Higher Education
Australian Mathematical Sciences Institute

Understanding the effect of individual variation in epidemics

Liam Hodgkinson
Professor Philip Pollett
University of Queensland

Australian Government
Department of Education

Australia devotes considerable resources to programs designed to protect its biodiversity and to limit the spread of disease. For example, the Australian National Health and Medical Research Council (NHMRC) provided $700 million in funding epidemiological research between 2002-2013 [13]. However, these programs can only be effective if the dynamics of the populations in question are well understood. An empirical study was conducted by Lloyd-Smith et al. in 2005 [8] to determine the effect of individual variation in conjunction with Galton-Watson branching processes. The authors' concluded that "emerging disease outbreaks cannot be fully understood if individual variation in infectiousness is neglected" [*ibid.*]. The primary reason for this is the natural emergence of so-called 'super-spreaders' which infect a significantly larger number of people than the average individual. This has become increasingly apparent since Dr Liu Jianlun spread the SARS virus to over 16 other guests on the same floor of a Hong Kong hotel he was staying at; which is believed to have triggered the 2003 global outbreak of the disease [17].

The goal of this project was to introduce individual variation in infectiveness into a more advanced, well-established model in ecology introduced by Hanski in 1994 [5], and apply it in the field of epidemic modelling, with some emphasis on implementing the model in practice, and reproducing the conclusions in [8].

1 The Model

Hanski's incidence function model is a discrete-time, time-homogeneous Markov chain on $\{0,1\}^n$ where n denotes the number of available patches. The model was first introduced in 1994 [5] to model butterfly metapopulations and has become the most commonly used stochastic patch occupancy model (SPOM) in the metapopulation literature [12]. I adopt the notation $\boldsymbol{X}_t^n = \left(X_{1,t}^n, \ldots, X_{n,t}^n\right)$ to denote this Markov chain where $X_{i,t}^n = 1$ if patch i is occupied at time t and $X_{i,t}^n = 0$ otherwise. In the case of epidemic modelling, an 'occupied' patch denotes an area with at least one infected individual present.

The general form of the model assigns two descriptive random variables to each patch i:

- $s_i \in [0,1]$ is the probability that an occupied patch will survive to the next time step
- $A_i > 0$ is a scalar weight which is often interpreted as the area of the patch

as well as a distance parameter $d_{i,j}$ between each two distinct patches i, j. The distribution of the i-th patch at time $t+1$ conditional on the occupation of the patch at time t and its descriptive variables is then given by:

$$X_{i,t+1}^n | \boldsymbol{X}_t^n, \boldsymbol{s}^n, \boldsymbol{A}^n, \boldsymbol{d}^n \sim \text{Bernoulli}\left[s_i X_{i,t}^n + f\left(C_{i,t}\right)\left(1 - X_{i,t}^n\right)\right] \tag{1.1}$$

where $f : [0,\infty) \to [0,1]$, $f(0) = 0$, and the *connectivity measure* $C_{i,t}$ is given by:

$$C_{i,t} = \sum_{j=1}^{n} X_{j,t}^n d_{i,j} A_j \tag{1.2}$$

McVinish and Pollett [10] assigned to each patch i a location $z_i \in \mathbb{R}^d$ in d-dimensional Euclidean space and then considered the logical case where $d_{i,j} = D\left(z_i, z_j\right)$ for some non-negative function D. To incorporate the effect of individual variation in infectiveness into this model, I assign to each patch i a random dispersal parameter $v_i > 0$ and set $d_{i,j} = D\left(v_j; z_i, z_j\right)$. The dispersal parameter v_i will be used to describe the size of the region about an infected individual present in patch i that can be most easily infected.

2 Simulating the Model

In order to properly apply the model in practice, it is necessary to find usable special cases of the model, create an algorithm to calibrate the model to empirical data, and develop efficient simulation strategies. To simplify matters, imagine that the infection evolves over a rectangular space divided into regular sub-rectangular patches. This translates to the following assumptions:

- The collection of all z_i forms a bounded d-dimensional regular lattice $L \subset \mathbb{R}^d$
- The collection of the z_i is distinct; i.e. $z_i \neq z_j$ for all $1 \leq i, j \leq n$, $i \neq j$
- The shortest distance between two consecutive points in L is denoted δ

In the important special case $d = 2$, (1.1) and (1.2) can be easily expressed as matrix equations, which allows for faster simulation using matrix-based computational software such as Matlab, Octave and Numpy.

