (MCMC) estimation. Further, as is evident from our plots in Fig. 8B, if we use the first T weeks of data, $T = 10$, 20, 30, and 40, to fit the model, we see that fitting the first 10 weeks (red curve) greatly overestimates the final size of the epidemic, while fitting the first 20 weeks somewhat underestimates the final number of cases. Before leaving this example, we note that in several runs of the optimization algorithm, different starting conditions converged to different solutions, thereby indicating that some of these solutions are local rather than global minima. When this happened, we selected the solution that gave the lowest log-likelihood value, but our searches were not sufficiently exhaustive for us to be sure that we had found the global minimum for each case.

5. Discussion and conclusion

What is the value of building SEIVD models in anticipation of, during, or after an epidemic outbreak, given the level of accuracy that can be expected from such models? As with all models of complex biological systems centered around organisms, populations or communities, the answer is the same: models provide a framework for obtaining insights into dynamic population processes that could not be obtained without them. Further, they provide a means for exploring and assessing the efficacy of interventions and other types of management actions designed to protect, conserve, or exploit the populations under consideration. In the context of epidemics this has certainly been true with regard to implementing vaccination (Dasbach et al., 2006; Edmunds et al., 1999; Getz et al., 2015) and quarantine programs (Bauch et al., 2005), assessing the effects of behavior (Funk et al., 2010), case detection (Dye and Gay, 2003), and treatment rates (Castillo-Chavez and Song, 2004; Salomon et al., 2006), managing the logistics of setting up treatment facilities during the course of epidemics, evaluating the efficacy of educational (Hadeler and Castillo-Chávez, 1995) and prophylactic campaigns (Williams et al., 2006), as well as drug-delivery programs that reduce the risk of producing drug-resistance pathogen strains (Blower and Chou, 2004).

Models are also the best tools for guiding our response to an outbreak once it has begun. Although, after 10 weeks, a fit to the Sierra Leone data set would have substantially underestimated the problem at hand, the fit at 20 weeks provided a much better ball park assessment of the final size of the Sierra Leone outbreak. This remained somewhat true at 30 weeks and certainly so at 40 weeks, despite all the known serious violations of an SEIVD model applied to an inhomogeneous, spatially-structured population. Thus SEIVD models remain an important tool for managing epidemics, provided we treat predictions from such models with circumspection. To this end, Numerus Model Builder provides a tool that allows non-expert coders to explore the behavior and response-to-interventions of SEIVD-elaborated systems in ways previously open only to those trained to code models themselves.

Contributions

The initial text was drafted by WMG and the Numerus model builder platform was created by RS. WMG, RS, OM, and KT built the models. WMG created the figures, HSY and KT created the videos. All authors read and edited the manuscript.

Acknowledgements

The development of NOVA, which is a precursor to Numerus Model Builder, was supported by NSF Grant CNS-0939153 to Oberlin College (PI: RS) and NSF-EEID Grant 1617982 (PI: WMG). We thank John Pataki of Logical Laboratories for his considerable help and input into creating and supporting the Numerus Model Builder website (URL: https://www.numerusinc.com/). We thank Juliet Pulliam for hosting WMG at the South African Center for Epidemiological Modeling at the University of Stellenbosch, South Africa, during the initial drafting of the ms. We thank two anonymous reviewers for extensive comments that have greatly improved this paper.

Supplementary Information.

To enable the reader to run all our models, we have provided links to the 6 models we used to generate the results presented in our paper, as well as links to videos that explain some of the details on how the models where built using the Numerus Model Builder (NMB). These can be found, in the online version, at https://doi.org/10.1016/j.epidem.2018.06.001. A free version of NMB software can be downloaded at https://www.numerusinc.com/.

