LETTERS IN BIOMATHEMATICS
An International Journal

Intercollegiate Biomathematics Alliance

RESEARCH ARTICLE

OPEN ACCESS

Mathematical Modeling and Analysis of Influenza In-Host Infection Dynamics

Blessing O. Emerenini[a], Reed Williams[b], Ricardo N. G. Reyes Grimaldo[c], Krista Wurscher[d], Ekeoma R. Ijioma[e]

[a]School of Mathematical Sciences, Rochester Institute of Technology, Rochester NY, USA; [b]State University of New York at New Paltz, NY, USA; [c]Department of Integrative Biology, Oregon State University, OR, USA; [d]University of Oregon, Eugene, OR, USA; [e]Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

ABSTRACT

Influenza is a viral infectious disease of high importance widely studied around the world. In this study we model within-host transmission of influenza in a continuous deterministic setting, a discrete stochastic framework and a spatial-temporal model. Previous models omit cellular restoration through cellular death, which is a key component for the possibility of chronic infections. We thus investigate the effect of cellular restoration on the spread of influenza within the host, through stability analysis of the deterministic model, the probability of state transitions in the stochastic model and the effect of mobility rates on disease spread in the spatial-temporal model. Using the Partial Rank Correlation Coefficient and the Latin Hypercube Sampling, we performed sensitivity analysis to determine which of the parameters are most influential to the model output.

ARTICLE HISTORY
Received December 25, 2020
Accepted September 17, 2021

KEYWORDS
mathematical modeling, influenza, in-host infection dynamics, stochastic

1 Introduction

Cellular regeneration is the regrowth of lost tissues or organs in response to injury. The ciliated cells are the target cells in the upper respiratory tract which are infected by the flu virus. The average ciliated cell population that die as a result of flu infection is huge and if more of the ciliated cells are infected or dead, this will increase the chance of the virus travelling freely through the bronchi to infect the lungs (Zanin et al., 2016; Matrosovich et al., 2004, 2007; Nicholls et al., 2007; van Riel et al., 2007; Yao et al., 2010). The inclusion of cellular regeneration is important in the study of disease state and transitions.

In this investigation, deterministic, stochastic and spatial-temporal models are used to study the effect of the cellular regeneration on the flu-infection dynamics within the upper respiratory tract, the peak infection and disease spread. Many investigations have focused on the dynamics (computational) without much emphasis on the steady states of the disease and the stability of such steady states. A few models, have investigated the stability analysis of flu-infection in population dynamics (Khanh, 2016; Kanyiri et al., 2018).

In order to account for cells which are exposed to the flu-virus but not replicating virus yet, we will consider such cells to be in the eclipse phase (or latent phase). This phase accounts for the time lag between the viral entry into the ciliated cells and active viral production. In general, the length of the eclipse phase is specific to each virus. For instance, the average length of the eclipse phase for human immunodeficiency virus-1 (HIV-1) infection is around 24 hours, whereas for different strains of influenza virus, the length of the eclipse phase can vary from 4 to 12 hours (Kakizoe et al., 2015).

One of the early deterministic within-host models by Perelson et al. (1993) considered the interaction of HIV with CD4 + T cells. The model contains an explicit compartment for uninfected, infected, eclipse and virus. The model exhibited 2 steady states: an uninfected state in which no virus is present and an endemically infected state in which virus and infected T cells are present. Inclusion of eclipse phase in the model has provided a better fit with both numerical and empirical data (Baccam et al., 2006; Beauchemin and Handel, 2011b; Boianelli et al., 2015). The deterministic model will help us to investigate the stability of the steady state of the different compartments.

We extend the ODE model to include variability in cells and virus, this allows us to investigate the future state of the disease in Discrete Time Markov Chains and Continuous-Time Markov Chains (CTMC). In the Discrete-Time Markov chains (DTMC),

CONTACT Blessing O. Emerenini boesma@rit.edu

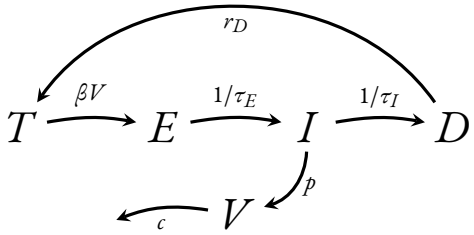


Figure 1: Compartmental diagram for model (1).

time is considered to be a discrete variable that can take values $t_0 + n\Delta t$ for $n \in \mathbb{N}$ and some initial time t_0, whereby Δt is assumed to be small enough so that most events occur during Δt (Allen, 2017; Malik, 2016). In the CTMC, the random variables are discrete value, and account for variability in viral establishment arising from small number of cells and virus at the initiation of an infection and the impact of the regenerated cells in determining the state of the disease. CTMC can be very computationally expensive, as they require generating exponential random variables that dictate how long a cell stays in a given state.

Numerical examples that show the infection dynamics by the various deterministic and stochastic modeling approaches are investigated in the context of Influenza A Virus (IAV) infection. Influenza A virus targets the epithelial cells or the upper respiratory tract with an incubation period of about 48 hours and a range of 24–96 hours (Taubenberger and Morens, 2008). Beauchemin and Handel report the life span of an infected cells from 12–48 hours, the average lifespan of a virion from 0.5 to 3 hours, the length of eclipse phase from 3 to 12 hours and a sufficient rate of cellular regeneration as $1d^{-1}$.

In order to account for the spread of the flu virus and the effect of the cellular regeneration in the disease spread, we formulate a density dependent diffusion coefficient. In the following Section 2, the deterministic within-host model is introduced. The basic reproduction number is calculated and used to study the stability of the steady states for disease free and endemic cases. In Section 3, stochastic models (DTMC and CTMC) are formulated with different but complementary objectives. In Section 4, a spatial-temporal model (partial differential equation) is formulated and studied to account for the disease spread; when the viral transport is dependent on the density of the infected and uninfected cells.

2 Deterministic Model

2.1 Governing equations

The core of our study is the deterministic ordinary differential equations (ODE) model which provides insights to study the disease interaction with cells . The cells are grouped into four classes: Target cells, T, Exposed cell, E, Infectious cells, I, and Dead cells D (see Figure 2.1). Target cell T, represent the cell population susceptible to infection. These cells transition to the exposed class at the rate β. Cells enter the target class at a rate r_D due to cellular restoration which is triggered by dead cells, D. Exposed cells, E represent the cells that have been infected but are not yet producing new virons. This class can also be referred to as the latent or eclipse class. This class gains cells from the target population and loses cells to the infectious class at a rate of $1/\tau_E$. Infectious cells, I, represent the class that actively produces new virons. It gains cells from the exposed class and loses cells to infection related death at a rate of $1/\tau_I$. Finally, Virus, V represents the virus. Infectious cells produce new virons at rate p and the virus is cleared at a rate c:

$$\begin{aligned}
\frac{\mathrm{d}T}{\mathrm{d}t} &= -\beta TV + r_D D \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \beta TV - \frac{E}{\tau_E} \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{E}{\tau_E} - \frac{I}{\tau_I} \\
\frac{\mathrm{d}V}{\mathrm{d}t} &= pI - cV
\end{aligned} \tag{1}$$

where

$$N = D + T + E + I \quad \Rightarrow \quad D = N - T - E - I.$$

Our biological constraints dictate that only non-negative cell populations are relevant. We briefly analyze the behavior of the system of differential equations on the T, E, I and V axes, all the combinations of planes, and all the combinations of hyper-

planes, in order to confirm that solutions always point from the boundaries inward towards the set of positive real numbers of the state space, $\Omega = \{(T,E,I,V) \in \mathbb{R}^4 : 0 \leq T,E,I,V;\ N \geq T+E+I\} \subset \mathbb{R}^4_+$.

In order to secure the existence and uniqueness of solutions for a one dimensional ODE, $\frac{\mathrm{d}}{\mathrm{d}t}y = f(y), f(y)$ must be continuous and $f'(y)$ must exist and be continuous. This can be extended to systems of differential equations, $\frac{\mathrm{d}}{\mathrm{d}t}F = F(x_1(t), x_2(t), \ldots, x_n(t))$. The system must be continuous on an open set $\mathcal{U}$, the Jacobian matrix for F must be defined in $\mathcal{U}$, and the elements of the Jacobian matrix must be continuous. In our case, $\mathcal{U} = \overset{\circ}{\Omega}$, the Jacobian exists and is continuous, over this region. Because $\mathcal{U} = \overset{\circ}{\Omega}$ does not account for the boundaries, the boundary behavior was analyzed to show that points on the boundary always go in towards the interior. The only exception is the T-axis which is invariant.

2.2 Stability analysis

2.2.1 Local stability of disease free equilibrium (DFE)

The model exhibits a DFE, when we set the right-hand sides of the equations of the model to zero. This is given as

$$(N, 0, 0, 0).$$

The local stability of the DFE is explored using the next generator method proposed by van den Driessche and Watmough (2002). The non-negative matrix $\mathcal{F}$, of the new infection terms and the M-matrix, $\mathcal{U}$, of the transition terms associated with the deterministic model are, respectively given by

$$\mathcal{F} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta N \\ 0 & \frac{1}{\tau_E} & 0 & 0 \\ 0 & 0 & p & 0 \end{pmatrix}, \qquad \mathcal{U} = \begin{pmatrix} r_D & r_D & r_D & \beta N \\ 0 & \frac{1}{\tau_E} & 0 & 0 \\ 0 & 0 & \frac{1}{\tau_I} & 0 \\ 0 & 0 & 0 & c \end{pmatrix}. \tag{2}$$

Definition 2.1. (van den Driessche and Watmough, 2002) The basic reproduction number ($\mathcal{R}_0$) is defined as the expected number of cells (secondary cases) that will be infected as a result of one infected cell and is denoted by $(\mathcal{R}_0) = \rho(\mathcal{F}\mathcal{U}^{-1})$ where ρ is the maximum eigenvalue of the matrix $\mathcal{F}\mathcal{U}^{-1}$.

The basic reproduction number of our model is given by

$$\rho(\mathcal{F}\mathcal{U}^{-1}) = \widetilde{\mathcal{R}_0} = \left(\frac{\beta p N \tau_I}{c}\right)^{\frac{1}{3}}. \tag{3}$$

Theorem 2.2. *The Disease Free Equilibrium is locally asymptotically stable when $\mathcal{R}_0 \leq 1$ and it is unstable when $\mathcal{R}_0 > 1$.*

Proof. Suppose that the eigenvalues of their Jacobian matrices have only negative real parts (Wiggins, 2003), then the equilibrium points are locally asymptotically stable, otherwise unstable.

Linearlizing our model, we have the Jacobian matrix evaluated at the disease free equilibrium, such that

$$\lambda I - J_{DFE} = \begin{pmatrix} \lambda + r_D & r_D & r_D & \beta N \\ 0 & \lambda + \frac{1}{\tau_E} & 0 & -\beta N \\ 0 & -\frac{1}{\tau_E} & \lambda + \frac{1}{\tau_I} & 0 \\ 0 & 0 & -p & \lambda + c \end{pmatrix}. \tag{4}$$

Computing the eigenvalues by $\det(\lambda I - J_{DFE})$, we obtain

$$\det(\lambda - J_{DFE}) = (\lambda + r_D)\left(\left(\lambda + \frac{1}{\tau_E}\right)\left(\lambda + \frac{1}{\tau_I}\right)(\lambda + c) - \frac{\beta N p}{\tau_E}\right) = 0. \tag{5}$$

We notice that the term $\beta N p/\tau_E$ can easily be expressed in terms of $\mathcal{R}_0$ and we find that $\frac{\beta N p}{\tau_E} = \frac{c\mathcal{R}_0}{\tau_E \tau_I}$. Substituting in this term and simplifying equation 5, we obtain the following:

$$\det(\lambda I - J_{DFE}) = (\lambda + r_D)\left(\lambda^3 + \lambda^2\left(\frac{1}{\tau_E} + \frac{1}{\tau_I} + c\right) + \lambda\left(\frac{1}{\tau_E \tau_I} + \frac{c}{\tau_E} + \frac{c}{\tau_I}\right) + \frac{c}{\tau_E \tau_I} \cdot (1 - \mathcal{R}_0)\right) = 0. \tag{6}$$

We see that $\lambda = -r_D$ is a negative root so it only remains for us to consider the third degree polynomial. The coefficients of λ^3, λ^2, and λ are all always positive due to our assumptions about the parameters. The term corresponding to λ^0 is positive when $\mathcal{R}_0 \leq 1$ and negative when $\mathcal{R}_0 > 1$.

When $\mathcal{R}_0 \leq 1$ and all the signs of the coefficients of λ are positive, there are zero sign changes in the sequence of coefficients. By Descartes' Rule of signs (Wiggins, 2003, p. 13), there are zero real positive roots. So, when $\mathcal{R}_0 < 1$, all the real components of the eigenvalues are negative and we have local asymptotic stability, as desired. Similarly, when $\mathcal{R}_0 > 1$, there is one sign change in the sequence of coefficients, meaning we have exactly one positive root, giving us instability at the disease free equilibrium.

We can confirm that there are no roots with positive real parts when $\mathcal{R}_0 < 1$ through the Routh Hurwitz test (see the supplementary materials for this article). □

2.2.2 Global stability of DFE

Theorem 2.3. *The disease free equilibrium of* (1) *is globally asymptotically stable whenever* $R_0 < 1$.

Proof. A comparison theorem will be used for the proof. The equations for the infected components of the model (1) can be written as

$$\frac{dE}{dt} = \beta TV - \frac{E}{\tau_E}, \qquad \frac{dI}{dt} = \frac{E}{\tau_E} - \frac{I}{\tau_I}, \qquad \frac{dV}{dt} = pI - cV.$$

These equations can be simplified as follows

$$\begin{pmatrix} \frac{dE}{dt} \\ \frac{dI}{dt} \\ \frac{dV}{dt} \end{pmatrix} = \left(\frac{T}{N}\right) F \begin{pmatrix} E \\ I \\ V \end{pmatrix} - V \begin{pmatrix} E \\ I \\ V \end{pmatrix}. \tag{7}$$

Since $T \leq N$; therefore,

$$\begin{pmatrix} \frac{dE}{dt} \\ \frac{dI}{dt} \\ \frac{dV}{dt} \end{pmatrix} \leq (F - V) \begin{pmatrix} E \\ I \\ V \end{pmatrix}. \tag{8}$$

Because the eigenvalues of the matrix $J = F - V$ have negative real parts when $R_0 < 1$, using comparison theorem (Neubert and Parker, 1989), we have $(E, I, V) \to (0, 0, 0)$. Now by substituting all infected classes equal to zero in the deterministic model, we get $T \to N$ at $t \to \infty$ for $R_0 < 1$.