The ecological and behavioural models (usually derived from random walk models) given in [3, 12, 15, 18] imply that D is often chosen in practice to be of the form:

$$D(v;\, z_1, z_2) = h(v)\, g(v \|z_1 - z_2\|) \tag{2.1}$$

where $\|\cdot\|$ is the Euclidean norm and $g, h : [0, \infty) \to [0, \infty)$ are continuous. Typically g is chosen to be monotone-decreasing which lends itself well to a truncation argument; it is possible to find R such that $g(vr) < \epsilon$, $\forall r > R$ for some chosen tolerance $\epsilon > 0$ and compute D only for z_i, z_j such that $\|z_i - z_j\| \leq R$. This reduces computation time from $O\left(n^2\right)$ to $O(n)$.

With these assumptions in mind, it is important to find suitable distributions for v. An often used parameter in epidemiology is the expected number of secondary infectives, also referred to as the *basic reproduction number* R_0. The problem of estimating R_0 from empirical data is well-studied, so it merely remains to relate the currently abstract dispersal parameter v with the expected number of secondary infectives. This is accomplished by approximating this model on an infinite d-dimensional regular lattice and considering only the expected number of secondary infectives from a single occupied patch; the result of this is found in Proposition 1.

Proposition 1. *Suppose that L is infinite with unit spacing, $A_i = 1$, $s_i = 0$, $v_i = v$ for all $i \in \mathbb{N}$, $X_{i,0} = \delta_{i,j}$ for some $j \in \mathbb{N}$ (where $\delta_{i,j}$ is the Kronecker delta symbol), D is given by (2.1) and $f : [0, \infty) \to [0, 1]$ is continuous. Define the random variable $N = \sum_{i=1}^{\infty} X_{i,1}$ as the number of newly colonised patches at the next time step. Then as $v \to 0^+$:*

$$\mathbb{E}[N] \sim \frac{2\pi^{\frac{d}{2}}}{v^d \cdot \Gamma\left(\frac{d}{2}\right)} \int_0^{\infty} f[h(v)\, g(\delta r)]\, r^{m-1} dr - f[h(v)\, g(0)] \tag{2.2}$$

Proof. (Outline) This result follows from the Riemann-integrability of f, g, h and the main theorem of [1]. □

I remark that Le Cam's theorem (see [2]) loosely implies that N is approximately Poisson with mean λ_v. Under the assumption that the mean number of secondary infectives in an epidemic is Poisson distributed, Lloyd-Smith et al. primarily considered the case where the mean rate was

Γ-distributed. Following this reasoning, one reasonable distribution for i.i.d. v_i can be found by generating $\mathbb{E}[N] \sim \Gamma(\alpha, \beta)$ and inverting (2.2).

3 A Deterministic Limit

When applying the incidence function model in epidemics, it is ideal to be able to increase the resolution of the model to where each patch corresponds to an individual. This is virtually impossible to accomplish via simulation, so analytical techniques are required. In accordance with the analysis in [9,10], we define the (discrete) random measures σ_n and $\mu_{n,t}$ for all $h \in C^+([0,1] \times (0,\infty) \times \Omega)$ by:

$$\int h(s,v,z)\,\sigma_n(ds,dv,dz) = \sum_{i=1}^{n} A_i h(s_i,v_i,z_i) \tag{3.1}$$

$$\int h(s,v,z)\,\mu_{n,t}(ds,dv,dz) = \sum_{i=1}^{n} A_i X_{i,t}^n h(s_i,v_i,z_i) \tag{3.2}$$

or alternatively for any $B \subseteq [0,1] \times (0,\infty) \times \Omega$:

$$\sigma_n(B) = \sum_{(s_i,v_i,z_i)\in B} A_i$$

$$\mu_{n,t}(B) = \sum_{(s_i,v_i,z_i)\in B | X_{i,t}^n = 1} A_i$$

To increase the resolution of the model, the number of patches must increase, and the area of the patches must decrease proportionally. With this in mind, we make the same assumptions as in [10]:

(a) The area parameters are inversely proportional to the number of patches, i.e. $A_i = a_i n^{-1}$ for $a_i \in (0, A]$

(b) The patch locations z_i are located almost surely in a compact set $\Omega \subset \mathbb{R}^d$

