Mathematical modelling of bacterial and viral infections
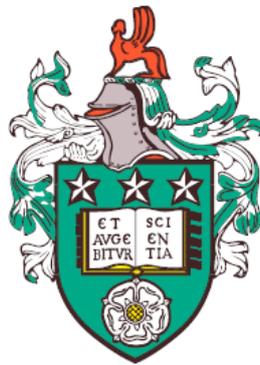
Jonathan Edward Carruthers

Department of Applied Mathematics

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

February, 2019

The candidate confirms that the work submitted is his own except where work which has formed part or jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Jonathan Edward Carruthers to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Joint Publications

All of the work in Chapter 3 has already been refereed and published, as follows:

- **Carruthers J**, López-García M, Gillard J.J, Laws T.R, Lythe G, Molina-París C, (2018). A novel stochastic multi-scale model of *Francisella tularensis* infection to predict risk of infection in a laboratory, *Frontiers in Microbiology*, **9**: 1165.

In addition, the work in Chapter 4 and Chapter 5 is in preparation:

- **Carruthers J**, Lythe G, López-García M, Gillard J.J, Laws T.R, Lukaszewski R, Molina-París C. Stochastic dynamics of intracellular infection and replication. *PLoS Computational Biology, to be submitted*.

- **Carruthers J**, Liao L.E, López-García M, Hiscox J.A, García-Dorival I, Smither S, Laws T.R, Lythe G, Molina-París C. A mathematical model for wild-type Ebola virus infection. *PLoS Computational Biology, to be submitted*.

Acknowledgements

Firstly, I would like to thank my supervisors, Carmen Molina-Parìs and Grant Lythe, who first introduced me to the field of mathematical biology during my undergraduate studies and have provided continuous support ever since. I would also like to thank Martin López-García who has offered valuable help and encouragement, particularly with the analysis of stochastic processes.

From Dstl, I would like to thank Joseph Gillard, whose enthusiasm and help in directing the research has always been appreciated, and Thomas Laws, whose knowledge of bacterial and viral infections has provided a basis for all of the mathematical models.

I would also like to thank John Paul Gosling for taking the time to discuss and advise about the Bayesian inference methods used throughout this thesis.

Finally, I would like to thank my friends and family, especially Helen, who have supported me throughout my studies.

Abstract

In this thesis, I present a series of mathematical models for describing the infection dynamics of two lethal intracellular pathogens, namely *Francisella tularensis* and Ebola virus. Through the construction of both stochastic and deterministic models, a multi-disciplinary approach is used to account for the key processes that dictate disease progression. Each model describes the dynamics of a population of host cells, with the analysis of interactions between individual cells and single bacteria or virus particles allowing for greater detail to be included. Throughout this thesis, the theory of birth-and-death processes, matrix analytic methods and numerical simulation algorithms are used to study the stochastic processes. In addition to these techniques, a Bayesian approach to statistical inference is applied in order to parametrise each of the models with relevant infection data.

Contents

CONTENTS

CONTENTS

List of Figures

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

1.1 Biological introduction

The first lines of defence against pathogens are physical barriers such as the skin and mucosa. Following the entry of a pathogen into the body, the second line of defence is provided by the innate immune response (Sompayrac (2008)). Although the innate immune response is a network of different cell types and proteins that depend on each other to quickly and effectively contain infection, the single most important cell is the macrophage (Murray & Wynn (2011)). Present in tissues throughout the body, macrophages detect foreign bodies and take them up into phagosomes through a process known as phagocytosis. Once internalised, these phagosomes fuse with lysosomes that contain chemicals capable of destroying the pathogen contained within (Underhill & Goodridge (2012)). In addition to actively killing invading pathogens, macrophages also produce cytokines, proteins that facilitate communication between cells of the immune system. It is these signals that recruit cells such as neutrophils to sites of infection (Murphy & Weaver (2016)). Neutrophils are the most abundant leukocyte in the body and are short-lived, surviving for around five days after they leave the bone marrow. Similar to macrophages, neutrophils phagocytose pathogens, exposing them to lethal chemicals. However, as neutrophils often release biocidal chemicals extracellularly, there is always damage to host tissues through this killing (Bardoel *et al.* (2014)). For this reason it is important that the turnover