The epidemiological interpretation of the above result is that if the value of R_0 is kept to less than or equal to unity, the disease can be eliminated. □

2.2.3 Local stability of the endemic equilibrium

Theorem 2.4. *The Endemic Equilibrium is locally asymptotically stable when* $\mathcal{R}_0 > 1$ *and it is locally asymptotically unstable when* $\mathcal{R}_0 \leq 1$.

Proof. Next we will look for local stability for the endemic equilibrium when $\mathcal{R}_0 > 1$. We will denote this point by $(T, E, I, V) = (T^*, E^*, I^*, V^*)$. Our Jacobian at the endemic equilibrium (J_{EE}) after taking the difference from λI is as follows:

$$\lambda I - J_{EE} = \begin{pmatrix} \lambda + \beta V^* + r_D & r_D & r_D & \beta T^* \\ -\beta V^* & \lambda + \frac{1}{\tau_E} & 0 & -\beta T^* \\ 0 & -\frac{1}{\tau_E} & \lambda + \frac{1}{\tau_I} & 0 \\ 0 & 0 & -p & \lambda + c \end{pmatrix}.$$

From this we can calculate the characteristic polynomial of J_{EE}:

$$\det(\lambda I - J_{EE}) = (\lambda + \beta V^* + r_D) \begin{vmatrix} \lambda + \frac{1}{\tau_E} & 0 & -\beta T^* \\ -\frac{1}{\tau_E} & \lambda + \frac{1}{\tau_I} & 0 \\ 0 & -p & \lambda + c \end{vmatrix} + \beta V^* \begin{vmatrix} r_D & r_D & \beta T^* \\ -\frac{1}{\tau_E} & \lambda + \frac{1}{\tau_I} & 0 \\ 0 & -p & \lambda + c \end{vmatrix}$$

$$= (\lambda + \beta V^* + r_D)\left(\left(\lambda + \tfrac{1}{\tau_E}\right)\left(\lambda + \tfrac{1}{\tau_I}\right)(\lambda + c) - \frac{\beta T^* p}{\tau_E} \right) + \beta V^* \left(r_D\left(\lambda + \tfrac{1}{\tau_I}\right)(\lambda + c) + \tfrac{1}{\tau_E}\left(r_D(\lambda + c) + p\beta T^*\right) \right).$$

This term is expanded and simplified because it cannot be factored. T^* is replaced by its equivalent form, $c/(p\beta\tau_I)$, to aid with the simplification. We are left with a characteristic polynomial with the coefficients listed below. We can confirm that there are no roots with positive real parts when $\mathcal{R}_0 > 1$ through the Routh Hurwitz test (see the supplementary materials for this article). □

2.2.4 Global stability of endermic equilibrium

Theorem 2.5. *Let $X^* = (T^*, E^*, I^*, V^*)$ be a positive steady state of the system (1). Then X^* is globally asymptotically stable with respect to initial conditions in $\overset{\circ}{\Omega}$ if $T^* < T, E^* < E, I^* < I, V^* < V$ and $V^* < 4V$, and $pI < cV$.*

Proof. Let us consider the system (1), we will prove the existence of a Lyapunov function. Let us notice that any steady state holds the following steady state equations:

$$\beta T^* V^* = r_D(N - T^* - E^* - I^*) = \frac{E^*}{\tau_E} = \frac{I^*}{\tau_I},$$
$$pI^* = cV^*$$

We define the function

$$L(\mathbf{X}) = \int_{T^*}^{T} 1 - \frac{T^*}{x}\,dx + \int_{E^*}^{E} 1 - \frac{E^*}{x}\,dx + \int_{I^*}^{I} 1 - \frac{I^*}{x}\,dx + \alpha \int_{V^*}^{V} 1 - \frac{V^*}{x}\,dx$$

where α is a positive constant, defined later. Notice that for any $X \in \Omega$ is such that $V(X) \geq 0$ where $V(X) = 0$ if and only if $X = X^*$. Furthermore, notice that

$$\begin{aligned} \dot{L(\mathbf{X})} &= \nabla(L(\mathbf{X}) \cdot \dot{\mathbf{X}} \\ &= \left(1 - \tfrac{T^*}{T}\right) r_D(T^* - T + E^* - E + I^* - I) + \tfrac{I^*}{\tau_I}\left(1 - \tfrac{T^*}{T} - \tfrac{TVE^*}{T^*V^*E} - \tfrac{I^*E}{IE^*} - \tfrac{I}{I^*}\right) + \tfrac{I^*}{\tau_I}\left(\tfrac{V}{V^*} + 2\right) + \alpha\left(1 - \tfrac{V^*}{V}\right)(pI - cV) \end{aligned}$$

where the first term $\left(1 - \frac{T^*}{T}\right) r_D(T^* - T + E^* - E + I^* - I) < 0$ by the conditions of the theorem. Looking at the second term, we want the magnitude of the negative terms to be greater than the positive terms. As in the local stability, we will utilize the Arithmetic Geometric Mean. Since there are 4 negative terms we know the following:

$$\sum_{i=1}^{4} \mathcal{N}_i \geq 4\left(\prod_{i=1}^{4} \mathcal{N}_i\right)^{\frac{1}{4}}. \tag{9}$$

So, it is enough to show that four times the fourth root of the product of the negative terms is greater than 1.

$$4\left(\prod_{i=1}^{4} \mathcal{N}_i\right)^{\frac{1}{4}} = 4\left(\frac{T^*}{T} \cdot \frac{TVE^*}{T^*V^*E} \cdot \frac{I^*E}{IE^*} \cdot \frac{I}{I^*}\right)^{\frac{1}{4}} = 4\left(\frac{V}{V^*}\right)^{\frac{1}{4}} \tag{10}$$

Setting this result greater than 1 and rearranging our terms, we obtain the inequality $V > \frac{1}{4}V^*$, which follows from our hypotheses. So, this term is negative.

We are now left with two terms, $\frac{I^*}{\tau_I}\left(\frac{V}{V^*} + 2\right)$ and $\alpha\left(1 - \frac{V^*}{V}\right)(pI - cV)$. Due to our conditions, we know that the term attached to α is negative. So, we define α so that we ensure the negative term overpowers the only remaining positive one. We define α as

$$\alpha = k \cdot \frac{I^*}{\tau_I}\left(\frac{V}{V^*} + 2\right) \tag{11}$$

where k is some constant that is larger than one when multiplied with $\left(1 - \frac{V^*}{V}\right)(pI - cV)$. □

3 Stochastic Model

In Discrete-Time Markov Chains, time is considered to be a discrete variable that can take on the values $t_0 + n\Delta t$ for $n \in \mathbb{N}$ and some initial time t_0. Thus, in order to make use of Discrete-Time Markov Chains, one must be able to reasonably assume that Δt can be chosen to be small enough so that at most one event occurs during Δt (Allen, 2017; Malik, 2016). On the other hand in Continuous-Time Markov Chains, $t \in [0, \infty)$, which frees us from needing to make such an assumption. However, Continuous-Time Markov Chains can be very computationally expensive, as they require generating exponential random variables that dictate how long a cell or person stays in a given state within the model (Hordijk et al., 1976; Sandmann, 2008). Additionally, given a population of size N, $N + 1$ Kolmogorov's differential equations would be needed for one of the most basic epidemiological models, an SIS model, in the case of Continuous-Time Markov Chains (Keeling and Rohani, 2008). Previous biological models have produced Discrete-Time Markov Chains that are more efficient than Continuous-Time models

Table 1: Transition events and their probabilities in the stochastic model.

Event	Event Description	Transitions	Probability of Event Occurring between time t and $t + \Delta t$
1	Target cell becomes exposed to the viron	$\mathcal{T} \to \mathcal{T} - 1$, $\mathcal{E} \to \mathcal{E} + 1$	$\beta \mathcal{T} \mathcal{V} \Delta t$
2	Exposed cell becomes infectious	$\mathcal{E} \to \mathcal{E} - 1$, $\mathcal{I} \to \mathcal{I} + 1$	$\frac{\mathcal{E}}{\tau_E} \Delta t$
3	Infectious cell dies	$\mathcal{I} \to \mathcal{I} - 1$, $\mathcal{D} \to \mathcal{D} + 1$	$\frac{\mathcal{I}}{\tau_I} \Delta t$
4	Cellular restoration	$\mathcal{D} \to \mathcal{D} - 1$, $\mathcal{T} \to \mathcal{T} + 1$	$r_D \mathcal{D} \Delta t$
5	No change	No transitions	$1 - \left(\beta \mathcal{T} \mathcal{V} + \frac{\mathcal{E}}{\tau_E} + \frac{\mathcal{I}}{\tau_I} + r_D \mathcal{D} \right) \Delta t$

and produce stochastically identical results in the case of biochemical network modeling (Sandmann, 2008) and genetic regulatory network modeling (Ivanov and Dougherty, 2006). However, these biological situations may lend themselves more easily to DTMC models due to having only one independent variable.

Continuous-Time Markov Chains are a popular and well-studied model for the spread of infectious diseases (Allen, 2017; Bai et al., 2019; Edholm et al., 2018; Yan et al., 2016). We have chosen instead to focus on a DTMC model in order to better study methods for using this type of model to predict the spread of infectious diseases and to develop a more computationally efficient model than ones used in the past.

We have chosen to focus on Discrete Homogenous-Time Markov Chains for the stochastic modeling of within-host dynamics of influenza. In this type of model, the classes a cell can be in and time are discrete variables. In our case, a cell may be a target, exposed, infectious, or dead cell in any of the time values $\{t_0,\ t_0 + \Delta t,\ t_0 + 2\Delta t, \ldots\}$. The homogeneous-time aspect of our model indicates that we are assuming that the probability of transitioning between the classes of our model does not depend on time. We will assume that we have a fixed number of cells that may fall in the classes of target cell, exposed cell, infectious cell, or dead cell. The random variables in the stochastic model will be denoted in calligraphic letters to avoid confusion with the non-random variables in the deterministic model. We let N be the total number of cells, $\mathcal{T}$ be the random variable representing the number of target cells, $\mathcal{E}$ be the random variable representing the number of exposed cells, $\mathcal{I}$ be the random variable representing the number of infectious cells, and $\mathcal{D}$ be the random variable representing number of dead cells, such that we have the dynamic states equation as

$$N = \mathcal{T} + \mathcal{E} + \mathcal{I} + \mathcal{D}.$$

Within this model, there are five events that could occur. Some of these events also affect the amount of virus present, represented by the random variable $\mathcal{V}$. These events and the probability of them occurring are summarized in Table 1. Notice that the transition probabilities are given by the transition rates seen in the deterministic ODE model multiplied by Δt.

Notice also that since N is a constant and due to the biological constraint $N = \mathcal{T} + \mathcal{E} + \mathcal{I} + \mathcal{D}$, we have one dependent variable and three near-independent variables considering that N is very large, albeit for simplification we will treat these variables as independent variables. We choose our dependent variable to be $\mathcal{D}$, leaving our independent variables as $\mathcal{T}$, $\mathcal{E}$, and $\mathcal{I}$. This set-up is a multivariate stochastic process $\{(\mathcal{T}(t), \mathcal{E}(t), \mathcal{I}(t))|_{t=0}^{\infty}\}$, that is time-homogeneous and should satisfy the Markov property, discussed below. Thus, we may write our joint probability density function as

$$P_{t,e,i}(t) := \Pr\left[(\mathcal{T}(t), \mathcal{E}(t), \mathcal{I}(t)) = (t, e, i)\right] \tag{12}$$

where $t, e, i \in \{0, 1, 2, \ldots, M\}$. We define M to be the total number of live cells which is bound by the total population size N such that $\mathcal{T}(t) + \mathcal{E}(t) + \mathcal{I}(t) = M \leq N$, this means that the sum of $\mathcal{T}(t), \mathcal{E}(t), \mathcal{I}(t)$ can never exceed the size of the entire population. We can assume that Δt can be sufficiently small such that at most one change in state occurs during the time interval Δt. The probability of transition from the state (t, e, i) to the state $(t + k, e + j, i + l)$ is defined (using the notation in Allen and Burgin, 2000, and Malik, 2016) by

$$P_{t+k,e+j,i+l}(\Delta t) = \Pr[(\Delta\mathcal{T}, \Delta\mathcal{E}, \Delta\mathcal{I}) = (k, j, l) \mid (\mathcal{T}(t), \mathcal{E}(t), \mathcal{I}(t)) = (t, e, i)]$$

where

$$\Delta\mathcal{T} = \mathcal{T}(t + \Delta t), \qquad \Delta\mathcal{E} = \mathcal{E}(t + \Delta t), \qquad \Delta\mathcal{I} = \mathcal{I}(t + \Delta t).$$

We assume that the joint probability density function (12) holds the Markov Property.

Definition 3.1. The *Markov Property* holds that the state at time $t + \Delta t$ is dependent only on the state at time t and is independent of all previous times. If X_t is a vector describing our state at time t, then the Markov Property may be written as

$$\Pr(X_{t+\Delta t} = b \mid X_t, X_{t-\Delta t}, \ldots, X_{\Delta t}, X_0) = \Pr(X_{t+\Delta t} = b \mid X_t).$$

If the Markov Property satisfies, then the values of $\mathcal{T}(t+\Delta t)$, $\mathcal{E}(t+\Delta t)$, $\mathcal{I}(t+\Delta t)$, and $\mathcal{D}(t+\Delta t)$ depend only on the values of $\mathcal{T}(t)$, $\mathcal{E}(t)$, and $\mathcal{I}(t)$.

