Notes on modelling vector-borne disease transmission

Sam Brand

February 14, 2018

MATHEMATICAL BACKGROUND

In this section I briefly outline (without proof) the results that will be used in modelling the transmission of infectious pathogens between vertebrate hosts (humans, livestock etc.) and arthropod vectors (mosquitoes, midges etc). These results can be found in pretty much any standard textbook on probability theory (e.g. [1]).

Multi-type Branching Processes
A discrete time stochastic process $\{\boldsymbol{Z}(n) \in \mathbb{N}^d,\ n = 0,1,2,...\}$ is called a *Galton-Watson multi-type branching process* (MTBP), with d types, if it obeys a *branching mechanism* for each time step $n = 1,2,3,...,$

$$\boldsymbol{Z}(n+1) = \sum_{i=1}^{d} \sum_{k=1}^{[\boldsymbol{Z}(n)]_i} \boldsymbol{O}_{i,k}.$$

Where $[\boldsymbol{Z}(n)]_i$ is the ith component of $\boldsymbol{Z}(n)$. Each random vector $\boldsymbol{O}_{i,k} \in \mathbb{N}^d$ models the number of offspring the kth individual of type i has amongst the d different types; each $\boldsymbol{O}_{i,k}$ is an independent realisation of the i-type offspring random vector $\boldsymbol{O}_i$. We will interpret the time steps n as generations of infected individuals. If $\boldsymbol{Z}(m) = \boldsymbol{0}$ for any generation $m < \infty$ then $\boldsymbol{Z}(n) = \boldsymbol{0}$ for all subsequent generations $n > m$, and we say that the MTBP goes *extinct*. The *next generation matrix* $\boldsymbol{K}$ for the branching process is defined as the expected number of offspring of type j generated from a single type i individual,

$$\boldsymbol{K} = (k_{ij})_{i,j=1,...,d} = (\mathbb{E}[[\boldsymbol{O}_i]_j])_{i,j=1,...,d}.$$

Assuming that the matrix $\boldsymbol{K}$ is irreducible, the *reproductive ratio*, R_0, for the MTBP is defined as the leading positive real eigenvalue of K[1].

The only result we will need for MTBPs is that if $R_0 \leq 1$ the MTBP will go extinct with probability one[2].

The Exponential Distribution
A non-negative random variable $X \sim \exp(\alpha)$ is *exponentially distributed* with rate parameter α if it has the distribution function,

$$F_X(x) = 1 - e^{-\alpha x}, \qquad x \geq 0.$$

USEFUL PROPERTIES OF THE EXPONENTIAL DISTRIBUTION

- *Memoryless*: $\mathbb{P}(X > h + x | X > x) = \mathbb{P}(X > h)$ for $x, h \geq 0$.
- *Constant hazard rate*: $\mathbb{P}(X \in (x, x+h] | X > x) = \alpha h + o(h)$ for $x, h \geq 0$.
- *Moment generating function*: $M_X(\theta) := \mathbb{E}[\exp(\theta X)] = \frac{\alpha}{\alpha - \theta}$, $\theta < \alpha$.

Each one of these properties can also be used to define an exponential random variable.

Finally, there is a useful result for the probability of another random variable, Y, being less than an exponential random variable:

$$\begin{aligned}
\mathbb{P}(Y < X) &= \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathbb{1}(y < x) \alpha e^{-\alpha x} \, \mathrm{d}x \, \mathrm{d}F_Y(y) \\
&= \int_{-\infty}^{\infty} \int_{y}^{\infty} \alpha e^{-\alpha x} \, \mathrm{d}x \, \mathrm{d}F_Y(y) \\
&= \int_{-\infty}^{\infty} e^{-\alpha y} \, \mathrm{d}F_Y(y) \\
&= M_Y(-\alpha).
\end{aligned}$$

Using $\mathrm{d}F_Y(y)$ is a technical point, it extends the integral definition to random variables which don't have a density function $f_Y = F_Y'$[3]. If Y **does** have a density function then the differential element $\mathrm{d}F_Y(y) = f_Y(y) \, \mathrm{d}y$. Note that M_Y is the moment generating function for the Y random variable, and we have assumed nothing about Y except that it is real valued (and that M_Y exists at $-\alpha$).

[1] The existence of such an eigenvalue r such that $|r| \geq |r'|$ for all other eigenvalues r' is guaranteed by the Perron-Frobenius theorem.