of neutrophils is controlled, otherwise neutrophils with a prolonged lifespan adopt a pro-inflammatory phenotype and cause further damage (Kolaczkowska & Kubes (2013)). A second role of macrophages is in presenting peptide molecules in order to initiate an adaptive immune response. Whilst the innate immune response acts immediately and is sometimes capable of resolving infection, in some cases the infection can only be contained and an adaptive immune response is required to clear the pathogen (Trinchieri (2003)). Although adaptive immune responses often take days to set in, the dependency on innate immune cells to present antigen results in a response that is specific to the invading pathogen (Hoebe et al. (2004)). Furthermore, adaptive immune responses have the advantage that they can also provide immunity to secondary infections.

In the chapters that follow, two specific pathogens are considered, the bacterium *Francisella tularensis* and the Ebola virus. An overview of each pathogen is provided in the following sections.

1.1.1   Francisella tularensis

Francisella tularensis is a gram-negative, facultative bacteria and the causative agent of tularemia (Oyston et al. (2008)). Of the four subspecies, *F. tularensis* subspecies *tularensis* is the most lethal, and the SCHU S4 strain of this subspecies is considered in Chapters 3 and 4. *F. tularensis* is currently listed as a category A bioterrorism agent by the Centers for Disease Control and Prevention (CDC), indicating that it can be easily disseminated and has the potential to have a major impact on public health (Center for Preparedness and Response (2017)). Previously, the concept of weaponising *F. tularensis* involved dispersing the bacterium as an aerosol that could be used to infect a population over a large area (Christopher et al. (1997)). For this reason, it is particularly important to study respiratory tularemia, with this route of infection also resulting in the highest case fatality rate of up to 30% when left untreated. The case fatality rate is, however, below 2% following antibiotic treatment. The main reason for *F. tularensis* being classified as a category A bioterrorism agent is due to the striking feature that the initial dose required for a 50% probability of infection, known
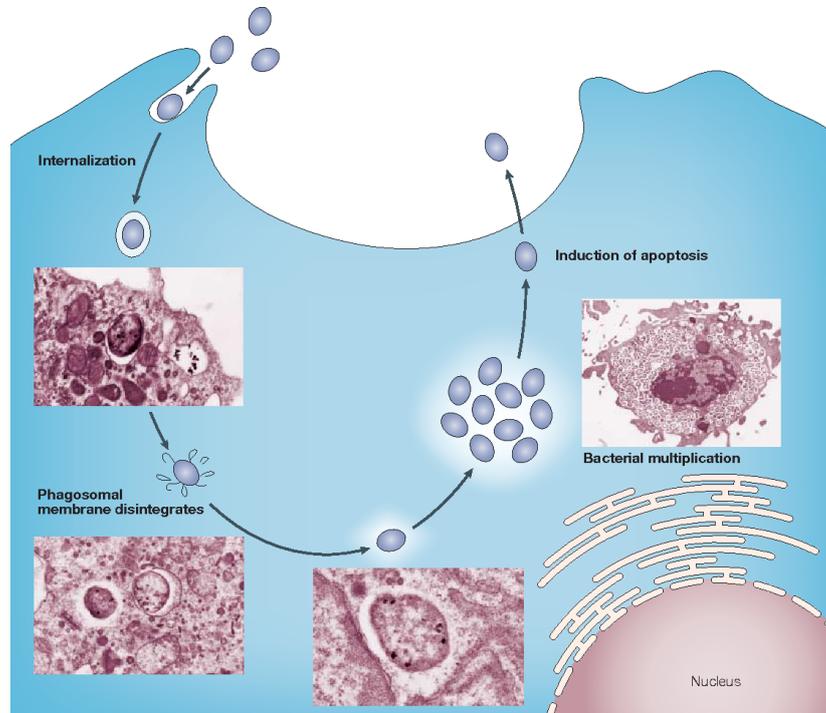
Figure 1.1: Intracellular life cycle of a *F. tularensis* bacterium by Oyston *et al.* (2004)

as the 50% infectious dose (ID_{50}), is fewer than 10 colony forming units (CFUs). Here, colony forming units are a measure of viable bacteria. As a comparison, between 8,000 and 50,000 spores must be inhaled to produce pulmonary anthrax (Oyston *et al.* (2004)). This, coupled with the method of dissemination, has the potential to result in a large number of cases, causing a strain on health services to provide antibiotic treatment quickly.