It is reasonable to assume that our model holds the Markov Property because we would not expect the probability that transition from a current disease state to a new disease state to depend on past disease states. The only factors that should affect the transition from the current disease state are how many virons, target cells, exposed cell, infectious cells, and dead cells there are at each moment; all of which are encompassed in the current disease state. The Markov Property is important for understanding a component of our stochastic model, the transition matrix. The transition matrix is a matrix that governs how values change from one time to the next. The Markov Property allows us to make use of the transition matrix in the following manner.

Denote the transition matrix as $\mathcal{M}$ and the random vector of values of describing the state at time t as $X(t)$. Then,

$$\begin{aligned} X(t_0 + \Delta t) &= \mathcal{M} \cdot X(t_0) \\ X(t_0 + 2\Delta t) &= \mathcal{M} \cdot X(t_0 + \Delta t) = \mathcal{M}^2 \cdot X(t_0) \\ &\vdots \\ X(t_0 + n\Delta t) &= \mathcal{M} \cdot X(t_0 + (n-1)\Delta t) = \mathcal{M}^n \cdot X(t_0). \end{aligned}$$

We give a more formal definition of the transition matrix.

Definition 3.2. (Allen, 2010, p. 47) The *transition matrix* of the Discrete Time Markov Chain $\{X_n\}_{n=0}^{\infty}$ with state space $\{1, 2, \ldots\}$ and one-step transition probabilities, $\{p_{ab}\}_{a,b=1}^{\infty}$, is denoted as $\mathcal{M} = (p_{ab})$, where

$$\mathcal{M} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

In this case, p_{ab} is the probability of transitioning from state a at time t to state b at time $t + \Delta t$ and $\sum_b p_{ab} = 1$.

Discrete Markov Chains are most commonly used when the change in state from a to b can be described by the change in only one variable. So, p_{01} may be thought of as the probability of having 0 cells in one class (say the infectious class) at time t and 1 cell in the same (infectious) class at time $t + \Delta t$. When there is only one independent variable to change between times, this approach is relatively straightforward. However, in our case, we have three independent variables that define our disease state: $\mathcal{T}$, $\mathcal{E}$, and $\mathcal{I}$.

In response to this, we first grouped the exposed and infectious classes together such that the state of the exposed/infectious class can be represented as an ordered pair of the form $(\mathcal{E}, \mathcal{I})$. Then we can form a transition matrix that describes the probability of transitioning in and out of this exposed/infectious class. There are $N + 1$ different matrices of this type that can be formed, where for each matrix the number of target cells, $\mathcal{T}(t) \in \{0, 1, 2, \ldots, N\}$, is fixed. The form of these matrices is described in more depth below. Each of these matrices should be a part of our transition matrix, so our final transition matrix will appear as block matrix that is composed of these smaller matrices on the diagonal. This allows for the transition matrix to account for changes in the number of exposed/infectious cells, but we also need to account for changes in the remaining independent variable, $\mathcal{T}$.

The diagonal of our transition matrix is composed of the blocks described above, so we use the off-diagonal blocks to include diagonal matrices, denoted as $D_{\mathcal{T}(t),\mathcal{T}(t+\Delta t)}$, that account for the probability of transitioning between the $\mathcal{T}(t)$ and $\mathcal{T}(t+\Delta t)$ number of target cells. Like so:

$$\mathcal{M} = \begin{bmatrix} \left[\mathcal{T}(t) = 0\right] & D_{0,1} & & & & 0 \\ D_{1,0} & \left[\mathcal{T}(t) = 1\right] & D_{1,2} & & & \\ & D_{2,1} & & & & \\ & & \ddots & \ddots & \ddots & \\ & & & & & D_{N-1,N} \\ 0 & & & & D_{N,N-1} & \left[\mathcal{T}(t) = N\right] \end{bmatrix}.$$

Table 2: Movements of the transition events of independent variables to attain the $(t,(e,i))$ state.

Event Description	State at t	State at $t+\Delta t$	(k,j,l)
Target cell becomes exposed	$(t+1,(e-1,i))$	$(t,(e,i))$	$(-1,1,0)$
Exposed cell becomes infectious	$(t,(e+1,i-1))$	$(t,(e,i))$	$(0,-1,1)$
Infectious cell dies	$(t,(e,i+1))$	$(t,(e,i))$	$(0,0,-1)$
Cellular restoration	$(t-1,(e,i))$	$(t,(e,i))$	$(1,0,0)$
No change	$(t,(e,i))$	$(t,(e,i))$	$(0,0,0)$

In order to compute the probabilities contained in the transition matrix, we need to determine how a cell can move between our independent classes. Given that we end up with $(\mathcal{T},(\mathcal{E},\mathcal{I}))=(t+k,(e+j,i+l))$ at time $t+\Delta t$, we want to find all the states we could have been in at time t. We assume Δt is very small so that $\mathcal{T}(t)$, $\mathcal{E}(t)$, and $\mathcal{I}(t)$ can change by at most 1 in time Δt; thus $k,j,l\in\{-1,0,1\}$ (see Table 2, for the events and how the event may lead to state $(t,(e,i))$).

The probabilities of the events described in Table 2, are already known, and can be found in the Table 1. Using our knowledge of states at time t that could result in the state $(t,(e,i))$ at time $t+\Delta t$, and the transition probability between these states, we are able to find $P_{ab}(t+\Delta t)$, which can be thought of as the sum of the probabilities of transitioning into the state $b=(t,(e,i))$. The probability of remaining in state b during the time Δt has the probability: 1-(probability of leaving state $(t,(e,i))$). Thus, we have the following lemma:

Lemma 1. *For $t+e+i+d=N$, the probability of leaving state a and entering state b while $\mathcal{T}(t)$ remains fixed in time Δt is given by*

$$P_{ab}=\sum_{\substack{e,i,j,l\geq 0\\ e+i=a\\ e+j+i+l=b}} P_{(e,i),(e+j,i+l)} \tag{13}$$

where

$$P_{(e,i),(e+j,i+l)}=\begin{cases}\frac{e}{\tau_E}\Delta t & (j,l)=(-1,1)\\ \frac{i}{\tau_I}\Delta t & (j,l)=(0,-1)\\ -\frac{e}{\tau_E}\Delta t-\frac{i}{\tau_I}\Delta t & (j,l)=(0,0)\\ 0 & \textit{otherwise.}\end{cases} \tag{14}$$

You may have observed that the transition probability associated with no change (i.e., when $(j,l)=(0,0)$) should be $1-\beta t\mathcal{U}\Delta t-r_D(N-t-e-i)\Delta t-\frac{e}{\tau_E}\Delta t-\frac{i}{\tau_I}\Delta t$ instead of $-\frac{e}{\tau_E}\Delta t-\frac{i}{\tau_I}\Delta t$; however, in some cases we may have several transitions in which $(j,l)=(0,0)$. In these cases, we only want the $1-\beta t\mathcal{U}\Delta t-r_D(N-t-e-i)\Delta t$ to appear in the sum once, where as $-\frac{e}{\tau_E}\Delta t-\frac{i}{\tau_I}\Delta t$ should appear as many times as the no change transition occurs. It is important to note that because of this discrepancy the 'probabilities' described above will be different from the actual probability when $a=b$. In this circumstance, the actual probability is

$$1-\beta t\mathcal{U}\Delta t-r_D(N-t-e-i)\Delta t+P_{aa},\quad \text{where } a \text{ is the state } (e,i).$$

We use the probabilities from Lemma 1 to determine the values in a matrix which we will call $M_{\mathcal{M}}$. To account for the extra added terms, we will form another matrix with the missing values, called M_R such that our transition matrix can be written as $\mathcal{M}=M_R+M_{\mathcal{M}}$.

We will denote our entire transition matrix as $\mathcal{M}$, where $\mathcal{M}$ is the sum of two matrices, M_R and $M_{\mathcal{M}}$. We let M_R contain the off-diagonal blocks of $\mathcal{M}$ and the remaining probability terms that are not generated in Lemma 1, these probabilities are all related to a change in the number of target cells. Each block describing the change in the number of target cells can be written in the form of a diagonal matrix denoted as $D_{\mathcal{T}(t),\mathcal{T}(t+\Delta t)}$.

In general, we have that

$$D_{\mathcal{T}(t),\mathcal{T}(t+\Delta t)}=\begin{bmatrix}\beta t\mathcal{U}\Delta t+r_D(N-t) & 0 & \dots & 0\\ 0 & \beta t\mathcal{U}\Delta t+r_D(N-t-1) & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \beta t\mathcal{U}\Delta t+r_D(0)\end{bmatrix}$$

where $t=\mathcal{T}(t)$. However, it is helpful to break $D_{\mathcal{T}(t),\mathcal{T}(t+\Delta t)}$ into the sum of two matrices, described below. Let $D_{\mathcal{T}(t),\mathcal{T}(t+\Delta t)}=B_{N-\mathcal{T}(t)+1}+C_{N-\mathcal{T}(t)+1}$, where the subscripts indicate the size of the matrix. $I_{N-\mathcal{T}(t)+1}$ is the $(N-\mathcal{T}(t)+1)\times(N-\mathcal{T}(t)+1)$

identity matrix, and

$$B_{N-\mathcal{T}(t)+1} = \begin{bmatrix} \beta t\mathcal{U}\Delta t & 0 & 0 & \dots & 0 \\ 0 & \beta t\mathcal{U}\Delta t & 0 & \dots & 0 \\ 0 & 0 & \beta t\mathcal{U}\Delta t & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \beta t\mathcal{U}\Delta t \end{bmatrix},$$

$$C_{N-\mathcal{T}(t)+1} = \begin{bmatrix} r_D(N-t)\Delta t & 0 & 0 & \dots & 0 \\ 0 & r_D(N-t-1)\Delta t & 0 & \dots & 0 \\ 0 & 0 & r_D(N-t-2)\Delta t & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_D(0)\Delta t \end{bmatrix}.$$

Thus, $B_{N-\mathcal{T}(t)+1}$ is a matrix containing the probabilities of one of the $\mathcal{T}(t)$ targets cell becoming exposed, and $C_{N-\mathcal{T}(t)+1}$ is a matrix containing the probabilities of a dead cell becoming a target cell via cellular restoration. Writing D as the sum of these two matrices is useful for writing M_R, because we are now able to define M_R in the following manner:

$$M_R = \begin{bmatrix} I_{N+1} - B_{N+1} - C_{N+1} & C_{N+1} & 0 & 0 & \dots & 0 \\ B_N & I_N - B_N - C_N & C_N & 0 & \dots & 0 \\ 0 & B_{N-1} & I_{N-1} - B_{N-1} - C_{N-1} & C_{N-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & C_2 \\ 0 & 0 & 0 & \dots & B_1 & I_1 - B_1 - C_1 \end{bmatrix}.$$

$M_{\mathcal{M}}$ contains the small transition matrices along the diagonal of $\mathcal{M}$ with the transition probabilities from Lemma 1. The smaller transition matrices for fixed $\mathcal{T}$ values which exist on the diagonals of $\mathcal{M}$ will be denoted as $\mathcal{M}_t$; where the subscript is the fixed number of target cells. Using the probabilities from Lemma 1, we can generate a general description of a $\mathcal{M}_t$ matrix:

$$\mathcal{M}_t = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{2}{\tau_I}\Delta t & -\frac{2}{\tau_I}\Delta t & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{3}{\tau_I}\Delta t & -\frac{3}{\tau_I}\Delta t & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{N-t}{\tau_I}\Delta t & -\frac{N-t}{\tau_I}\Delta t \end{bmatrix} = \begin{bmatrix} P_{00} & P_{01} & \dots & P_{0(N-t)} \\ P_{10} & P_{11} & \dots & P_{1(N-t)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{(N-t)0} & P_{(N-t)1} & \dots & P_{(N-t)(N-t)} \end{bmatrix}.$$

Using the notation described above, we are able to write our finished transition matrix, $\mathcal{M}$ as the sum of the matrices M_R and $M_{\mathcal{M}}$, where M_R is as described above, and

$$M_{\mathcal{M}} = \begin{bmatrix} \mathcal{M}_0 & 0 & 0 & \dots & 0 \\ 0 & \mathcal{M}_1 & 0 & \dots & 0 \\ 0 & 0 & \mathcal{M}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathcal{M}_N \end{bmatrix}.$$

Notice that all rows of any $\mathcal{M}_t$ matrix sum to 0. Since the remaining entries in $M_{\mathcal{M}}$ are all 0, we can see that any row in $M_{\mathcal{M}}$ sums to 0. In the matrix M_R, B_{N+1} and C_1 are both Zero-matrices, so adding matrices row-wise sums to the identity matrix, which has 1's along the diagonal and 0 everywhere else. Thus, every row in M_R sums to 1. Since every row in M_R sums to 1 and every row in $M_{\mathcal{M}}$ sums to 0, we can see that every row in $\mathcal{M}$ sums to 1+0 = 1. So, we have confirmed that $\mathcal{M}$ meets the condition to be a transition matrix.

To use $\mathcal{M}$ as our transition matrix, there is still one problem we need to address. The way $\mathcal{M}$ is described above still contains a random variable $\mathcal{U}$. Since we do not want any random variables in our matrix, this is something that needs to be changed. Note that $\mathcal{U}$ represents the amount of virus that is present, and because the virus is produced by infectious cells, we would expect the amount of virus to be a function of both the number of infectious cells, and the production rate at which infectious cells produce new virons. Thus, we let $\mathcal{U} = \mathcal{U}(i) = p \cdot i$, where $i = \mathcal{I}(t)$. When we go to replace $\mathcal{U}$ with $p \cdot i$, we run into a new problem;

Table 3: Possible transition events starting with $e + i = y$ exposed/infectious cells.