(c) $D(v;\, z, \tilde{z})$ is symmetric in $z, \tilde{z}$ and defines a class of functions:

$$\mathscr{A} = \{f :\, (0,\infty) \times \Omega \to [0,\infty) \,|\, f(v,z) = D(v;\, z, \tilde{z}) \text{ for some } \tilde{z} \in \Omega\}$$

which is uniformly bounded and equicontinuous, i.e. $\exists M > 0$ such that $\forall f \in \mathscr{A}$ and $(v,z) \in (0,\infty) \times \Omega$, $|f(v,z)| \leq M$ and for each $\epsilon > 0$ and $(v_0, z_0) \in (0,\infty) \times \Omega$ there exists a neigh-

bourhood N of (v_0, z_0) such that for all $(v, z) \in N$:

$$\sup_{\tilde{z} \in \Omega} |D(v;\, z, \tilde{z}) - D(v_0;\, z_0, \tilde{z})| < \epsilon$$

(d) f is Lipschitz continuous, i.e. $\exists L \geq 0$ such that $\forall x, y \in [0, \infty)$, $|f(x) - f(y)| \leq L|x - y|$

(e) As $n \to \infty$, $\sigma_n \xrightarrow{\mathcal{D}} \sigma$ for some non-random measure σ [by Theorem 16.16 of [7], this holds if (z_i, a_i, s_i, v_i) are i.i.d.]

With these assumptions, the existence of a deterministic limit μ_t and a limiting Markov chain $\boldsymbol{X}_t$ as $n \to \infty$ can be shown by applying Theorem 2 below.

Theorem 2. *Suppose that (a)-(e) hold and that $\mu_{n,0} \xrightarrow{\mathcal{D}} \mu_0$ for some non-random measure μ_0. Then $\mu_{n,t} \xrightarrow{\mathcal{D}} \mu_t$ for all $t = 0, 1 \ldots$ where μ_t is defined by the recursion equation:*

$$\begin{aligned}\int h(s, v, z)\, \mu_{t+1}(ds, dv, dz) = &\int sh(s, v, z)\, \mu_t(ds, dv, dz) \\ &+ \int h(s, v, z)\, f\left[\int D(\tilde{v};\, z, \tilde{z})\, \mu_t(d\tilde{s}, d\tilde{v}, d\tilde{z})\right] \sigma(ds, dv, dz) \\ &- \int h(s, v, z)\, f\left[\int D(\tilde{v};\, z, \tilde{z})\, \mu_t(d\tilde{s}, d\tilde{v}, d\tilde{z})\right] \mu_t(ds, dv, dz) \qquad (3.3)\end{aligned}$$

for all $h \in C^+([0, 1] \times (0, \infty) \times \Omega)$. Moreover, if $X^n_{i,0} \xrightarrow{p} X_{i,0}$ for any $i \in \{1, \ldots, n\}$ then $X^n_{i,t} \xrightarrow{p} X_{i,t}$ for all $t \geq 0$ where:

$$\Pr(X_{i,t+1} = 1 | X_{i,t}, z_i, s_i) = s_i X_{i,t} + f\left[\int D(v;\, z_i, z)\, \mu_t(ds, dv, dz)\right](1 - X_{i,t}) \qquad (3.4)$$

From Theorem 2, if μ_0 is absolutely continuous with respect to σ then μ_t is absolutely continuous with respect to σ for all $t \geq 0$. In this case, the Radon-Nikodym derivative of μ_t with respect to σ evaluated at (s_i, v_i, z_i) can be interpreted as the probability of occupation of patch i at time t in the limiting metapopulation.

With this result, it is now possible to turn to the problem of identifying equilibria of the system, with a view to identifying conditions under which the disease becomes endemic. Let μ_∞ denote a fixed point of the recursion (3.3) and define $\psi(z) = \int D(\tilde{v};\, z, \tilde{z})\, \mu_\infty(d\tilde{s}, d\tilde{v}, d\tilde{z})$. Then:

$$\int h(s, v, z)\, [1 - s + f(\psi)]\, \mu_\infty(ds, dv, dz) = \int h(s, v, z)\, f(\psi)\, \sigma(ds, dv, dz)$$

For the Radon-Nikodym derivative to be defined and to establish results dealing with the space $C(\Omega)$, we require another assumption:

(f) There exists an open neighbourhood V about 1 such that $\sigma(V \times \Omega) = 0$ (i.e. $s_i \notin V$ almost surely; any patch can undergo local extinction)

With this assumption, we have:

$$\int h(s,v,z)\,\mu_\infty(ds,dv,dz) = \int h(s,v,z)\,\frac{f(\psi)}{1-s+f(\psi)}\sigma(ds,dv,dz)$$

and so our desired fixed point equation becomes $\psi = R\psi$ where:

$$R\psi = \int D(\tilde{v};\,z,\tilde{z})\,\frac{f(\psi)}{1-\tilde{s}+f(\psi)}\sigma(d\tilde{s},d\tilde{v},d\tilde{z})$$

Clearly $\psi = 0$ (total extinction) is always a solution of this equation. However we are primarily interested in the existence of non-zero (endemic) equilibria. Following the same procedure as in [10], using the additional assumptions:

(g) f is increasing, strictly concave, and twice differentiable with bounded second derivative in a neighbourhood of 0

(h) $\sigma([0,1]\times N_z) > 0$ for every open neighbourhood N_z of every $z \in \Omega$ (i.e. the z_i are almost surely dense in Ω)

(i) $D \neq 0$ almost surely

the question of the existence of non-zero equilibria is settled by the following theorem:

Theorem 3. *Suppose that assumptions (a)-(i) hold. Let $\mathcal{A} : C(\Omega) \mapsto C(\Omega)$ be the bounded linear operator:*

$$\mathcal{A}\phi(z) = f'(0)\int \frac{D(\tilde{v};\,z,\tilde{z})}{1-\tilde{s}}\cdot\phi(\tilde{z})\,\sigma(d\tilde{s},d\tilde{v},d\tilde{z}) \tag{3.5}$$

for $\phi \in C(\Omega)$. There exists a unique non-zero solution to $\psi = R\psi$ if and only if the spectral radius $r(\mathcal{A}) > 1$.

4 Analysing the Spectral Radius

It has become clear that application of the analytical results in the previous section is reliant upon knowledge of the spectral radius of an associated bounded linear operator $\mathcal{A}$. However it cannot be computed analytically in the majority of cases, and so we require optimal methods to bound and approximate it.

The spectral radius can be bounded using the following mean-value-type result. A short proof of the result has also been provided.

Proposition 4. *There exists some $\xi \in \Omega$ such that:*

$$r(\mathcal{A}) = f'(0) \int \frac{D(v;\xi,z)}{1-s} \sigma(ds, dv, dz) \tag{4.1}$$

Proof. Let $R(\xi)$ be defined as the right-hand side of (4.1). Since D is continuous in ξ and bounded over $\xi \in \Omega$, by the Lebesgue Dominated Convergence Theorem, R is uniformly continuous over the compact set Ω, attaining both its supremum and infimum. It is clear from the definition of $\mathcal{A}$ that $\|\mathcal{A}\|_\infty = \sup_{z\in\Omega} R(z)$. From Proposition 2.1 of [4], for any $\phi \in C(\Omega)$, if $\lambda\phi \leq \mathcal{A}\phi$ then $\lambda \leq r(\mathcal{A})$. Setting ϕ to be constant in Ω yields $\inf_{z\in\Omega} R(z) \leq r(\mathcal{A})$. Since $r(\mathcal{A}) \leq \|\mathcal{A}\|_\infty$, the result follows from the intermediate value theorem. □

The integral (4.1) can be further simplified in the following case where the dispersal parameters are Γ-distributed.

Corollary 5. *Suppose that v_i, z_i are all i.i.d. with z_i distributed in Ω with distribution function ζ, $v_i \sim \Gamma(\alpha, \beta)$ and $s_i = s$ is constant. Set $D(v;\, z_1, z_2) = \exp(-v\|z_1 - z_2\|)$. Then $\exists \xi \in \Omega$ such that:*

$$r(\mathcal{A}) = \frac{f'(0)}{1-s} \cdot \int_\Omega (1 + \beta\|z - \xi\|)^{-\alpha}\, d\zeta$$