[2] The sole exception is if each individual has exactly one offspring at each generation.

[3] For example if $Y = \bar{Y}$ deterministically, then $F_Y(y) = 1$ if $y \geq \bar{Y}$ and 0 otherwise. This 'not random' random variable doesn't have a density function.

Counting Processes, and their use as integrators
A *counting process*, $\{B(t),\ t \geq 0\}$ is determined by an increasing sequence of random *epoch times* $(T_n)_{n\geq 0}$:

$$B(t) = \sum_{n\geq 0} \mathbb{1}(T_n \leq t)$$

The *waiting durations* between events are defined as $X_n = T_n - T_{n-1}$ for $n = 1,2,3,....$ If each X_n is an independent realisation of $X \sim \exp(\alpha)$ we call the counting process a *Poisson process* with rate α, which is denoted $B \sim PP(\alpha)$. Counting processes can be used as integrators for stochastic integrals with bounded integrands $g(\cdot)$,

$$\int_0^T g(t)dB(t) = \sum_{n:\{0\leq T_n \leq T\}} g(T_n).$$

If the integrand $g(\cdot)$ is predictable[4], and $B(t)$ is a Poisson process with rate α, we have that the average of the stochastic integral is the time integral of the average integrand rescaled by the rate of the Poisson process,

$$\mathbb{E}\Big[\int_0^T g(t)dB(t)\Big] = \alpha \int_0^T \mathbb{E}[g(t)]\,\mathrm{d}t.$$

This is essentially a consequence of the constant hazard rate property of exponential random variables. This integration property can also be used to define Poisson processes via their compensator[5]. This gives the thinning and compound properties of Poisson processes:

- Thinning: Suppose events *can* happen at the epoch times of a rate α Poisson process $B(t)$, but each event only *does* happen with the success of an independent Bernoulli trial at probability p (one for each epoch time). The number of events that do happen is a Poisson process with rate $p\alpha$.
- Compound: Suppose events happen at the epoch times of a set of Poisson processes $(i = 1,...,N)$ each at rate α_i. The total number of events is a Poisson process with rate $\sum_i \alpha_i$.

VECTOR-BORNE DISEASES

Diseases spread by insect biting (i.e. vector-borne diseases or VBDs) include some of the most devastating suffered by humans, e.g. malaria, spread by *Anopheles gambiae spp.* and many other mosquito species and *Flaviviruses* such as Dengue fever, yellow fever and Zika spread by (predominantly) the mosquitos *Aedes aegypti* and *Aedes albopictus*. VBDs can also

[4] Essentially, if g is left-continuous.

[5] A compensator for a stochastic process is the unique predictable process such that the difference between the process and its compensator is a (local) martingale. One definition of a Poisson process at rate α is that it is the only counting process with the simple deterministic compensator αt.

effect commercial livestock causing economic damage, e.g. the diseases of ruminants *Bluetongue virus* spread by *Culicoides* genus biting midges. Other VBDs are zoonotic, affecting wild animals, commercial livestock and humans such as *West Nile virus* and *Rift Valley fever*.

MODEL DESCRIPTION

This lecture presents a modelling approach for VBDs based on the early stages of a VBD outbreak, which captures the generic features of vector life histories that are relevant to transmission via biting. There are two types of individual: hosts and vectors. Each individual is also described by a disease state:

For hosts these disease states are

- **susceptible** (if bitten by an infectious vector could contract the VBD),
- **latent** (infected but not yet infectious),
- **infectious** (if bitten by a susceptible vector then could transmit to that vector), and
- **recovered** (the host has recovered, or died, from the VBD, therefore she no longer transmits, and has at least temporary immunity to re-infection).

The disease states for vectors are

- **susceptible** (biting an infectious host could cause infection),
- **latent** (infected but the infectious pathogen that causes the VBD has not yet escaped the gut of the biting insect and invaded the proboscis, therefore bites do not cause infection), and
- **infectious** (bites from the insect can cause transmission to hosts).