(e, i) at time t	$(e + j, i + k)$ at time $t + \Delta t$	Event Description
$(y, 0)$	$(y, 0)$	No change
	$(y - 1, 1)$	Exposed cell becomes infectious
$(y - 1, 1)$	$(y - 1, 1)$	No change
	$(y - 2, 2)$	Exposed cell becomes infectious
$\vdots$	$\vdots$	$\vdots$
$(1, y - 1)$	$(1, y - 1)$	No change
	$(0, y)$	Exposed cell becomes infectious
$(0, y)$	$(0, y)$	No change

what is the value of $\mathcal{I}(t)$? Recall that the exposed and infectious classes were grouped together, so at any point in the transition matrix, if we have a exposed/infectious cells at time t and b, exposed/infectious cells at time $t + \Delta t$, how many infectious cells are there at time t? The term $\mathcal{U}$ only appears on the main diagonal of $\mathcal{M}$ and in the B_{N-t} off-diagonal matrices. Recall also that each row of the transition matrix must sum to 1, so the value of $\mathcal{U}$ must be the same in both places it appears in one row; this allows us to focus on the value of $\mathcal{I}(t)$ on the main diagonal.

Let $y = e + i$, then y is the total number of exposed/infectious cells at time t, which functions as the index of the rows of $\mathcal{M}_t$. Consider the diagonal entry of an arbitrary yth row of $\mathcal{M}_t$. The value of this entry represents the probability of transitioning from y exposed/infectious cells to y exposed/infectious cells in time Δt, when there are $\mathcal{T}(t) = t$ target cells. We make a Table of all transitions of y exposed/infectious cells to y exposed/infectious cells that have a non-zero probability (see Table 3).

We wish to find the expected value of i at time t when we transition from y to y exposed/infectious cells. This is given by the lemma below.

Lemma 2. *The expected number of infectious cells at time t when transitioning from y exposed/infectious cells to y exposed/infectious cells is denoted $\mathbb{E}(i)$, and is given by*

$$\mathbb{E}(i) = \frac{y^2}{2y + 1}$$

(see the supplementary materials for the derivation).

Implications of Lemma 2 Any time we know the combined number of exposed/infectious cells, y, we can calculate the expected number of infectious cells. We may also use this to replace the random variable $\mathcal{U}$ with the virus production rate per infectious cell times the expected number of infectious cells, $V(y) = p \cdot \frac{y^2}{2y+1}$. Thus, we are able to eliminate all random variables from our matrix. Lemma 2 also provides us with the ability to approximate the number of infectious cells at any time point in our stochastic simulations.

3.1 Computational realization

Recall that given a transition probability matrix, $\mathcal{M}$, and a vector describing the state of the disease at time t_0, which we will call $X(t_0)$, one may determine the vector describing the likely state of the disease at time $t_0 + n\Delta t$, $X(t_0 + n\Delta t)$ by

$$X(t_0 + n\Delta t) = \mathcal{M}^n X(t_0).$$

To carry out these sorts of computations, we need to know the form of the vector $X(t)$ that describes the state of the disease at time t. Since $X(t)$ will be multiplied by $\mathcal{M}$, and $\mathcal{M}$ is a $\frac{(N+1)(N+2)}{2} \times \frac{(N+1)(N+2)}{2}$ matrix, we know that $X(t)$ must be a vector of length $(N + 1)(N + 2)/2$. This makes sense because there are $(N + 1)(N + 2)/2$ possible states of the form (t, y, d), where $t = \mathcal{T}(t)$, $y = e + i$, and $d = \mathcal{D}(t)$; and $X(t)$ is a vector containing the probabilities of being in each of these states.

Recall also that $\mathcal{M}$ is a block-diagonal matrix, where each sub-matrix on the diagonal is similar to a transition probability matrix where the number of target cells is fixed. If the fixed number of target cells is $\mathcal{T}(t)$, then the size of sub-matrix corresponding to this number of target cells is $(N - \mathcal{T}(t)) \times (N - \mathcal{T}(t))$. There should be a corresponding section of the vector $X(t)$ that has a length of $N - \mathcal{T}(t)$, as demonstrated in Figure 2.

The initial state vector, $X(t_0)$ should be able to specify the initial probabilities of the disease at a given state. Since we want to begin with full certainty as to which state we are in at t_0, we know that a block of $X(t_0)$ should be a vector containing the

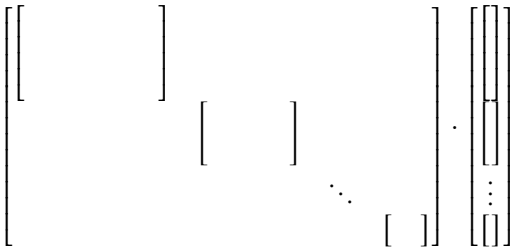


Figure 2: Basic form of Transition Matrix and Disease-State Vector. The boxed areas indicate regions corresponding to a fixed number of target cells.

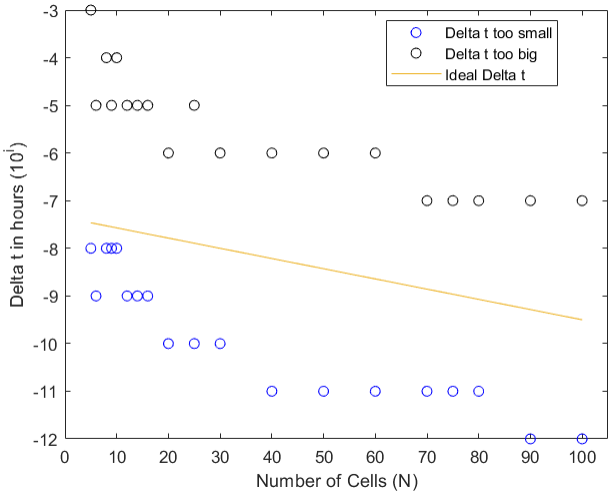


Figure 3: Scatterplot of the boundaries for Δt for $0 \leq N \leq 100$ with function for the ideal Δt.

distributions of that state and adding up to one. For simplicity we let one 1 at the entry that represents part of our vector that defines the initial disease state, and 0's everywhere else. Thus, if we start with $N-3$ target cells, then the vector $X(t_0)$ should contain only 0's except for the third to last section of the vector. This section of the vector should be the same size as the third to last block of the matrix $\mathcal{M}$. This matrix has a size of $(N-(N-3)+1)\times(N-(N-3)+1)$ or 4×4, thus the corresponding part of $X(t_0)$ has four entries.

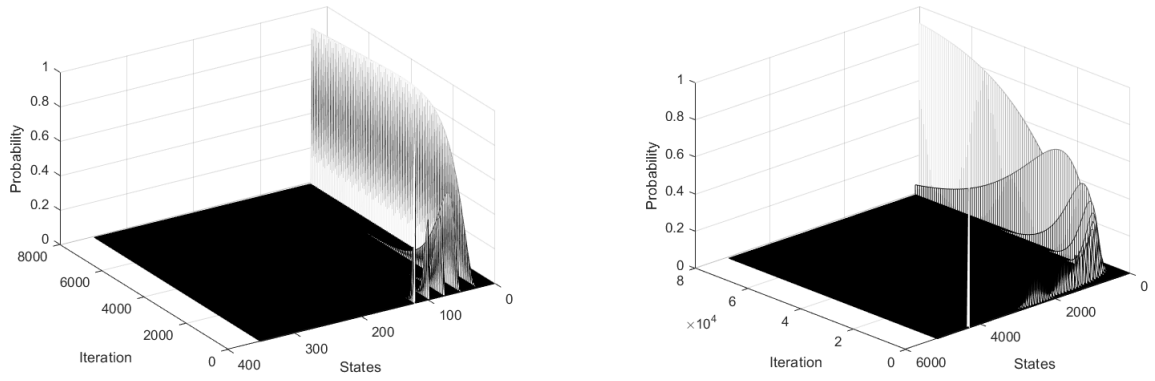
If there are $N-3$ target cells, then there are 3 cells remaining that fall into some other class; namely the exposed/infectious class, or the dead class. There could be 0, 1, 2, or 3 exposed/infectious cells. Since $y+d=3$ in this case, we have four possible states of the form (t,y,d): $(N-3,0,3)$, $(N-3,1,2)$, $(N-3,2,1)$, or $(N-3,3,0)$. In fact, in general, if there are $N-n$ target cells, then there can be $0,1,2,\ldots,n$ exposed/infectious cells, forming $n+1$ possible states that are represented by an $n+1$ length section of $X(t_0)$.

Since the section of $X(t_0)$ that corresponds to $\mathcal{T}(t)$ has the same number of rows as $\mathcal{M}_{\mathcal{T}(t)}$, we would expect the rows of that part of $X(t_0)$ and the rows of $\mathcal{M}_{\mathcal{T}(t)}$ to represent the same thing. Say, we wanted to start with 1 exposed/infectious cell. Since there are $N-3$ target cells, this means that there are 2 dead cells. The rows of $\mathcal{M}_{\mathcal{T}(t)}$ represent the number of exposed/infectious cells before the time Δt has passed. So, having 1 exposed/infectious cell would put us in the second row, because the first row refers to having 0 exposed/infectious cells. Thus, the corresponding part of $X(t_0)$ is $\left[\left[\,0\,1\,0\,0\,\right]^T\right]$.

Thus, if we wanted our initial disease state to be $N-3$ target cells, 1 exposed/infectious cell, and 2 dead cells, then we have

$$X(t_0) = [\,0\,0\,\cdots\,0\,1\,0\,0\,\cdots\,0]^T.$$

Using our transition matrix and a vector describing an initial disease state, we can simulate the expected trajectory of the disease by repeatedly multiplying the initial state vector by the transition matrix. Every time we multiply by the transition matrix, we move forward by Δt in time. As stated previously Δt must be chosen to be small enough so that multiple events do not occur simultaneously. When Δt is chosen to be too large, the simulation breaks down and produces impossible results (such as probabilities as large as 10^{136}), so this problem is quite easy to spot. When Δt is chosen to be too small, then the probability of an event occurring during the time Δt is too small and the simulation indicates that there is no change from the initial state, even



(a) (b)

Figure 4: (a) Probability of each state for 25 cells over 8000 iterations (approximately 2.88 seconds) starting at the disease state $(4, 21, 0)$. (b) Probability of each state for 100 cells over 80,000 iterations (approximately 0.43 seconds) starting at the disease state $(60, 5, 35)$ and approaching the state $(0, 65, 35)$.

after a large number of iterations. Thus, there is an ideal range for the value of Δt. Due to our complexity, the ideal size of Δt depends on the number of cells that are being simulated, so a relationship between N and Δt needs to be established. In order to determine the appropriate form of this relationship, we first collected data on the bounds of Δt for each N to produce the scatterplot on Figure 3.

If we let $\Delta t = 10^{-i}$ for some $i \in \mathbb{N}$, then the ideal region for i appears to follow an approximately linear path in relation to N. We let Δt be on the larger end of the ideal range for two reasons: problems due to Δt being too large are easy to spot and a larger Δt saves computation. This linear path is approximated from the points (100, 9.5) and (30, 8) from Figure 3, and we obtain $-(\frac{3}{140}N + \frac{103}{14})$. So, we let $\Delta t = 10^{-\lfloor \frac{3}{140}N + \frac{103}{14} \rfloor}$ to simulate our stochastic model.

There are several patterns that appear in the stochastic simulations that occur at different time scales. We will begin with the smallest time scale and work our way to a larger scale.

Simulation using 80,000 or less iterations

We begin by looking at what occurs when we simulate $\mathcal{M}^n \cdot X(0)$ for $0 \leq n \leq 80{,}000$. This produces a wave-like pattern, which can be seen in Figure 4.

We observe that as the probability of one state begins to decline, the probability of one other state rises. This propagates the most likely state down to an eventual end state, at which point, the probability of this last state appears to approach 1.

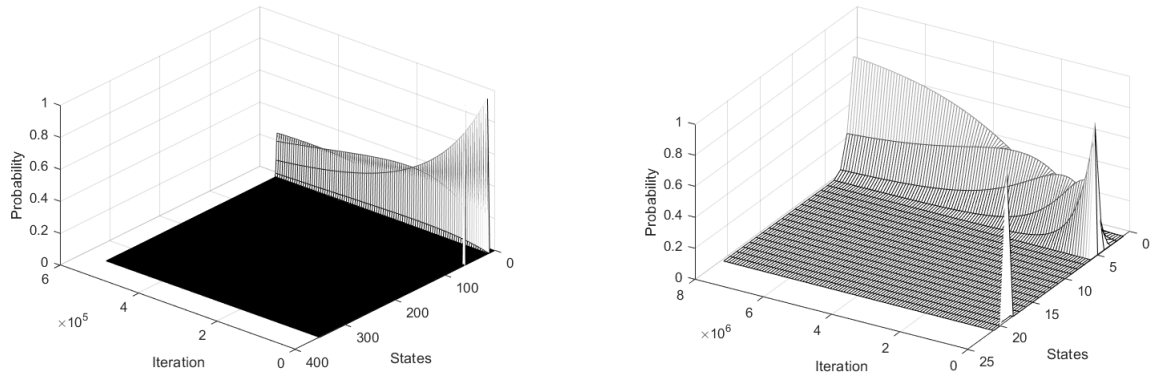
Upon close examinations of the most probable states at any given time, we see that it is easy to predict the course of the simulation, given the initial vector. We give an example of a simulation with a relatively small number of cells. In Figure 4(a), we see that there are five states with non-zero probabilities. The initial state (the spike with probability 1 seen at iteration 0) can be represented by the ordered triplet $(4, 21, 0)$, which is of the form (t, y, d). The next state that emerges after approximately 1000 iterations is $(3, 22, 0)$. The next state that peaks in probability is $(2, 23, 0)$, then $(1, 24, 0)$, and finally, the last state that appears is $(0, 25, 0)$. So, we have the following sequence between states:

$$(4, 21, 0) \to (3, 22, 0) \to (2, 23, 0) \to (1, 24, 0) \to (0, 25, 0).$$

Notice that the initial number of dead cells is always the most likely number of dead cells, meaning that if the initial disease state is (t_0, y_0, d_0), then states with non-zero probabilities will be of the form (t_n, y_n, d_0) for $0 \leq n \leq 80,000$. One can also see that the target cells tend to become exposed/infectious cells while the number of dead cells remain constant. Thus, on this time scale, we see that initial states of the form (t_0, y_0, d_0) go to $(0, y_0 + t_0, d_0)$ within 0 to 80,000 iterations. This pattern can be viewed with a larger number of cells as shown in Figure 4(b), which appears same regardless of initial disease state. Without looking further, we might be led to believe that given an initial disease state of (t_0, y_0, d_0), that the disease approaches the state $(0, y_0 + t_0, d_0)$ as $t \to \infty$; however, this does not appear to be the case.