Following inhalation of *F. tularensis* bacteria, the primary cell type that is initially infected is alveolar macrophages, with this cell type representing almost 80% of all infected cells in the lungs of mice after 24 hours (Hall *et al.* (2008)). Following phagocytosis, bacteria are able to escape the phagosome before fusion with lysosomes can occur, thus preventing their killing. Once bacteria are present in the cytosol, they undergo multiple rounds of replication until the rupturing of the cell releases these bacteria into the extracellular environment, where they can continue to infect cells and disseminate to other organs (Jones *et al.* (2012)). The

intracellular life cycle of *F. tularensis* is depicted in Figure 1.1. For approximately 72 hours, these events go undetected by the host immune response, however, this is then followed by a sharp increase in the production of pro-inflammatory cytokines such as IFN-γ and TNF-α. This sudden upregulation of pro-inflammatory cytokines is reminiscent of a 'cytokine storm', preventing the clearance of infection and inducing widespread necrosis that is instead damaging to the host (D'Elia *et al.* (2013)).

Neutrophils are among the first circulating immune cells to be recruited to the site of infection and are present in large numbers in the lungs of mice, comprising over half of all infected cells at three days post infection (Hall *et al.* (2008)). However, instead of helping to resolve infection, *F. tularensis* bacteria infect neutrophils and prolong the lifespan of these usually short lived cells to 48 hours, such that they adopt a pro-inflammatory phenotype and contribute to further damage to the host (Allen (2013); Schwartz *et al.* (2012)). The unregulated inflammatory response results in widespread organ and system failure and is ultimately the cause of death.

1.1.2 Ebola virus

Ebola virus (EBOV) is a negative-sense single-strand RNA virus that causes Ebola virus disease (EVD). This lethal disease often results in severe haemorrhagic fever characterised by multiple organ failure and has a case fatality rate of between 45% and 90% (Prescott *et al.* (2017)). Most significantly, a large scale epidemic in West Africa between 2014 and 2016 resulted in almost 30,000 cases with over 11,000 fatalities (CDC (2017)). Although this case fatality rate is not consistent with the range previously quoted, this is likely due to the difficulty in calculating such rates. It is often not possible to accurately estimate the number of individuals who succumb to infection and not due to other causes. Furthermore, factors such as access to healthcare can lead to vastly different case fatality rates between populations. Despite this, the scale of the West African epidemic is indicative of the infectious nature of EBOV and, along with the high mortality rate, is the reason why EBOV is also classified as a category A bioterrorism agent by the CDC.

1. INTRODUCTION

The structure of an EBOV virus particle consists of an RNA molecule that comprises seven genes, each encoding a different viral protein. A depiction of this structure is provided in Figure 1.2. Nucleoprotein (NP) and VP30 form the capsid that contains the RNA molecule and together are referred to as the nucleocapsid. The L and VP35 proteins form the polymerase complex which transcribes and replicates the EBOV genome, whilst VP24 and VP40 are involved in the assembly and budding of new virus particles. Glycoproteins (GP) form 'spikes' on the envelope of the virus particle that are important for virus entry into cells. Figure 1.2 depicts the intracellular life cycle of an Ebola virus particle. The virion first enters cells through a vesicle by endocytosis or macropinocytosis. Following acidification of the vesicle, its membrane and that of the virus fuse to allow the release of the nucleocapsid into the cytosol and transcription of the viral genome can begin. Initially, transcription of the viral genome from a negative to positive sense enables viral protein synthesis, particularly NP, by host cell ribosomes. Accumulation of NP then triggers a switch from transcription to replication. Due to a lack of proofreading capabilities, the RNA polymerase is notoriously error-prone when replicating the viral genome. This has advantages for the virus, allowing genetically diverse populations to form that enable the virus to adapt to different environments (Barr & Fearns (2010)). However, it can also result in the formation of defective particles that may have deficiencies in both their ability to enter host cells, as well as their ability to replicate once inside. When sufficient levels of nucleocapsids and envelope-associated proteins have been synthesised, virus assembly occurs at the cell membrane and is followed by the release of virus particles (Knipe et al. (2007)).