Female insect vectors bite throughout their lives, potentially transmitting disease, using the bloodmeals from hosts for sustenance and in order to bring their eggs to maturity. Between bites female vectors are either producing new eggs (oogenesis), laying eggs (oviposition) or seeking for new hosts in order to bite; a full bite-to-bite cycle is called the gonotrophic period for the vector. For each biting vector we consider their biting process $B(t)$ as being a counting process with waiting durations X_n (these model the duration of the gonotrophic cycle) being independent and identically distributed with mean duration for the gonotrophic cycle $\bar{G}$. Vector lives are short compared to hosts, and they undergo constant hazard of death from predation (and other sources), therefore, in accordance with the constant hazard property, their life duration is modelled as $L \sim \exp(\mu)$ where μ is called the mortality rate of the vector. Not all bites on infectious hosts will cause infection in the biting vector, the probability of transmission from host to vector is denoted $P_{\overrightarrow{HV}}$, similarly not every bite from an infectious vector causes infection, the vector to host transmission probability is denoted $P_{\overrightarrow{VH}}$. The latency period for vectors is called the *extrinsic incubation period* (EIP), or sporogony period

for malaria[6], which is modelled as a random non-negative duration E. We assume that hosts are long lived compared to their duration of latency and therefore all hosts live at least until they become infectious. The difference between the onset time for infectiousness after infection for a host (T_{inf}) and the recovery time for the host (T_{rec}) is called the infectious duration for hosts and is a non-negative random variable D.

Reproductive Ratio for VBDs

I don't define a full population transmission model, which should include dynamics for the numbers of individuals of both types and each disease state. Instead, I rely on a result due to Ball and Donnelly [2]; that the early stages of an outbreak are essentially equivalent to an approximating (multi-type) branching process when the number of susceptible individuals is large. The generations of the approximating branching process represent the number of individuals (of each type) infected in that generation, where the n^{th} generation of the process are all individuals n infections away from the initial cases (generation 0). In particular Ball and Donnelly demonstrate that the probability of the branching process approximation going extinct is the same as that of the full epidemic model when the number of susceptible individuals go to infinity.

Therefore, $R_0 \leq 1$ for the approximating branching process always leads to extinction of the VBD; this kind of argument has been generalised to many types of epidemics [3, 4].

Leading eigenvalue of next generation matrix

The branching mechanism for the approximating branching process can be constructed from our model of the biting process of the vectors, and by assuming that (i) every bite on an infected host is from a susceptible vector, and (ii) every bite by an infectious vector is on a susceptible host. This assumption is reasonable when the number of infected individuals is small compared to the number of susceptibles. I denote the average number of vectors infected by a single infected host during her complete infectious duration $R_{\overrightarrow{HV}}$, and the average number of hosts infected by a single infected vector during her complete infectious duration $R_{\overrightarrow{VH}}$. Therefore the next generation matrix for the VBD is:

$$\boldsymbol{K} = \begin{pmatrix} 0 & R_{\overrightarrow{HV}} \\ R_{\overrightarrow{VH}} & 0 \end{pmatrix}.$$

Which has two distinct eigenvalues: $\pm\sqrt{R_{\overrightarrow{HV}} R_{\overrightarrow{VH}}}$; this implies that

$$R_0^2 = R_{\overrightarrow{HV}} R_{\overrightarrow{VH}}.$$

[6] Malaria gets to be special.

The random number of bites the infected vector makes after her EIP and before her death can be written as a stochastic integral with respect to the biting process

$$N_{\overrightarrow{VH}} = \int_0^\infty \mathscr{H}(t-E)(1-\mathscr{H}(t-L))\,\mathrm{d}B(t).$$

Where $\mathscr{H}$ is the left-continuous version of the Heaviside function[7]. Vectors are infected with the VBD by biting an infected host, therefore by our model assumption (iid waiting times between bites) the time until the next bite (and subsequent bites) is a realisation of the waiting time random variable X. In the lecture, we only got far enough to consider the model choice where $B(t)$ is a Poisson process with biting rate $\alpha = 1/\bar{G}$. Since, the vector life duration is exponentially distributed $L \sim \exp(\mu)$, by the memoryless property of exponential random variables, the vector's remaining life after the infecting bite is also distributed according to L. These two model choices justify treating the infection time for the vector as 'time zero', and ignoring the life-history of the vector before it got infected.