Simulations using 80,000 to 600,000 iterations

If we look at time values beyond 80,000 iterations, we see that the probability of the $(0, y_0 + t_0, d_0)$ state decreases, as the probability of $(0, y_0 + t_0 - 1, d_0 + 1)$ increases. This new transition in probabilities occurs much slower, and the next state does



(a) (b)

Figure 5: (a) Probability of each state for 25 cells over 550,000 iterations (approximately 3.3 minutes) starting at the disease state $(2, 8, 15)$, (b) Probability of each state for 5 cells over 8,000,000 iterations (approximately 48 minutes) starting at the disease state $(4, 0, 1)$.

not fully surpass the $(0, y_0 + t_0, d_0)$ state until approximately $t_0 + 550{,}000\Delta t$ for the simulations displayed in Figure 5.

In the simulation in Figure 5, we use $N = 25$ cells in 5(a) and $N = 5$ cells in 5(b) respectively. The first several spikes demonstrate the pattern described above. The initial state $(2, 8, 15)$ transitions to $(0, 10, 15)$, which completes the sequence of states described above. The end of this sequence is recognizable because the state that it leads to, the state $(0, 10, 15)$ (all target cells having become infected) occurs with very high probability. This high probability is sustained for some time, but as one can see in the above figure, that a new state $(0, 9, 16)$ eventually surpasses this peak disease state in probability, which is in turn surpassed by the state $(0, 8, 17)$.

Subsequently, if we focus on a larger time scale, we see a new wave of high-probability states begin, forming a progression from $(0, 8, 17)$ to $(0, 7, 18)$, $(0, 6, 19)$, $(0, 5, 20)$, $(0, 4, 21)$, $(0, 3, 22)$, $(0, 2, 23)$, $(0, 1, 24)$ and eventually $(0, 0, 25)$. Once there are 0 target and 0 exposed/infectious cells, the probability of state $(0, 0, d_0 + t_0 + y_0)$ appears to approach 1. Only the start of this progression can be seen in Figure 5(a), however, this full pattern can be seen in Figure 5(b). It becomes very computationally expensive to run simulations involving more than 8 million iterations, especially for large N, thus it is unclear if this the state $(0, 0, d_0 + t_0 + y_0)$ persists, or if new state with higher probabilities eventually arise. This is a source for future research.

4 Spatial-Temporal Model

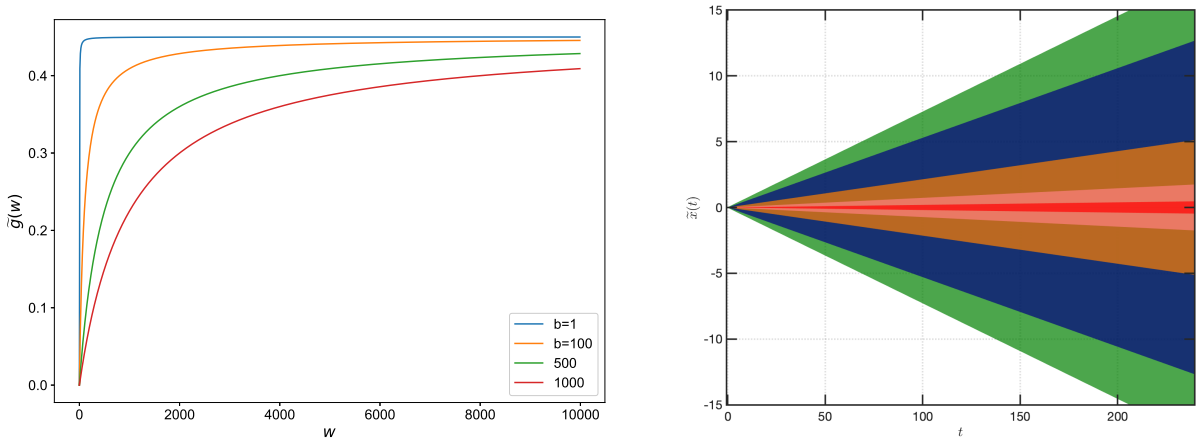
It has been reported experimentally that the spread of some species of virus, the vaccinia virus, depends on its interaction with infected cells through a phenomenon referred to as the repulsion effect (Doceul et al., 2010), which has been modeled mathematically to provide insights on the spread of virus within-host. The reaction-diffusion model of Lai and Zou (2014) captures all the dynamics described by Doceul et al. (2010) and predicts that viruses spread more rapidly when the repulsion effect is activated in the diffusion model, which is contrary to previous modeling efforts, where either diffusion is assumed to be infinite (Baccam et al., 2006) or isotropic following the Fickian law (Quirouette et al., 2020). The model by (Doceul et al., 2010) depends both on the Fickian diffusion as well as on the concentration of infected cells with the assumption that the flux of free viral particles depends on the concentration gradient of the virus and the concentration of infected cell:

$$\vec{J}_v = D_v(w)\Big(-\frac{\partial v}{\partial x}\Big), \tag{15}$$

where $D_v(w)$ is taken as an increasing function of w. Its form is given by

$$D_v(w) = D_{PCF} + g(w), \quad \text{with} \quad g(w) = \frac{aw}{b + w}, \tag{16}$$

where D_{PCF} is the Fickian diffusion coefficient of free virions, the parameter a represents saturation level for the infected cells and b indicates how quickly the cells get saturated. Setting $a = 0$ in $g(w)$ implies no infected cell and as a consequence there will be no repulsion effect. In the case of infected cells acting as chemorepellant for virus, the repulsion effect should force the viruses down the concentration gradient of infected cells, this is nicely captured in (Wang et al., 2017a,b) as a chemotactic flux of virus which depends on their own density, the density of infected cells, and the concentration gradient of infected cells.



(a) (b)

Figure 6: (a) Variation of $\widetilde{g}(w) = aw/(b+w)$ as function of w for $a = 4.5 \times 10^{-1}\mathrm{m}^2\mathrm{d}^{-1}$. (b) Contours of the spatial spread of V(t, x) at distinct mobility rates. The coloured areas mark the regions in space where the viral population size is larger than $\widetilde{V} = 0.1$ for various values of a. The boundaries of these areas have slopes equal to the rate of spread, which correspond to $c = 0.0019\ \mathrm{m}\cdot\mathrm{d}^{-1}$ for $a = 0.0$ (red), $c = 0.0073\ \mathrm{m}\cdot\mathrm{d}^{-1}$ for $a = 4.5 \times 10^{-6}\ \mathrm{m}^2\cdot\mathrm{d}^{-1}$ (pink), $c = 0.0214\ \mathrm{m}\cdot\mathrm{d}^{-1}$ for $a = 4.5 \times 10^{-5}\ \mathrm{m}^2\cdot\mathrm{d}^{-1}$ (orange), $c = 0.0529\ \mathrm{m}\cdot\mathrm{d}^{-1}$ for $a = 4.5 \times 10^{-4}\ \mathrm{m}^2\cdot\mathrm{d}^{-1}$ (black), and $c = 0.0726\ \mathrm{m}\cdot\mathrm{d}^{-1}$ for $a = 4.5 \times 10^{-3}\ \mathrm{m}^2\cdot\mathrm{d}^{-1}$ (green).

We adopt the framework of Quirouette et al. (2020) without the advection term and with the inclusion of density dependence in the diffusion term. We assume that the viruses can spread in a self-organized manner within a host and could be influenced by some components of the host environment (Galasso and Sharp, 1962). Thus, based on the considerations of previous efforts, we adopt a functional form for the density-dependent diffusion, which is similar to the model proposed by Lai and Zou (2014) as a basis for our present analysis. However, to be consistent with the dimensional analysis of our virus model, we interpret the density-dependent part, $g(w)$, of our diffusion model in terms of a Michaelis-Menten-type model (Ingalls, 2013):

$$\widetilde{g}(w) = \frac{aw}{b+w}, \tag{17}$$

where a is the maximum mobility rate in units of $\mathrm{m}^2\mathrm{d}^{-1}$ achieved by a virion due to its interaction with the infected cells, and b, the number of uninfected cells at which the mobility rate is half of a.

The function $g(w)$ is graphically presented in Figure 6(a) for variations of b with fixed a, and shows that the ability of virions to move due to interaction with infected cells is faster when there are lesser number of uninfected cells in their vicinity, this also means that higher mobility rates could result to faster rate of diffusion of free virions.

The spatial-temporal model that captures the influenza infection kinetics is a system of partial differential equations written as follows:

$$\begin{aligned}
\frac{\partial T(t,x)}{\partial t} &= -\beta T(t,x)V(t,x) + r_D D(t,x)\\
\frac{\partial E(t,x)}{\partial t} &= \beta T(t,x)V(t,x) - kE(t,x)\\
\frac{\partial I(t,x)}{\partial t} &= kE(t,x) - \delta I(t,x)\\
\frac{\partial V(t,x)}{\partial t} &= pI(t,x) - cV(t,x) - A(t) + \frac{\partial}{\partial x}\left(D_V(I)\frac{\partial V(x,t)}{\partial x}\right)\\
D(t,x) &= N(x) - T(t,x) - E(t,x) - I(t,x).
\end{aligned} \tag{18}$$

When virus becomes available to target cells $T(x,t)$ they become infected by virus $V(x,t)$ at a rate β and enter into the eclipse phase $E(x,t)$ where no viral production takes place. The eclipse cells turn into infectious cells $I(x,t)$ at a rate k which in turn produce virus at a rate p and die at a rate δ. The produced virus continues to infect new target cells and is cleared at a rate c due to loss of infectivity or non-specific clearance.

4.1 Spreading speed of influenza virus in $\mathbb{R}$

To quantify the spreading speed of waves, we consider computing the asymptotic spreading speed in $\mathbb{R}$. Since our model is based on a reaction-diffusion equation, we adopt the technique introduced by Neubert and Parker (2004) and which has also been used for a spreading virus population (Lai and Zou, 2014; Wang et al., 2017b). This technique requires fixing a threshold value, say $\widetilde{V}$, for the virus concentration $V(t,x)$, below which virus is not detectable. Then, we find the position, $\widetilde{x}(t)$, which corresponds to $\widetilde{V}$, i.e. $V(t,\widetilde{x}(t)) = \widetilde{V}$. The asymptotic spreading rate is then given by the rate of change of $\widetilde{x}(t)$ as t becomes very large, i.e. $c = \lim_{t\to\infty} \frac{d\widetilde{x}(t)}{dt}$, where c is the spreading speed (see Figure 6(b)).

The upper respiratory tract is represented as a one-dimensional grid with the top of the respiratory tract at $x = 0$ cm and the bottom at $x = 0.3$ m. In this model of the respiratory tract we include the conducting airways only up to the end of the bronchi while excluding the nostrils. The length of the naso-pharyngeal tract was estimated by dividing the volume by the surface area as estimated by Guilmette et al. (1997) for a human of 170 cm height, which is the height of the reference man (ICRP, 1975), the consecutive lengths of the nostrils, larynx, trachea, and bronchi were obtained from morphometric quantification performed by Raabe et al. (1976) and Stoneham (1993), the total length was found to be ≈ 0.3 m.

4.2 Computer realization

The objective of our numerical simulation experiments for the spatial-temporal model will be to better understand the dynamics of the influenza virus infection in-host for various values of the mobility rate a in the presence or absence of cellular regeneration and immune response respectively. For the simulations, we choose $b = 500$, the basic reproduction number

$$\mathcal{R}_0 = \left(\frac{\beta p N \delta}{c}\right) = \left(\frac{3.2\times10^{-5}\cdot 0.046\cdot 4\times10^{8}\cdot 5.2}{5.2}\right) = 5.89\times10^{2} \gg 1.$$

Result I: Effect of variations of mobility rates on infection dynamics with no cellular regeneration

In this simulation experiment, the mobility rate was varied in the absence of cell regeneration. Virus is first deposited initially at $x_d = 0.15$ m, $x_d = 0.01$ m and $x_d = 0.25$ m respectively (see Figure 7. The mobility rate is varied in increasing steps of one order of magnitude from 4.5×10^{-6} m$^2\cdot$d^{-1} to 4.5×10^{-1} m$^2\cdot$d^{-1}, and corresponds to the colors blue, green, cyan, magenta, dashed red and dashed blue (coincides with black line). We investigated the infection dynamics by plotting the fractions of target cells and the concentration of the virus for the variations of the mobility rates at different deposition depths. Spatial simulations show that virus becomes available to target cells gradually as it diffuses away from the site of deposition such that neighbouring cells become infected first, and further cells follow (see Figure 9. When compared to the ODE model, target cell infection takes place at a slower pace depending on the mobility rate. The extreme values $a = 0$ and $a = \infty$ corresponding to red and black lines respectively behave like the ordinary differential equations. Due to asymmetry in the site of deposition with respect to the respiratory tract when $x_d = 0.01$ m, virus diffuses out of the top of the respiratory tract before it gets to the bottom of the respiratory tract. This can be seen as a slight drop in the concentration of the virus followed by a decay once virus has crossed the lower end of the respiratory tract as well. When the target cells are completely exposed or infected, viral titer decay is observed and occurs sooner as mobility rate increases. When virus is deposited at $x_d = 0.15$ m, it reaches opposite ends of the respiratory tract at the same time and one phase of decay is observed, this is even more obvious when virus is deposited at $x_d = 0.25$ m.

Result II: Variations in mobility rates with cellular regeneration

The inclusion of the cellular regeneration into the spatial-temporal model represents a kind of reservoir of uninfectible and pluripotent cells that are responsible for repairing epithelial damage due to the influenza virus infection, which means that there is no damage that cannot be repaired by the reservoir cells. We find that a regeneration rate $r_d = 10d^{-1}$ is sufficient to maintain a chronic infection in the case of immediate regeneration of dead cells as shown in Figure 8 for an infection peaking at day 2 for extreme mobility rates of $a = 0$ and $a = \infty$ respectively. We observed that the fraction of target cells balances (or plateaus) after few days, the earliest was after 2 days. Cell regeneration has the effect of persisting the propagation of the virus population.