References

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J.J., et al., 2010. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. BMC Infect. Dis. 10 (1), 190.
Allen, L.J., 2017. A primer on stochastic epidemic models: formulation, numerical simulation, and analysis. Infect. Dis. Model.
Althaus, C.L., 2014. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS Curr. 6.
Andersen, P.K., Geskus, R.B., de Witte, T., Putter, H., 2012. Competing risks in epidemiology: possibilities and pitfalls. Int. J. Epidemiol. 41 (3), 861–870.
Anderson, D.A., Watson, R.K., 1980. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika 67 (1), 191–198.
Backer, J.A., Wallinga, J., 2016. Spatiotemporal analysis of the 2014 Ebola epidemic in West Africa. PLoS Comput. Biol. 12 (12), e1005210.
Balcan, D., Gonçalves, B., Hu, H., Ramasco, J.J., Colizza, V., Vespignani, A., 2010. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. J. Comput. Sci. 1 (3), 132–145.
Bame, N., Bowong, S., Mbang, J., Sallet, G., Tewa, J., 2008. Global stability analysis for SEIS models with n latent classes. Math. Biosci. Eng. 5 (1), 20.
Bauch, C.T., Lloyd-Smith, J.O., Coffee, M.P., Galvani, A.P., 2005. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. Epidemiology 16 (6), 791–801.
Bauer, A.L., Beauchemin, C.A., Perelson, A.S., 2009. Agent-based modeling of host–pathogen systems: the successes and challenges. Inf. Sci. 179 (10), 1379–1389.
Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41, 379–406.
Bellan, S.E., Fiorella, K.J., Melesse, D.Y., Getz, W.M., Williams, B.G., Dushoff, J., 2013. Extra-couple HIV transmission in sub-Saharan Africa: a mathematical modelling study of survey data. Lancet 381 (9877), 1561–1569.
Blower, S.M., Chou, T., 2004. Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance. Nat. Med. 10 (10), 1111.
Blythe, S., Anderson, R., 1988. Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). Math. Med. Biol. 5 (1), 1–19.
Britton, T., 2010. Stochastic epidemic models: a survey. Math. Biosci. 225 (1), 24–35.
Burke, D.S., Epstein, J.M., Cummings, D.A., Parker, J.I., Cline, K.C., Singa, R.M., et al., 2006. Individual-based computational modeling of smallpox epidemic control strategies. Acad. Emerg. Med. 13 (11), 1142–1149.
Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media.
Camacho, A., Kucharski, A., Funk, S., Breman, J., Piot, P., Edmunds, W., 2014. Potential for large outbreaks of Ebola virus disease. Epidemics 9, 70–78.
Castillo-Chavez, C., Song, B., 2004. Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. 1 (2), 361–404.
Castillo-Chavez, C., Hethcote, H.W., Andreasen, V., Levin, S.A., Liu, W.M., 1989. Epidemiological models with age structure, proportionate mixing, and cross-immunity. J. Math. Biol. 27 (3), 233–258.
Choi, B., Rempala, G.A., 2011. Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling. Biostatistics 13 (1), 153–165.
Clancy, D., 2014. SIR epidemic models with general infectious period distribution. Stat. Probab. Lett. 85, 1–5.
Dasbach, E.J., Elbasha, E.H., Insinga, R.P., 2006. Mathematical models for predicting the epidemiologic and economic impact of vaccination against human papillomavirus infection and disease. Epidemiol. Rev. 28 (1), 88–100.
Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, vol. 5 John Wiley & Sons.
Dye, C., Gay, N., 2003. Modeling the SARS epidemic. Science 300 (5627), 1884–1885.
Edmunds, W., Medley, G., Nokes, D., 1999. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. Stat. Med. 18 (23), 3263–3282.
Eisenberg, M.C., Eisenberg, J.N., D'Silva, J.P., Wells, E.V., Cherng, S., Kao, Y.H., et al., 2015. Forecasting and uncertainty in modeling the 2014–2015 Ebola epidemic in West Africa. arXiv preprint. arXiv:150105555.
Fares, A., 2011. Seasonality of tuberculosis. J. Global Infect. Dis. 3 (1), 46.
Fulford, G., Roberts, M., Heesterbeek, J., 2002. The metapopulation dynamics of an infectious disease: tuberculosis in possums. Theor. Popul. Biol. 61 (1), 15–29.
Funk, S., Salathé, M., Jansen, V.A., 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J. R. Soc. Interface rsif20100142.
Getz, W.M., Lloyd-Smith, J.O., 2006. Basic methods for modeling the invasion and spread of contagious diseases. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 71. pp. 87.
Getz, W.M., Pickering, J., 1983. Epidemic models: thresholds and population regulation. Am. Nat. 121 (6), 892–898.
Getz, W.M., Schreiber, S.J., 1999. Multiple time scales in consumer–resource interactions. Annales Zoologici Fennici. JSTOR 11–20.
Getz, W.M., Gonzalez, J.P., Salter, R., Bangura, J., Carlson, C., Coomber, M., et al., 2015. Tactics and strategies for managing Ebola outbreaks and the salience of immunization. Comput. Math. Methods Med. 2015.
Getz, W., Muellerklein, O., Salter, R., Carslon, C., Lyons, A., Seidel, D., 2017. A web app for population viability and harvesting analyses. Nat. Resource Model. 30 (2).
Getz, W.M., 1996. A hypothesis regarding the abruptness of density dependence and the growth rate of populations. Ecology 77 (7), 2014–2026.
Gibson, M.A., Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A 104 (9), 1876–1889.
Gilchrist, M.A., Sasaki, A., 2002. Modeling host–parasite coevolution: a nested approach