During human infection, EBOV initially associates with macrophages, with this cell type able to support high levels of viral replication. Infected macrophages are activated and produce large amounts of pro-inflammatory cytokines, such as IFN-γ, but are impaired in their ability to present antigen (McElroy et al. (2018)). Despite this, fatal cases of EVD show indicators of early T-cell activation, with increased levels of IFN-γ and IL-2. However, this early activation is followed by a sharp increase in lymphocyte apoptosis which is unusual given that lymphocytes themselves have not been shown to become infected. Instead it is believed

Figure 1.2: Intracellular life cycle of an Ebola virus particle by Messaoudi *et al.* (2015)

that increased levels of the anti-inflammatory cytokine IL-10 might prevent a sustained T-cell response. On the other hand, non-fatal cases correlate to a balanced immune response, with lower but more robust levels of T-cell activation in the earlier stages of infection. Ultimately, death of the host is likely to result from the build up of necrotic debris as a result of high levels of virus replication, and coagulation abnormalities due to the pro-inflammatory response (Prescott *et al.* (2017)).

1.2   Objectives of this thesis

It is evident, given the description of *F. tularensis* and Ebola virus, that both pathogens pose a significant risk to a population. The aim of this thesis is therefore to use mathematical modelling as an approach to help mitigate this risk. With *F. tularensis*, when considering isolated cases, the majority of infections can be resolved following antibiotic treatment. However, since infection is still

lethal in approximately 2% of treated individuals, this suggests that early detection and administration of treatment is required to ensure survival. The difficulty therefore arises when considering a population of individuals who are exposed to airborne *F. tularensis* bacteria. To optimise treatment strategies, it is first necessary to understand the amount of bacteria each individual is exposed to. Provided with this information, it must then be determined how likely it is that the immune response of the individual is capable of clearing the infection and, in the case where it cannot, the time period during which treatment will remain most effective. The answers to these questions are best provided by a dose-and-time response model. For this reason, a stochastic within-host model of *F. tularensis* infection that can describe this dose dependency is introduced in Chapter 3. This then forms the central level of a multi-scale model that connects the intracellular, within-host and population-level scales, thereby using an understanding of the exact biological mechanisms to predict the effects on a population. Where possible, stochastic descriptors offer an approach to analytically study the dynamics of the model that does not require the solution of the often intractable Kolmogorov differential equations (Section 2.2.4). The matrix analytic methods that allow these descriptors to be evaluated have become increasingly effective in recent years, due in part to computational improvements, and provide an efficient algorithmic approach that exploits specific underlying structures of the stochastic process (Section 2.3). However, when the space of states becomes too large, even these algorithms become difficult to implement and approximative simulation techniques are instead employed (Section 2.4). The use of *in vitro* and *in vivo* infection data allows each level of the model to be parametrised in a Bayesian setting, an approach that allows for uncertainty in parameter estimates to be quantified (Section 2.5).

This dose-and-time response approach can be used to identify the average length of the therapeutic window given a known initial dose of pathogen. Naturally, however, the progression of disease differs between individuals and therefore so does the length of this therapeutic window. An alternative approach could therefore be to find methods of extending this window of opportunity. For this, a detailed understanding of the early progression of the disease prior to symptom onset is required. In Chapter 4, this level of detail is accounted for in an agent

based model (ABM) that describes a murine response to the inhalation of *F. tularensis* bacteria. It is demonstrated how, in the form of birth-and-death processes with catastrophe, some of the most simple stochastic processes can be used to approximate the more complex dynamics of the ABM. These approximations allow parameter inference to be performed efficiently, which as a result yields an accurate estimate for the intracellular replication rate of *F. tularensis* bacteria, a parameter that is identified as being crucial for dictating how fast the infection progresses.