It is worth noting as the right-hand side decreases monotonically as the variance of the v_i increases. This reflects the behaviour indentified by Lloyd-Smith et al., and supports the notion that their findings hold for more advanced models. Using simulations, this behaviour can be demonstrated when the distribution of the v_i is more complex. In particular, the infimum and supremum of (4.1) over ξ are typically easy to compute analytically for simple cases of Ω and bound the spectral radius. These bounds can be easily shown to be computed in $O(n)$ time. For a direct $O(n^2)$ approximation to the spectral radius, observe that for the collection of matrices defined by:

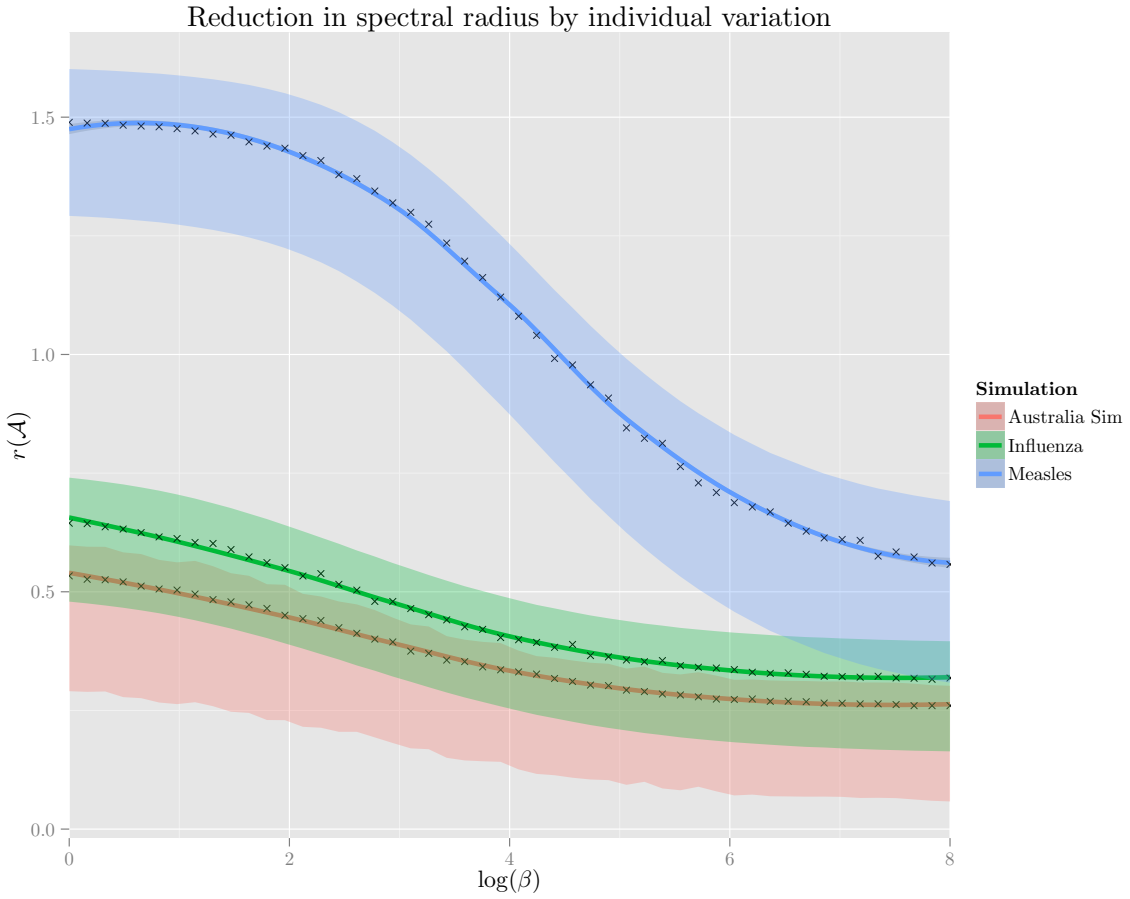
$$(A_n)_{i,j} = \frac{f'(0)}{n} \cdot \frac{D(v_j; z_i, z_j)}{1-s_j} a_j^b \tag{4.2}$$

it can be found that the i-th element of $A_n(\phi(z_1), \ldots, \phi(z_n))^T$ weakly converges to $\mathcal{A}\phi(z_i)$ as $n \to \infty$. Because of this, $\|A_n\|_\infty$ is naturally suited to approximate $\|\mathcal{A}\|_\infty$ as is $r(A_n)$ for $r(\mathcal{A})$. It would be interesting as a future problem to verify whether $r(A_n) \xrightarrow{p} r(\mathcal{A})$ as $n \to \infty$.