Result III: Effect of mobility rates in virus spreading

In the first simulation experiments, we investigate how the mobility rate affects the infection dynamics in the absence of cellular regeneration and so we focus on the structure of the emerging virus wave patterns. Most of the parameters used are adopted from Quirouette et al. (2020) as presented in the Table 4. The concentration of virus in the system was measured over a period of time. The virus was deposited initially at $x_d = 0.15$m and once infection and proliferation starts, the virions spread on either side of the respiratory tract depending on the value of the mobility rate a. We investigated the case when the mobility rate is zero ($a = 0$) (see Figure 9(a) and the case when the mobility rate is set to $a = 4.5\times10^{-6}$m^{2}d^{-1} (see Figure 9(c)). For better

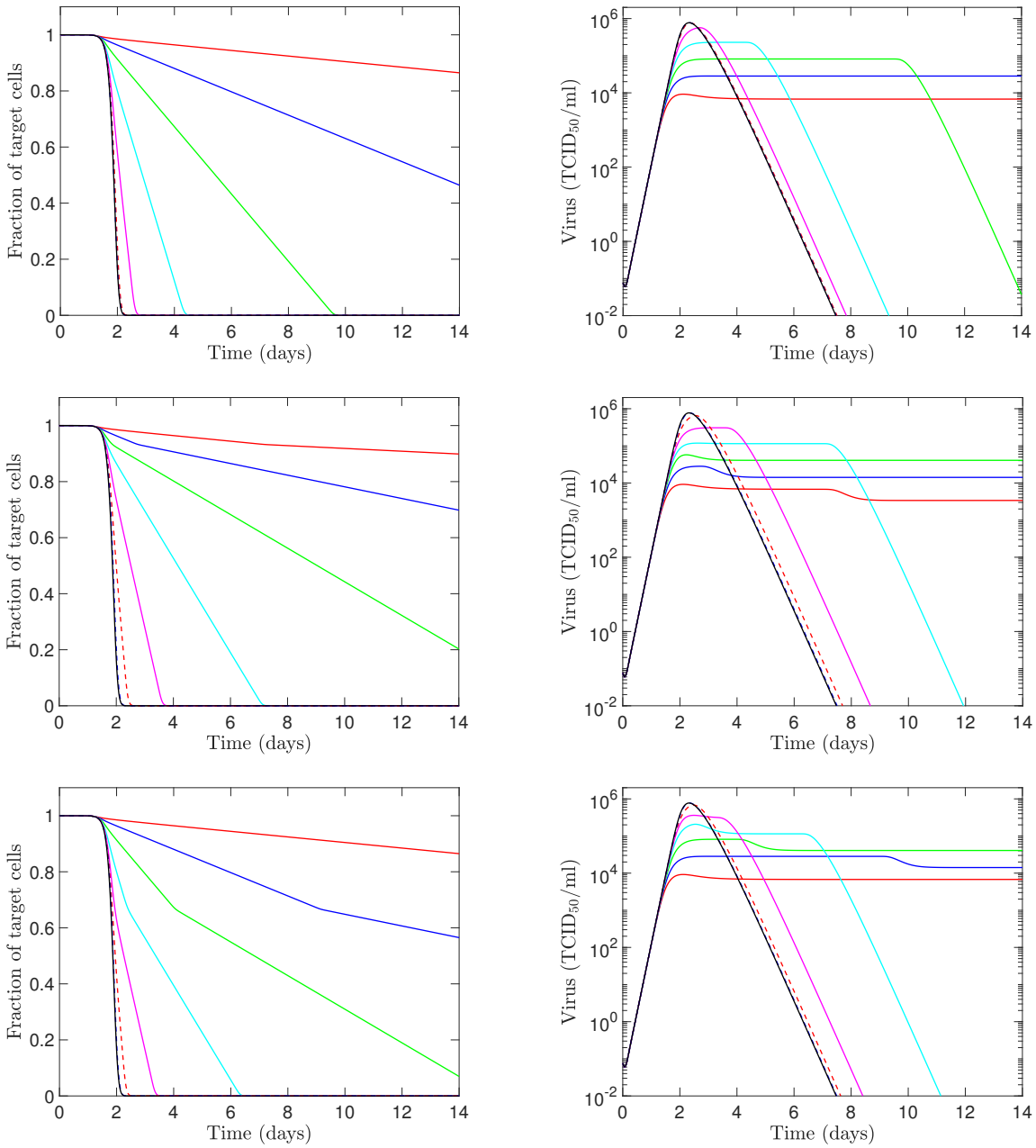


Figure 7: Effect of varying the mobility rate in the absence of cell regeneration and immune response. Virus is deposited initially at $x_d = 0.15$ m (top row), $x_d = 0.01$ m (middle row) and $x_d = 0.25$ m (bottom row). The ordinary differential equations (ODEs) assume infinite diffusion ($a = \infty$). The mobility rate is varied in increasing steps of one order of magnitude from 4.5×10^{-6} m$^2\cdot$d^{-1} to 4.5×10^{-1} m$^2\cdot$d^{-1}, and corresponds to the colors blue, green, cyan, magenta, dashed red and dashed blue (coincides with black line). The extreme values $a = 0$ and $a = \infty$ (ODE) correspond to red and black lines respectively.

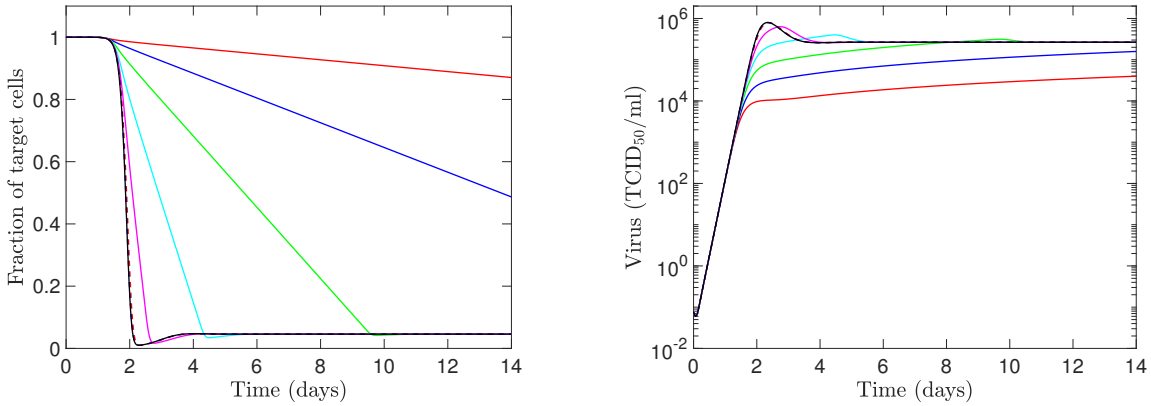


Figure 8: Effect of varying the mobility rate in the presence of cell regeneration. Virus is deposited initially at $x_d = 0.15$ m. The mobility rate is varied in increasing steps of one order of magnitude from 4.5×10^{-6} m$^2 \cdot$d^{-1} to 4.5×10^{-1} m$^2 \cdot$d^{-1}, and corresponds to the colors in blue, green, cyan, magenta, dashed red and dashed blue (coincides with black line). The extreme values $a = 0$ and $a = \infty$ (ODE) correspond to red and black lines respectively.

Table 4: Default initial conditions and parameter values of the target cell limited model. All parameters are from Baccam et al. (2006) except for the viral production p that was evaluated in this work.

symbol	parameter	value
T_0	number of initially available target cells	4×10^8 cells
E_0, I_0	number of initially infected cells	0
D_0	number of initially dead cells	0
V_0	initial viral inoculum	7.6×10^{-2}TCID$_{50}$/mL
$1/\tau_E$	lifespan of exposed cells E	6 h
$1/\tau_I$	lifespan of infectious cells I	4.6 h
c	viral clearance rate	4.6 h
β	infection rate of cells by virus	3.2×10^{-5} (TCID$_{50}$/mL)$^{-1} \cdot$d^{-1}
p	virus production rate	0.49 TCID$_{50}$/mL$\cdot$d^{-1}

comparison, the two simulation experiments where presented in one plots (see Figure 9(e)). By setting the mobility rate to zero, the wave front is narrow with delayed proliferation. On the other hand when the mobility rate is set to $a = 4.5 \times 10^{-6}$m^{2}d^{-1}, the wave front becomes wide with proliferation and spreading of the virus beyond the initial deposition dept. In the second simulation experiments, we investigate how the mobility rates affects the infection dynamics when there is cellular regeneration in the system, we investigate the structure of the emerging virus wave patterns. We also investigated the case when the mobility rate is zero ($a = 0$) (see Figure 9(b) and the case when the mobility rate is set to $a = 4.5 \times 10^{-6}$m^{2}d^{-1} (see Figure 9(d)). For better comparison, the two simulation experiments where presented in one plots (see Figure 9(f)). We observed that when the mobility rate is set to zero, the wave front is narrow but with proliferation underneath the waves, this is similar in the case of $a = 4.5 \times 10^{-6}$m^{2}d^{-1} except that the virus wave pattern is wide with rapid spreading of the virus beyond the initial deposition depth.

5 Sensitivity Analysis

Sensitivity analysis allows us to determine which parameters are most influential to the model output. We use this information to examine biological implications such as model corroboration, research prioritization, model simplification, identification of critical areas, and baseline parameter estimates (Saltelli et al., 2007, p. 34). In this study, we applied the Latin Hypercube Sampling (LHS) method which will give us the necessary information to rank our parameters' degree of importance through the Partial Rank Correlation Coefficient (PRCC) method. LHS is a stratified sampling without replacement technique, first introduced by McKay et. al. in 1979, that allows us to sample the entire parameter space more efficiently than similar Monte Carlo methods. LHS splits each of the k parameter distributions into N equally probable parts where N is at least greater than

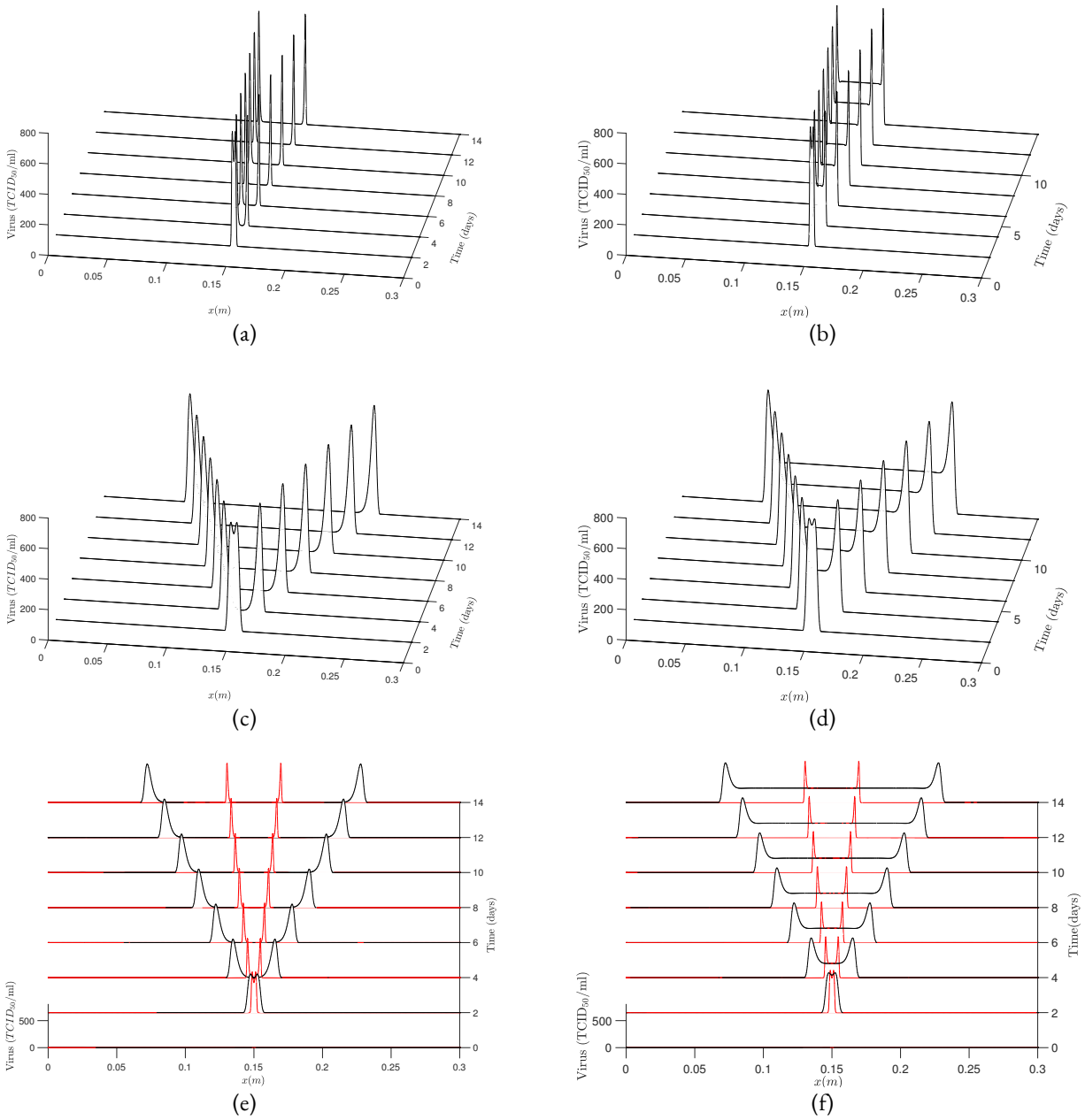


Figure 9: Spatial-temporal profiles of $V(t,x)$ for varying mobility rate. Virus is deposited initially at $x_d = 0.15$ and once infection and proliferation starts, the virions diffuse on either side of the RT depending on the value of a, this is shown in the absence of cellular regeneration (a,c,e) and with cellular regeneration (b,d,f). Comparison of the extent of spread in the two profiles shown in (a),(c) and (b), (d) re presented in (e) and (f) respectively. The wave profile in red indicates the case when $a = 0$ and in black when $a = 4.5 \times 10^{-6} \mathrm{m}^2\mathrm{d}^{-1}$.