The average number of infections caused amongst hosts per infected vector in the MTBP model[8] is the mean number of bites the infected vector makes after her EIP (see figure 1 for a schematic) and before her death ($\mathbb{E}[N_{\overrightarrow{VH}}]$) multiplied by the per bite probability of transmission from vector to host ($P_{\overrightarrow{VH}}$). From the result on stochastic integrals with respect to Poisson processes we have,

$$\begin{aligned}
R_{\overrightarrow{VH}} &= P_{\overrightarrow{VH}}\mathbb{E}\Big[\int_0^\infty \mathscr{H}(t-E)(1-\mathscr{H}(t-L))\,\mathrm{d}B(t)\Big] \\
&= \alpha P_{\overrightarrow{VH}}\int_0^\infty \mathbb{E}\Big[\mathscr{H}(t-E)(1-\mathscr{H}(t-L))\Big]\,\mathrm{d}t \\
&= \alpha P_{\overrightarrow{VH}}\int_0^\infty F_E(t)e^{-\mu t}\,\mathrm{d}t \\
&= \alpha P_{\overrightarrow{VH}}\Big[\frac{-e^{-\mu t}F_E(t)}{\mu}\Big|_0^\infty + \frac{1}{\mu}\int_0^\infty e^{-\mu t}\,\mathrm{d}F_E(t)\Big] \\
&= \frac{\alpha P_{\overrightarrow{VH}}}{\mu}M_E(-\mu).
\end{aligned}$$

This result can be easily interpreted, it is the probability that the vector survives her EIP ($\mathbb{P}(E < L) = M_E(-\mu)$ see above) multiplied by the average number of bites a vector would make in her lifetime (α/μ) and the probability of transmission per bite ($P_{\overrightarrow{VH}}$). In general, the biting of a vector will not be memoryless and therefore the distribution of E will also 'interfere' with the expected number of bites made post EIP, this result is particular to the Poisson process biting assumption.

[7] $\mathscr{H}(x) = 0$ for $x \le 0$ and 0 otherwise.

[8] Recall the MTBP model assumes that the probability that the vector bites a susceptible host is one.

CALCULATING $R_{\overrightarrow{HV}}$: AVERAGE NUMBER OF VECTORS INFECTED BY A SINGLE INFECTED HOST

Suppose there are N_H hosts available for the N_V size vector population to bite. We assume that there is no reason for the vectors to prefer biting any particular single host, and the desire to bite a particular host is independent of the host's previous bites, therefore each bite from the whole vector population has probability $1/N_H$ of occurring on that particular host. Since the MTBP model is only appropriate when the populations of hosts and vectors is large, the biting rate from the whole population will be very large whilst the probability $1/N_H$ will be vanishingly small. The law of rare events gives that the number of 'successes' from a large numbers of independent trials with very small probabilities of success will be Poisson distributed, therefore the biting rate on an individual host will be (at least approximately) a Poisson process, *even if we don't believe that the biting of any individual vector is a Poisson process.*

We didn't progress past modelling the biting of a vector as a rate α Poisson process in the lecture. Therefore, instead of calculating the rate of biting from a vector population at demographic equilibrium[9], we can calculate the biting process from the whole vector population on a single host (denoted $B^*(t)$) using the compound and thinning properties of Possion processes. $B^*(t)$ is the Poisson process derived from compounding over the biting processes of each vector in the vector population and thinned by considering the chance of selecting one individual to bite; that is

$$B^*(t) \sim PP(\alpha N_V/N_H).$$

Therefore, as before we write the random number of potential infections amongst the vectors due to biting a single infected host as a stochastic integral (see figure 1 for a schematic),

$$N_{\overrightarrow{HV}} = \int_0^\infty \mathbb{1}(T_{inf} < t < T_{rec})\, \mathrm{d}B^*(t).$$

The average number of vectors infected by a single infected host is:

$$\begin{aligned} R_{\overrightarrow{HV}} &= P_{\overrightarrow{HV}} \mathbb{E}\Big[\int_0^\infty \mathbb{1}(T_{inf} < t < T_{rec})\, \mathrm{d}B^*(t) \Big] \\ &= P_{\overrightarrow{HV}} \mathbb{E}\Big[\int_{T_{inf}}^{T_{rec}} \alpha \frac{N_V}{N_H}\, \mathrm{d}t \Big] \\ &= P_{\overrightarrow{HV}} \alpha \frac{N_V}{N_H} \mathbb{E}[D] \end{aligned}$$

CONTROLLING VECTOR-BORNE DISEASES

The full expression for the reproductive ratio is

$$R_0^2 = R_{\overrightarrow{HV}} R_{\overrightarrow{VH}} = V \frac{N_V}{N_H} \frac{\alpha^2 M_E(-\mu)}{\mu} \mathbb{E}[D].$$

Where $V = P_{\overrightarrow{HV}} P_{\overrightarrow{VH}}$ is called the *vector competence* of the vector for the transmissible pathogen. The expression for R_0^2 has a few notable features:

[9] See this paper for more details on this if you're interested [5].