Unlike *F. tularensis*, there is currently no approved vaccine or antiviral treatment for Ebola virus, which is reflected in the extremely high case fatality rates. Furthermore, due to this high mortality rate and the potential for person-to-person transmission, experiments involving live EBOV must be conducted in Containment Level 4 (CL-4) laboratories, often restricting the types of experiments that can be performed (Knipe *et al.* (2007)). Minigenome replicon systems offer an approach to study the transcription and replication of the EBOV genome by using a shortened genome that does not contain the viral genes, and can therefore be carried out in CL-2 laboratories (Tao *et al.* (2017)). However, the safest approach yet to study virus kinetics is using mathematical modelling. In order to develop therapeutics, it is necessary to first understand how the infection progresses. For this reason, a model of EBOV is developed in Chapter 5 that builds on traditional models of viral kinetics by also accounting for the stochastic intracellular dynamics of infected cells. In doing so the model can be compared to multiple *in vitro* experiments, and from this the types of data that provide the most learning can be identified. Chapter 5 concludes with the derivation of the basic reproduction number, $\mathcal{R}_0$, an indicator of whether or not the infection will spread. Here, a stochastic approach is used to represent the intracellular dynamics of a single infected cell, with this being the first known effort to compute $\mathcal{R}_0$ for EBOV in this manner.

Chapter 2

Mathematical background

In many biological processes, the presence of small numbers of molecules means that, from a modelling perspective, it is no longer appropriate to describe these processes using a deterministic approach. One of the most striking features of *F. tularensis* infection is that an initial dose of as few as ten bacteria is sufficient to cause infection. In such a case, deterministic models would only be able to describe the polarised scenarios where infection either takes hold or the bacterial population becomes extinct. For this reason, stochastic processes are primarily the tool used throughout this thesis to model the response to infection. By capturing the natural randomness associated with small populations, it is possible to gain insight into features, such as the probability of population extinction, that cannot be accurately studied using deterministic models alone. In saying this, it is important to understand that stochastic models can introduce unnecessary mathematical detail and complexity, and a balance between both modelling techniques is required. Although the majority of models described here are stochastic, deterministic approaches have also been applied, reflecting the necessity for this balance.

This chapter begins with an introduction to probability and an overview of stochastic processes required for the development of the mathematical models detailed in the remaining chapters. Analytical approaches used to study such processes are outlined, and when these methods are difficult to apply, two different stochastic simulation algorithms are also described. Standard results are

provided for a well-known type of stochastic process, known as a birth-and-death process, with these results extended in Chapter 4 to account for catastrophes within a population. Bayesian approaches to parameter inference are also introduced, thereby facilitating a method for comparing experimental data with the mathematical models developed.

2.1   Introduction to probability

In this section I introduce ideas from probability theory necessary for defining and interpreting stochastic processes. The definitions provided here are equivalent to those in Allen (2003) and Taylor & Karlin (1998).

A random variable, X, can be described as a variable that takes on its value by chance, where this value is determined by the outcome of a random phenomenon. Let $\mathcal{S}$ be the set of values X can assume, then the probability that X takes a value less than or equal to a real number, $x \in \mathcal{S}$, defines the *distribution function* $F_X(x)$, that is,

$$F_X(x) = \Pr(X \leq x), \qquad x \in \mathcal{S}.$$

If the number of possible values that X can assume is finite, the random variable is discrete and the *probability mass function* (p.m.f), $f(x)$, defines the probability that X assumes a specific value

$$f(x) = \Pr(X = x), \qquad x \in \mathcal{S}.$$

However, if $\Pr(X = x) = 0$ for every $x \in \mathcal{S}$, X is a continuous random variable. In this case, if there exists a non-negative function $f_X(x)$ such that

$$\Pr(a < X \leq b) = \int_a^b f_X(x)\, dx, \qquad a, b \in \mathcal{S}, \ \ a < b,$$

then $f_X(x)$ is called the *probability density function* (p.d.f) for the random variable X. A well-known example of a continuous random variable that has particular importance for stochastic processes is described by the exponential distribution.

2. MATHEMATICAL INTRODUCTION

Definition 1. *A non-negative random variable X is said to be exponentially distributed with parameter $\lambda > 0$ if the probability density function is*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0\,, \\ 0 & x < 0\,. \end{cases}$$

Suppose now that two independent random variables, X_1 and X_2, have probability density functions $f_{X_1}(x)$ and $f_{X_2}(x)$ respectively. The probability density function of their sum, $Z = X_1 + X_2$ is given by the convolution of the individual probability density functions

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X_1}(z - x)\, f_{X_2}(x)\, dx = \int_{-\infty}^{+\infty} f_{X_1}(x)\, f_{X_2}(z - x)\, dx\,.$$

By considering the sum of n exponential random variables, each with equal rate, the Erlang distribution can be defined.