based on mechanistic models. J. Theor. Biol. 218 (3), 289–308.
Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22 (4), 403–434.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.
Hadeler, K.P., Castillo-Chávez, C., 1995. A core group model for disease transmission. Math. Biosci. 128 (1), 41–55.
Heffernan, J., Smith, R., Wahl, L., 2005. Perspectives on the basic reproductive ratio. J. R. Soc. Interface 2 (4), 281–293.
Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Rev. 42 (4), 599–653.
Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models With Data, vol. 28 Princeton University Press.
Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., et al., 1991. Common West African HLA antigens are associated with protection from severe malaria. Nature 352 (6336), 595–600.
Ionides, E.L., Bretó, C., King, A., 2006. Inference for nonlinear dynamical systems. Proc. Natl. Acad. Sci. 103 (49), 18438–18443.
Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19 (2), 101–108.
Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. Lond. B: Biol. Sci. 266 (1421), 859–867.
Klepac, P., Caswell, H., 2011. The stage-structured epidemic: linking disease and demography with a multi-state matrix approach model. Theor. Ecol. 4 (3), 301–319.
Koelle, K., Cobey, S., Grenfell, B., Pascual, M., 2006. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. Science 314 (5807), 1898–1903.
Kramer, A.M., Pulliam, J.T., Alexander, L.W., Park, A.W., Rohani, P., Drake, J.M., 2016. Spatial spread of the West Africa Ebola epidemic. Open Sci. 3 (8), 160294.
Kurts, C., Robinson, B.W., Knolle, P.A., 2010. Cross-priming in health and disease. Nat. Rev. Immunol. 10 (6), 403.
Leclerc, P.M., Matthews, A.P., Garenne, M.L., 2009. Fitting the HIV epidemic in Zambia: a two-sex micro-simulation model. PLoS ONE 4 (5), e5439.
Leclerc, M., Doré, T., Gilligan, C.A., Lucas, P., Filipe, J.A., 2014. Estimating the delay between host infection and disease (incubation period) and assessing its significance to the epidemiology of plant diseases. PLoS ONE 9 (1), e86568.
Lele, S.R., Dennis, B., Lutscher, F., 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecol. Lett. 10 (7), 551–563.
Li, M.Y., Muldowney, J.S., 1995. Global stability for the SEIR model in epidemiology. Math. Biosci. 125 (2), 155–164.
Li, M.Y., Wang, L., 2002. Global stability in some SEIR epidemic models. In: Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory. Springer. pp. 295–311.
Lloyd, A.L., Jansen, V.A., 2004. Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. Math. Biosci. 188 (1), 1–16.
Lloyd-Smith, J.O., Galvani, A.P., Getz, W.M., 2003. Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. Proc. R. Soc. Lond. B: Biol. Sci. 270 (1528), 1979–1989.
Lloyd-Smith, J.O., Cross, P.C., Briggs, C.J., Daugherty, M., Getz, W.M., Latto, J., et al., 2005a. Should we expect population thresholds for wildlife disease? Trends Ecol. Evol. 20 (9), 511–519.
Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M., 2005b. Superspreading and the effect of individual variation on disease emergence. Nature 438 (7066), 355.
Lyte, M., 2004. Microbial endocrinology and infectious disease in the 21st century. Trends Microbiol. 12 (1), 14–20.
Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 100 (26), 15324–15328.
McCallum, H., Barlow, N., Hone, J., 2001. How should pathogen transmission be modelled? Trends Ecol. Evol. 16 (6), 295–300.
Merl, D., Johnson, L.R., Gramacy, R.B., Mangel, M., 2009. A statistical framework for the adaptive management of epidemiological interventions. PLoS ONE 4 (6), e5807.
Myung, I.J., 2003. Tutorial on maximum likelihood estimation. J. Math. Psychol. 47 (1), 90–100.
Perelson, A.S., Weisbuch, G., 1997. Immunology for physicists. Rev. Mod. Phys. 69 (4), 1219.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 2007. Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press, Cambridge, England.
Roberts, G.O., Rosenthal, J.S., et al., 2004. General state space Markov chains and MCMC algorithms. Probab. Surv. 1, 20–71.
Salomon, J.A., Lloyd-Smith, J.O., Getz, W.M., Resch, S., Sánchez, M.S., Porco, T.C., et al., 2006. Prospects for advancing tuberculosis control efforts through novel therapies. PLoS Med. 3 (8), e273.
Van den Driessche, P., Watmough, J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180 (1), 29–48.
Vestergaard, C.L., Génois, M., 2015. Temporal Gillespie algorithm: fast simulation of contagion processes on time-varying networks. PLoS Comput. Biol. 11 (10), e1004579.
Williams, B.G., Lloyd-Smith, J.O., Gouws, E., Hankins, C., Getz, W.M., Hargrove, J., et al., 2006. The potential impact of male circumcision on HIV in Sub-Saharan Africa. PLoS Med. 3 (7), e262.