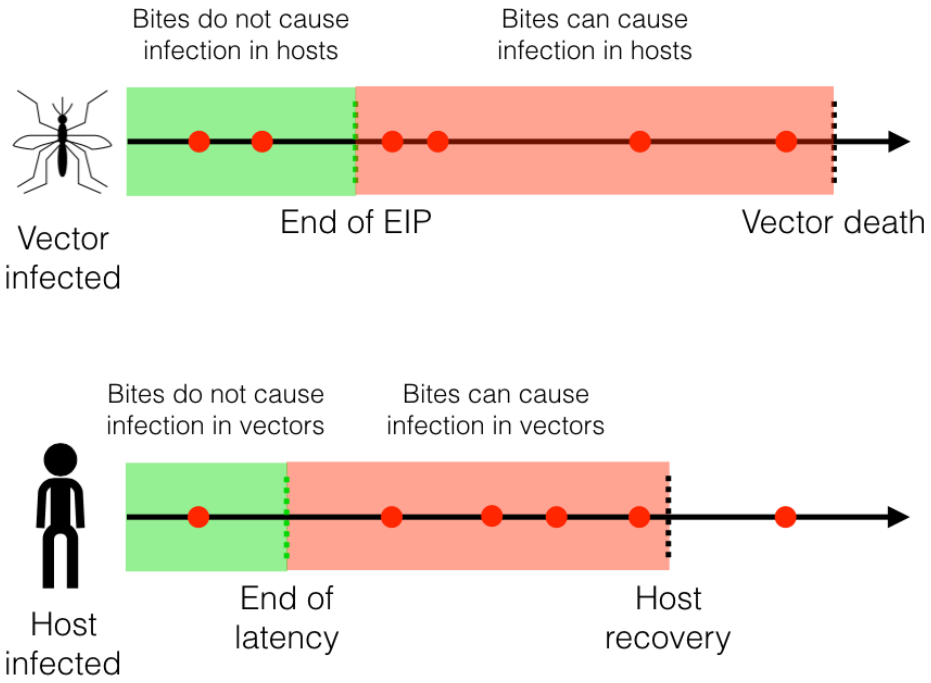

Figure 1: Schematic of transmission from infected vectors and hosts. *Top:* Red dots represent bites from the infected vector onto hosts, but only those after the end of the vector EIP (shaded red; also note that the vector might not survive the EIP) can potentially cause infection. *Bottom:* Red dots represent all bites from vectors onto the host, but only those during the infectious period of the host (shaded red) can potentially cause infection to the biting vector.

- R_0^2 doesn't depend on the absolute numbers of either vectors or hosts, but rather on the vector-to-host ratio. From an epidemiological perspective this is because to be part of a chain of transmission a host must be bitten at least *twice*, the odds of this happening decrease if the number of vectors per host is smaller (even if there were more vectors in absolute numbers).

- R_0^2 is insensitive to the host's epidemiology except for the mean duration of infectiousness; for example the duration of latency for hosts doesn't effect R_0^2, or the distribution of the infectiousness period. This is because we are assuming that the latency period for hosts is short compared to the host's life.

- In contrast, R_0^2 can depend sensitively on the EIP distribution for vectors, because the probability of vectors dying before they become infectious is usually fairly high.

If we can reduce R_0^2 to less than 1 the VBD must be eliminated (see result on MTBPs), therefore the reproductive ratio calculation gives an estimate of the effort that would be required to eliminate a VBD that depends on measurable entomological/epidemiological quantities such as vector biting rate, or the average infectious duration of hosts. We (briefly) consider three broad categories of control each parameterised by a control 'effort' parameter $\chi_c \in [0,1]$:

1. Reducing the vector population. If the vector population can be reduced by some proportion χ_c (maybe by spraying sites where vectors lay eggs with larvicidal treatments) then we will reduce R_0^2 by a factor $(1-\chi_c)$. The critical level of vector population reduction (χ_c^*) is when a post-control $R_0^2 = 1$ is achieved:

$$(1-\chi_c^*)R_0^2 = 1 \implies \chi_c^* = 1 - \frac{1}{R_0^2}.$$

   For example, if $R_0^2 = 10$ before control we would need to reduce the vector population by 90% to achieve elimination.