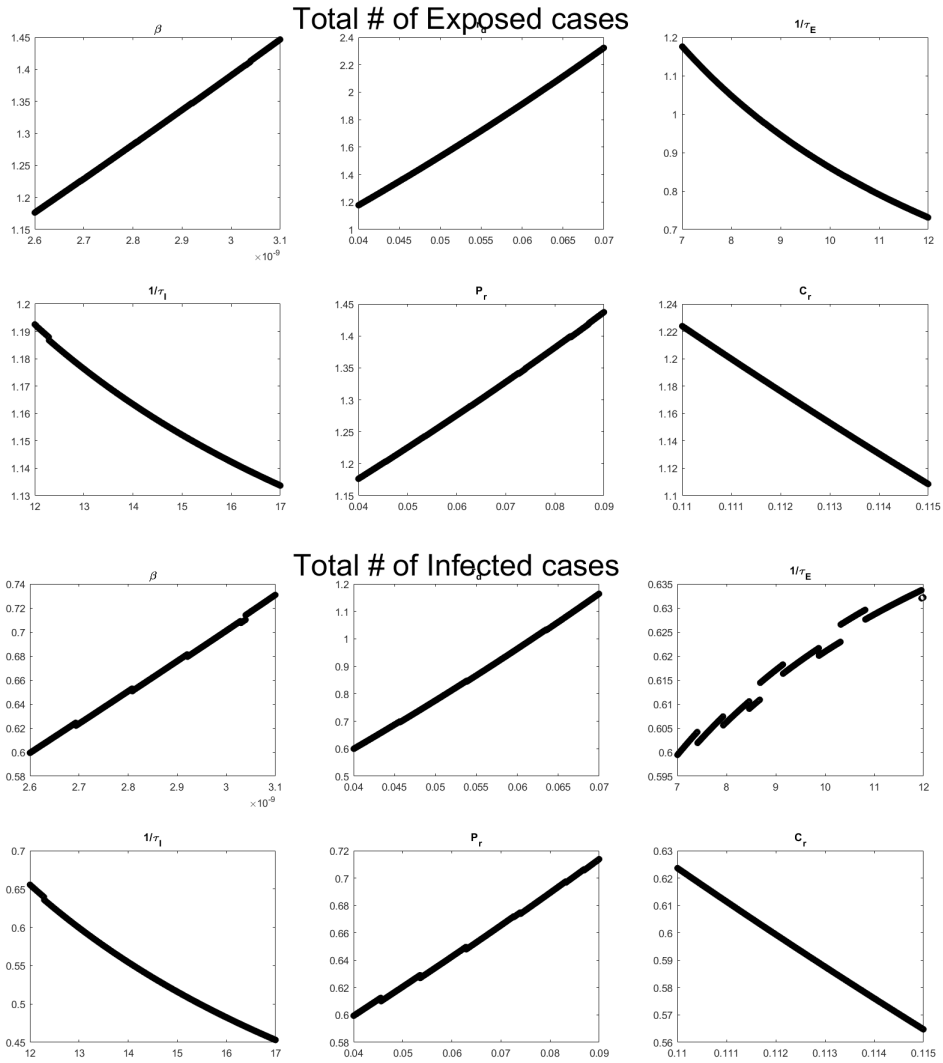


Figure 10: The title of the given section of graphs is the output measure for the given case and is along every y-axis. The parameter above every individual graph lies on the x-axis.

$k+1$, although it is often much greater for the sake of accuracy. These distributions are used to create the LHS input matrix. This matrix has N rows, one for each simulation, and k columns, one for each parameter. The k^{tb} entry of each row randomly selects an interval from the given parameter's probability distribution without replacement. This process is repeated to create N unique combinations of the parameters that spans the entire parameter space.

From this, a $N \times 1$ output matrix is generated where each entry is the output value from the corresponding simulation. Both the input and output matrices are then rank transformed according to the magnitude of the values along a column. The rank transformed matrices replace the raw data with values from one to N. We complete this transformation in order to compute the PRCC values which are explained below (Marino, 2008).

Before choosing PRCC, it is important to check for monotonicity between parameters and outputs. If we lack this, PRCC values are not accurate. If monotonicity does not hold, it is sometimes possible to truncate the LHS parameter ranges into monotonic regions (Gomero, 2012). PRCC uses the rank transformed data, not the raw data, to provide a measure of the linear association between a specific parameter and the output after the linear effects from the remaining inputs are removed. PRCC values range from -1 to 1, where a positive sign indicates a direct relationship between the parameter and the output value while a negative sign indicates an inverse relationship between the two. The magnitude of the PRCC value represents the importance of the parameter to the model output. The further the value is from zero, the more influential it is (Marino, 2008).

The *E Cases* output measure is defined by $\beta TV - \frac{E}{\tau_E}$. When we look at these graphs shown in Figure 10, we are considering

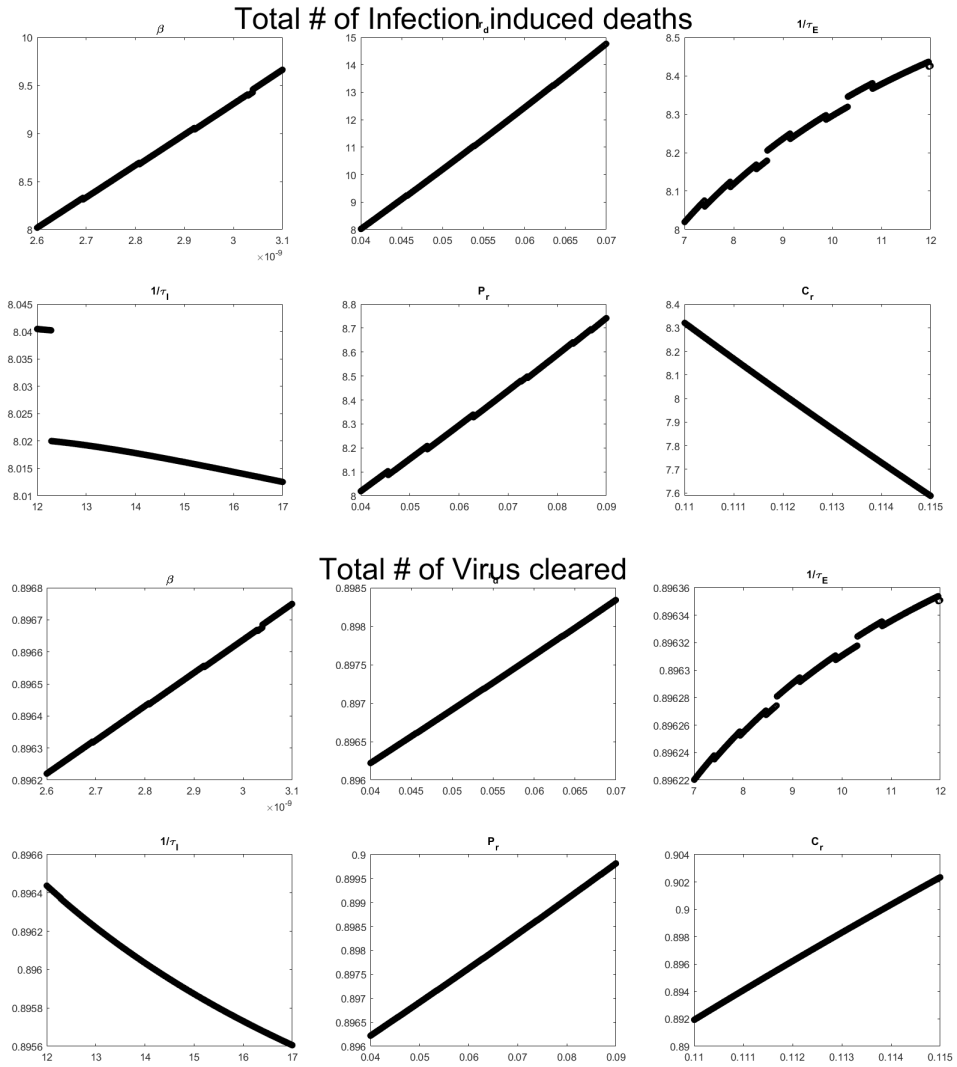


Figure 11: The set up is the same at the above graphs.

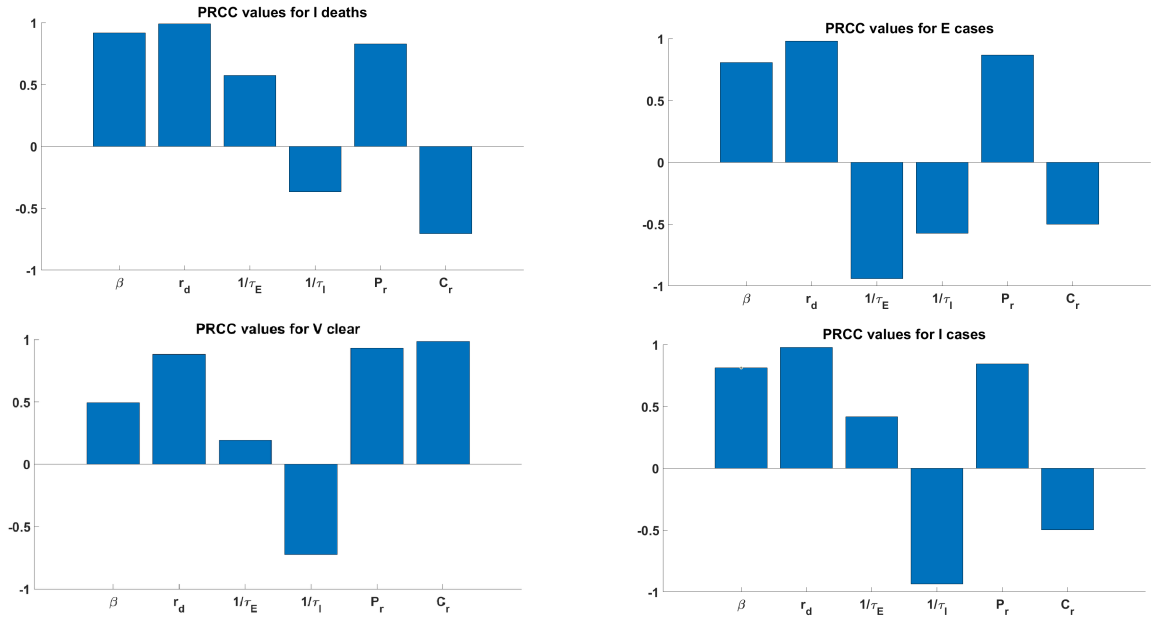


Figure 12: On the y-axis we have the PRCC values and along the x-axis each bar corresponds with the given parameter.

the relationship between the given parameter and the difference between cells entering the E compartment, βTV, and the cells exiting the E compartment, E/τ_E. The parameter lies on the x-axis in the shown graphs. It is important to note that all these plots are monotonic, so we can go on to consider the PRCC values as accurate. From Figure 12, the parameters' degree of influence on the *E Cases* output measure ranked from least to greatest are: $c < \frac{1}{\tau_I} < \beta < p < \frac{1}{\tau_E} < r_D$. We note that $1/\tau_E, 1/\tau_I$, and c have an inverse relationship with the output while β, r_D, and p have a direct relationship on the data. It is interesting to note that the parameters with positive PRCC values all fall into the numerator of $\mathcal{R}_0$ while the parameters with negative PRCC values fall into the denominator.

The number of infectious cases is measured by $\frac{E}{\tau_E} - \frac{I}{\tau_I}$, this is referred to as *I cases* in the associated plots shown in Figure 10. From Figure 12, the parameters that impact the number of infectious cases listed from least to most impact are: $\frac{1}{\tau_E} < c < \beta < p < \frac{1}{\tau_I} < r_D$, where the parameters r_D, p, and β are directly proportional to the number of infectious cases and the parameters $1/\tau_I$ and c are inversely proportional to the number of infectious cases. We notice that for the number of infectious cases, parameters that are directly proportional to *I cases* that also appear in $\mathcal{R}_0$ are the parameters that appear in the numerator of $\mathcal{R}_0$; namely p and β. Similarly, the parameters that are inversely proportional to *I cases* that also appear in $\mathcal{R}_0$ are the parameters that appear in the denominator of $\mathcal{R}_0$; specifically $1/\tau_I$ and c.

The number of infection induced deaths is calculated using I/τ_I which is shown in Figure 11. The parameters that impact the number of infection induced deaths, written as *I deaths*, ranked from least to most impactful to get $\frac{I}{\tau_I} < \frac{1}{\tau_E} < c < p < \beta < r_D$, this is again from Figure 12. Parameters that are directly proportional to the number of infection induced deaths are r_D, β, p, and $1/\tau_E$, while parameters that are inversely proportional to the number of infection induced deaths are c and $1/\tau_I$. The same relationship can be seen between the proportionality of parameters and $\mathcal{R}_0$ in *E cases* and *I cases* can be seen in *I deaths*.

The number of virons cleared is given by cV, shown in Figure 11, is effected by the following parameters, ranking from least to the most impactful, as shown in Figure 12: $\frac{1}{\tau_E} < \beta < \frac{1}{\tau_I} < r_D < p < c$. The parameters c, p, r_D, β, and $1/\tau_E$ are all directly proportional to the amount of virus cleared, whereas $1/\tau_I$ is inversely proportional to the number of virons cleared. Since c is directly proportional to this measure, we do not see the same pattern related to $\mathcal{R}_0$ as we did in the previous measures. This change follows our assumption, however, due to the amount of virus cleared not being directly related to the trajectory of the virus. When *V clear* is very high, it could either be due to the host being very efficient at clearing out the virus, or it could be because the host is sustaining the virus for a long time. Thus, the value of *V clear* is not clearly related to whether the state of the virus is approaching endemic or disease-free equilibrium.