Using these methods, the spectral radius was bounded and approximated in modelling the spread of measles and influenza/SARS; this information is presented in Figure 4.1.

It is worth noting that both the construction of A_n and the power method used to calculate its spectral radius are $O(n^2)$. Unfortunately, in most cases, convergence tends to be rather slow. It may be possible to reduce this time in the case where it is only necessary to check whether $r(\mathcal{A}) < 1$ by taking advantage of Jury's test [6].

Figure 4.1: The reduction in the spectral radius $r(\mathcal{A})$ for v_i i.i.d. such that $\mathbb{E}[N] \sim \Gamma(\mu/\beta, \beta)$ (found using Proposition 1) using recorded R_0 parameters $[11, 14, 16]$ of measles (blue; $\mu = 15$), and influenza/SARS ($\mu = 3$) in the maximal spread (green) case and with area parameters derived from Australian population density data (red). The shaded area denotes the bounded region from Proposition 4 while the marked points (with corresponding local regression lines) were computed as the spectral radius of (4.2) with $n = 50$.



5 Conclusion

I have adapted the analytical methods of McVinish and Pollett [9, 10] to analyse a more general form of Hanski's incidence function model which accounts for variation in individual infectiveness. I have also developed an efficient simulation procedure and methods of implementing the model in practice. In future studies, it would be interesting to expand upon Proposition 1, show convergence in probability of parameters in (4.2), and investigate Jury's test for checking $r(\mathcal{A}) < 1$ and the stability problem mentioned in [10].

I would like to thank AMSI for this wonderful opportunity to investigate a fascinating modern ecological model and learn about the exciting work being done in finding new analytical tools to solve probabilistic problems. This summer has given me a valuable insight into the process and current state of academic research in probability theory, and the chance to interact with other like-minded people and learn about other fascinating areas of mathematics. I would also like to thank my supervisor Prof. Philip Pollett as well as Dr Ross McVinish, Dr Thomas Taimre and the other summer scholarship researchers for their assistance and support throughout the course of this project.

References

[1] John A. Baker. Integration of radial functions. *Mathematics Magazine*, 72:392–395, 1999.

[2] A. D. Barbour and Peter Hall. On the rate of poisson convergence. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95:473–480, 1984.

[3] S. R. Broadbent and David G. Kendall. The random walk of trichostrongylus retortaeformis. *Biometrics*, 9:460–466, 1953.

[4] K.-H. Forster and B. Nagy. On the Collatz-Wielandt numbers and the local spectral radius of a nonnegative operator. *Linear Algebra and its Applications*, 120:193–205, 1989.

[5] Ilkka Hanski. A practical model of metapopulation dynamics. *Journal of Animal Ecology*, 63:151–162, 1994.

[6] E.I. Jury. *Inners and the Stability of Linear Systems*. Wiley, 1982.

[7] Olav Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, 2002.

[8] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438:355–359, 2005.

[9] Ross McVinish and Philip Pollett. The limiting behaviour of a stochastic patch occupancy model. *Journal of Mathematical Biology*, 67:693–716, 2013.

[10] Ross McVinish and Philip Pollett. The limiting behaviour of Hanski's incidence function metapopulation model. *Journal of Applied Probability*, 51:297–316, 2014.

[11] Christina E. Mills, James M. Robins, and Marc Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432:904–906, 2004.

[12] Atte Moilanen. SPOMSIM: software for stochastic patch occupancy models of metapopulation dynamics. *Ecological Modelling*, 179:533–550, 2004.

[13] National Health and Medical Research Council. *NHMRC research funding datasets based on burden of disease and health issues*, 2013.

[14] The CDC and the World Health Organization. History and epidemiology of global smallpox eradication. In *Smallpox: Disease, Prevention, and Intervention*, pages 16–17, 2014.

[15] Peter Turchin. *Quantitative Analysis of Movement: Measuring and Modelling Population Redistribution in Animals and Plants*. Sinauer Associates Inc., 1998.

[16] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.*, 160(6):509–516, 2004.

[17] World Health Organization. How SARS changed the world in less than six months. *Bulletin of the World Health Organization*, 2003.

[18] Norikazu Yasuda. The random walk model of human migration. *Theoretical Population Biology*, 7:156–167, 1975.