2. Reducing bites from the vector population. If we can reduce the bites on the host population by a factor χ_c, for example by deploying bed nets to impede nighttime feeding by the vectors on the particular hosts we want to protect (humans), then we will reduce R_0^2 by a factor $(1-\chi_c)^2$. The critical reduction in biting to achieve elimination is:

$$(1-\chi_c^*)^2 R_0^2 = 1 \implies \chi_c^* = 1 - \frac{1}{R_0}.$$

   Note that if $R_0^2 = 10$ before control we would need to decrease biting by 68.4% to achieve elimination, so in this sense elimination by reducing biting is 'easier' than elimination by reducing vector population. This is because reducing biting reduces the number of hosts infected per infected vector **and** the number of vectors infected per infected host. In contrast, reducing the vector population only reduces the number of vectors infected per infected host.

3. Reducing the life expectancy of biting vectors. Often it is not possible to reduce the vector population size either because of logistical constraints or because of larvicidal resistance. However, it is possible to deploy insecticidal poisons close to the hosts such that vectors that bite hosts are exposed and have a reduced life expectancy from $\mathbb{E}[L] = 1/\mu$ before control to the new post-control life expectancy $\mathbb{E}[L_c] = (1-\chi_c)/\mu$. This decrease in life expectancy is equivalent to an increase in the mortality rate $\mu \to \mu/(1-\chi_c)$. We assume that this increase in mortality doesn't affect the number of vector arriving to bite hosts but does reduce the transmission potential of infected vectors, therefore the critical increase in mortality obeys the equation,

$$\frac{R_0^2\mu}{M_E(-\mu)}\frac{M_E(-\mu/(1-\chi_c^*))}{\mu(1-\chi_c^*)} = \frac{R_0^2}{M_E(-\mu)}\frac{M_E(-\mu/(1-\chi_c^*))}{1-\chi_c^*} = 1.$$

This is different from the last two cases: (i) there isn't a simple closed form solution, a numerical solution requires using a root finding solver, and (ii) the estimate of critical life expectancy reduction to achieve elimination will depend not just on R_0^2 (before control), but also the pre-control mortality rate for vectors and the distribution of the EIP.

Traditionally, in mathematical VBD analysis, it was most common to assume that the EIP was the same for each infected vector; that is $E = \bar{E}$ deterministically. In this case, the mean duration of life for the vector *after her EIP* is $M_E(-\mu)/\mu = \exp(-\mu\bar{E})/\mu$. Therefore, the deterministic EIP assumption implies that increasing the mortality rate for vectors exponentially decreases R_0^2, and that mortality rate increase nearly always looks like a favourable control strategy. As computer simulations of VBDs have become more common, its become more common in the literature to assume that the EIP is exponentially distributed[10] with incubation rate $\nu = 1/\bar{E}$. If $E \sim \exp(\nu)$ then the mean duration of life after EIP is $M_E(-\mu)/\mu = \nu/\mu(\mu+\nu)$ and increasing the mortality rate of vectors decreases R_0^2 at sub-exponential rate. If $\bar{E} \ll \mathbb{E}[L]$ then nearly all infected vectors will survive their EIP, and the details about EIP distribution become irrelevant to predicting R_0^2, just as the mean latency period of hosts does not appear in our final expression. However, for a lot of VBDs the mean EIP duration $\bar{E}$ is actually longer than the life expectancy of the vector $\mathbb{E}[L]$... the distribution details about the EIP become significant factors!

This is as far as we got in the lecture.

REFERENCES

[1] Grimmett GR, Stirzaker DR. Probability and Random Processes. 3rd ed. New York: Oxford University Press; 2001.

[10] In this case the VBD model is a simple Markov process and can be simulated using the popular Gillespie algorithm.

[2] Ball F, Donnelly P. Strong approximations for epidemic models. Stochastic Processes and their Applications. 1995;55(1):1–21.

[3] Diekmann O, Heesterbeek J, Metz JA. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology. 1990;28(4):365–382.

[4] Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases. model building, analysis, and interpretation. Wiley; 2000.

[5] Brand SPC, Brand S, Rock KS, Rock KS, Keeling MJ, Keeling MJ. The Interaction between Vector Life History and Short Vector Life in Vector-Borne Disease Transmission and Control. PLoS Computational Biology. 2016 Apr;12(4):e1004837.