Sensitivity analysis revealed that r_D, p, and β are directly proportional to all four measures of the disease. $1/\tau_I$ was found to be inversely proportional to all four measures of the disease, and clearance rate, c, was found to be inversely proportional to the number of exposed cases, the number of infectious cases, and the number of disease-induced deaths. In the monotonicity plots, we can observe that the plot corresponding to $1/\tau_E$ is not strictly monotonic. Sensitivity analysis using PRCC requires

monotonicity, thus this may be a source of error in our analysis. The ranking of how much each parameter affects the number of exposed cases, infectious cases, cell deaths, and virons cleared is consistent regardless of the number of runs used. However, the magnitude of the PRCC values does change depending on the number of runs used. In some cases, this difference is extreme; such as when comparing the PRCC values of $1/\tau_E$ and β for the number of virons cleared using 100 runs to the PRCC values of $1/\tau_E$ and β for the number of virons cleared using 2000 runs. When using 100 runs, $1/\tau_E$ has a PRCC value of 0.033 and β has a PRCC value of 0.523, making for a 0.490 difference. When using 2000 runs, $1/\tau_E$ has a PRCC value of 0.192 and β has a PRCC value of 0.493, making a difference of 0.301. The degree to which the difference between the PRCC values of $1/\tau_E$ and β changes depending on the number of runs is concerning and may be the result of the non-monotonicity of $1/\tau_E$. Since the largest number of simulations used (in this case $N = 2000$) should be the most accurate, we will examine the results when that number of runs is used. For information regarding the 100, 500, and 1000 runs, see the supplementary materials for this article.

For 2000 simulations, the PRCC values of each parameter all have $p < 1.2 \cdot 10^{-9}$, so each parameter significantly effects the number of expose cases, infectious cases, deaths due to infection, and virons cleared.

6 Discussion

The Basic Reproduction Number, $\mathcal{R}_0$ was calculated to be $\beta pN/\left(c\frac{1}{\tau_I}\right)$. The numerator of $\mathcal{R}_0$, βpN, contains values related to the transmission and production of the virus, and the denominator, $c\frac{1}{\tau_I}$, contains values related to the death and clearance of the virus. Thus, when the production and transmission of the virus is greater than the death and clearance of the virus, we see a spread in infectious and a tendency to the locally asymptotically stable endemic equilibrium. On the other hand, when production and transmission is less than the death and clearance of the virus, the disease tends to the locally asymptotically stable disease-free equilibrium. The goal in influenza treatment then, would be to manipulate the conditions of the disease in order to produce a situation that approaches the disease-free equilibrium. Thus, it is important to determine what parameters contribute the most to hallmarks of the disease.

Partial Rank Correlation Coefficients were used to determine how different parameters contributed to important measures of the disease including, the number of exposed cases, the number of infectious cases, the number of infection-induced deaths, and the number of virons cleared. Health professionals would ideally like to minimize the first three of these measures. Interestingly, it was found that cellular restoration had the strongest positive relationship with the number of exposed cases, the number of infectious cases, and the number of cell deaths due to infection. This finding may indicate that in some cases it would be helpful to slow down cellular restoration as it is positively associated with the measures we wish to minimize. No causation is established in this study, however, the detrimental role of cellular restoration on the trajectory of the spread of influenza has been documented in other cases (Deecke and Dobrovolny, 2018). In particular, chronic cases of influenza cannot occur without cellular restoration, so this parameter clearly plays a role in the perpetuation of the influenza infection.

It was also noted from the sensitivity analysis that parameters that were directly proportional to the number of exposed cases, the number of infectious cases, and the number of disease-induced deaths tend to appear in the numerator of $\mathcal{R}_0$, whereas parameters that are indirectly proportional to these measures tend to appear in the denominator of $\mathcal{R}_0$. This makes sense because each of these three measure impact the trajectory of the disease (whether it tends to endemic state or disease-free), which is precisely what $\mathcal{R}_0$ measures.

Results from the stochastic simulations indicate that influenza first spreads extremely rapidly, achieving a state in which all target cells are either exposed or infectious very quickly. Comparatively, it takes a much longer time for the exposed and/or infectious cells to die. To illustrate; for a population of 5 cells with initially 0 exposed/infectious cells it takes approximately 10-11 seconds for all the original target cells to become exposed/infectious, and it is not until at least 3 minutes have passed since the initial time that the first cells begin to die. These findings indicate that any treatment for influenza intended to curtail the spread of the virus within the host would need to be administered and be effective very soon after initial exposure to the influenza virus. These findings are supported by biological evidence that early administration of antiviral treatment for patients with influenza is an important factor of recovery (Zheng et al., 2018).

From the simulation experiments of the spatial-temporal model, we conclude that (i) For any fixed b, the wave speed depends on the mobility rate, a; (ii) as $a \to \infty$, the virus reaction-diffusion model is equivalent to the corresponding ODE model; (iii) the initial deposition position, x_d, affects the infection kinetics, but has no consequence when a is large; (iv) increase in diffusion rate increases the consumption of target cells; and (v) the wave profiles of the solutions in Figure 9 suggest the existence of traveling wave solutions.

7 Future Directions

In the future, it would be important to compare the models developed in this article with data collected regarding the spread of influenza within-host. Doing so would allow us to verify the accuracy of our model as well as refine our models using parameter

estimation. Currently, we are using values for our parameters that previous studies have used in their models (Bai et al., 2019; Beauchemin and Handel, 2011a; Quirouette et al., 2020); however, it may be more beneficial to use collected biological data to estimate our parameters through a process known as parameter estimation. It would be possible to utilize the results from the stochastic simulations to determine the time until peak infection (when the maximum number of exposed and infectious cells is achieved). This time appears to depend both on the number of cells present, N, as well as the initial state. In particular, it may be useful to examine how the initial number of infectious cells impacts the time until peak infection and if/how the relationship is mediated by N. The tools developed in this study are sufficient to investigate this question, but due to the computationally expensive nature of the simulations, this question remains unanswered.

References

Allen, L. J. (2010). *An Introduction to Stochastic Processes with Applications to Biology, Second Edition* (Second ed.). Chapman and Hall/CRC. 235

Allen, L. J. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *2*(2), 128–142. 230, 233, 234

Allen, L. J. and A. M. Burgin (2000). Comparison of deterministic and stochastic sis and sir models in discrete time. *Mathematical Biosciences 163*, 1–33. 234

Baccam, P., C. Beauchemin, C. A. Macken, F. G. Hayden, and A. S. Perelson (2006). Kinetics of influenza a virus infection in humans. *80*(15), 7590–7599. 229, 241, 245

Bai, F., K. E. S. Huff, and L. J. S. Allen (2019). The effect of delay in viral production in within-host models during early infection. *Journal of Biological Dynamics 13*(sup1), 47–73. PMID: 30021482. 234, 251

Beauchemin, C. A. and A. Handel (2011a). A review of mathematical models of influenza a infections within a host or cell culture: lessons learned and challenges ahead. *11*(1), S7. 251

Beauchemin, C. A. A. and A. Handel (2011b). A review of mathematical models of influenza a infections within a host or cell culture: lessons learned and challenges ahead. *BMC public health 11*(S1), S7. 229

Boianelli, A., V. K. Nguyen, T. Ebensen, K. Schulze, E. Wilk, N. Sharma, S. Stegemann-Koniszewski, D. Bruder, F. R. Toapanta, C. A. Guzmán, et al. (2015). Modeling influenza virus infection: a roadmap for influenza research. *Viruses 7*(10), 5274–5304. 229

Deecke, L. and H. M. Dobrovolny (2018). Intermittent treatment of severe influenza. *442*, 129–138. 250

Doceul, V., M. Hollinshead, L. van der Linden, and G. L. Smith (2010). Repulsion of superinfecting virions: a mechanism for rapid virus spread. *Science 327*(5967), 873–876. 241

Edholm, C. J., B. O. Emerenini, A. L. Murillo, O. Saucedo, N. Shakiba, X. Wang, L. J. S. Allen, and A. Peace (2018). Searching for superspreaders: Identifying epidemic patterns associated with superspreading events in stochastic models. In A. Radunskaya, R. Segal, and B. Shtylla (Eds.), *Understanding Complex Biological Systems with Mathematics*, Association for Women in Mathematics Series, pp. 1–29. Springer International Publishing. 234

Galasso, G. and D. Sharp (1962). Virus particle aggregation and the plaque-forming unit. *The Journal of Immunology 88*(3), 339–347. 242

Gomero, B. (2012). Latin hypercube sampling and partial rank correlation coefficient analysis applied to an optimal control problem. 247

Guilmette, R. A., Y. S. Cheng, and W. C. Griffith (1997). Characterising variability in adult human nasal airway dimensions. *Ann. occup. Hyg 41*, 491–496. 243

Hordijk, A., D. L. Iglehart, and R. Schassberger (1976). Discrete time methods for simulating continuous time markov chains. *8*(4), 772–788. 233

ICRP (1975). Report of the task group on reference man. (23). 243

Ingalls, B. P. (2013). *Mathematical modeling in systems biology: an introduction*. MIT press. 242

Ivanov, I. and E. R. Dougherty (2006). Modeling genetic regulatory networks: continuous or discrete? *14*(2), 219–229. 234

Kakizoe, Y., S. Nakaoka, C. A. Beauchemin, S. Morita, H. Mori, T. Igarashi, K. Aihara, T. Miura, and S. Iwami (2015). A method to determine the duration of the eclipse phase for in vitro infection with a highly pathogenic shiv strain. *Scientific reports 5*(1), 1–14. 229

Kanyiri, C. W., K. Mark, and L. Luboobi (2018). Mathematical analysis of influenza a dynamics in the emergence of drug resistance. *Computational and mathematical methods in medicine 2018*. 229

Keeling, M. J. and P. Rohani (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press. OCLC: ocn163616681. 233

Khanh, N. H. (2016). Stability analysis of an influenza virus model with disease resistance. *Journal of the Egyptian Mathematical Society 24*(2), 193–199. 229

Lai, X. and X. Zou (2014). Repulsion effect on superinfecting virions by infected cells. *Bulletin of mathematical biology 76*(11), 2806–2833. 241, 242, 243

Malik, T. (2016). Discrete time markov chain of a dynamical system with a rest phase. *International Journal of Applied Nonlinear Science 2*, 137. 230, 233, 234

Marino, S. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology 254*, 178–196. 247

Matrosovich, M., T. Matrosovich, J. Uhlendorff, W. Garten, and H.-D. Klenk (2007). Avian-virus-like receptor specificity of the hemagglutinin impedes influenza virus replication in cultures of human airway epithelium. *361*(2), 384–390. 229

Matrosovich, M. N., T. Y. Matrosovich, T. Gray, N. A. Roberts, and H.-D. Klenk (2004). Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *101*(13), 4620–4624. 229

Neubert, M. G. and I. M. Parker (1989). *Theory of impulsive differential equations*, Volume 6. World scientific. 232

Neubert, M. G. and I. M. Parker (2004). Projecting rates of spread for invasive species. *Risk Analysis: An International Journal 24*(4), 817–831. 243

Nicholls, J. M., M. C. W. Chan, W. Y. Chan, H. K. Wong, C. Y. Cheung, D. L. W. Kwong, M. P. Wong, W. H. Chui, L. L. M. Poon, S. W. Tsao, Y. Guan, and J. S. M. Peiris (2007). Tropism of avian influenza a (h5n1) in the upper and lower respiratory tract. *13*(2), 147. 229

Perelson, A. S., D. E. Kirschner, and R. De Boer (1993). Dynamics of hiv infection of cd4+ t cells. *Mathematical biosciences 114*(1), 81–125. 229

Quirouette, C., N. P. Younis, M. B. Reddy, and C. A. A. Beauchemin (2020). mathematical model describing the localization and spread of influenza a virus infection within the human respiratory tract. Master's thesis. 241, 242, 243, 251

Raabe, O., H. Yeh, G. Schum, and R. F. Phalen (1976). Tracheobronchial geometry: Human, dog, rat, hamster. 243

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2007). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd. 245

Sandmann, W. (2008). Discrete-time stochastic modeling and simulation of biochemical networks. *32*(4), 292–297. 233, 234

Stoneham, M. D. (1993). The nasopharyngeal airway, assessment of position by fibreoptic laryngoscopy. *48*, 575–580. 243

Taubenberger, J. K. and D. M. Morens (2008). The pathology of influenza virus infections. *Annual review of pathology 3*, 499–522. 230

van den Driessche, P. and J. Watmough (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci. 180*, 29–48. 231

van Riel, D., V. J. Munster, E. de Wit, G. F. Rimmelzwaan, R. A. M. Fouchier, A. D. M. E. Osterhaus, and T. Kuiken (2007). Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *171*(4), 1215–1223. 229

Wang, W., W. Ma, and X. Lai (2017a). A diffusive virus infection dynamic model with nonlinear functional response, absorption effect and chemotaxis. *Communications in Nonlinear Science and Numerical Simulation 42*, 585–606. 241

Wang, W., W. Ma, and X. Lai (2017b). Repulsion effect on superinfecting virions by infected cells for virus infection dynamic model with absorption effect and chemotaxis. *Nonlinear Analysis: Real World Applications 33*, 253–283. 241, 243

Wiggins, S. (2003). *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Second ed.), Volume 2 of *Texts in Applied Mathematics*. Springer. 231, 232

Yan, A. W. C., P. Cao, and J. M. McCaw (2016). On the extinction probability in models of within-host infection: the role of latency and immunity. *73*(4), 787–813. 234

Yao, L., C. Korteweg, W. Hsueh, and J. Gu (2010). Avian influenza receptor expression in h5n1-infected and noninfected human tissues. *22*(3), 733–740. 229

Zanin, M., P. Baviskar, R. Webster, and R. Webby (2016). The interaction between respiratory pathogens and mucus. *Cell host & microbe 19*(2), 159–168. 229

Zheng, S., L. Tang, H. Gao, Y. Wang, F. Yu, D. Cui, G. Xie, X. Yang, W. Zhang, X. Ye, Z. Zhang, X. Wang, L. Yu, Y. Zhang, S. Yang, W. Liang, Y. Chen, and L. Li (2018). Benefit of early initiation of neuraminidase inhibitor treatment to hospitalized patients with avian influenza a(h7n9) virus. *66*(7), 1054–1060. 250