Definition 2. *Let $X_1, X_2, ..., X_n$ be n independent exponentially distributed random variables, each with rate $\lambda > 0$. The sum $Z = X_1 + X_2 + ... + X_n$ is an Erlang(n, λ) distributed random variable with probability density function*

$$f_Z(z) = \begin{cases} \dfrac{\lambda^z z^{n-1} e^{-\lambda z}}{(n-1)!} & z \geq 0\,, \\ 0 & z < 0\,. \end{cases}$$

An important concept that can help to characterise the probability density function is the expected value of X, or the average value that X assumes. In general, for a function, g, of a discrete random variable, the expectation of $g(X)$ is given by

$$\mathbb{E}\left[g(X)\right] = \sum_{x \in \mathcal{S}} g(x) f(x)\,,$$

whilst for a continuous random variable, this same quantity is given by

$$\mathbb{E}\left[g(X)\right] = \int_{x \in \mathcal{S}} g(x) f_X(x)\, dx\,.$$

With this, probability generating functions can also be defined. These are an important concept when studying stochastic processes as they provide a method

for characterising the process that avoids solving an often intractable system of differential equations. For the purpose of this thesis, it is only necessary to define the probability generating function of a discrete random variable.

Definition 3. *The probability generating function (p.g.f), $P_X(z)$, of a discrete random variable X is*

$$P_X(z) = \mathbb{E}\left[z^X\right] = \sum_{j=0}^{+\infty} \Pr(X = j)\, z^j\,,$$

for some $z \in \mathbb{R}$.

The probability generating function can be used to compute the probability that X assumes a specific value,

$$\Pr(X = x) = \frac{1}{x!} \frac{d^{(x)}}{dz^{(x)}} P_X(z) \bigg|_{z=0}\,,$$

as well as defining the kth factorial moment of X

$$\mathbb{E}\left[\frac{X!}{(X-k)!}\right] = \frac{d^{(k)}}{dz^{(k)}} P_X(z) \bigg|_{z=1}\,, \qquad k \geq 0\,.$$

2.2 Stochastic processes

In outlining key properties of stochastic processes, this section makes use of definitions and results from Allen (2003); He (2014); Latouche & Ramaswami (1999); Renshaw (2011) and Taylor & Karlin (1998).

Stochastic processes are simply a collection of random variables but can be more formally defined as follows:

Definition 4. *A stochastic process is a collection of random variables $\mathcal{X} = \{X(t) : t \in T\}$ where T is some index set and $X(t)$ denotes a single random variable defined on a state space $\mathcal{S}_{\mathcal{X}}$. A stochastic process may also be a collection of n random vectors, $\mathcal{X} = \{X_1(t), X_2(t), ..., X_n(t) : t \in T\}$.*

The index set, T, is often used to denote time, where time is treated here as a continuous entity, such that $T = \{t : t \in [0, +\infty)\}$. Although the index set is

continuous, each of the stochastic processes described throughout this thesis have a discrete state space, $\mathcal{S}_{\mathcal{X}}$. Furthermore, many of these stochastic processes satisfy the *memoryless* Markov property, and are therefore referred to as continuous time Markov processes (CTMP). The Markov property intuitively states that the future state of a stochastic process depends only on its current state and not any previous states.

Definition 5. *The stochastic process* $\mathcal{X} = \{X(t) : t \in [0, +\infty)\}$ *is called a continuous time Markov process (CTMP) if the following condition holds true: For any sequence of real numbers satisfying* $0 \leq t_0 < t_1 < ... < t_n < t_{n+1}$,

$$\Pr(X(t_{n+1}) = i_{n+1} \,|\, X(t_0) = i_0, ..., X(t_n) = i_n) = \Pr(X(t_{n+1}) = i_{n+1} \,|\, X(t_n) = i_n).$$

2.2.1 Transition probabilities

Transition probabilities provide a way of relating the state of a stochastic process at different time points and are defined as

$$p_{i,j}(s, t) = \Pr(X(t) = j \,|\, X(s) = i) \qquad s < t.$$

If the transition probabilities depend only on the length of the interval, $t - s$, rather than s and t explicitly, that is,

$$p_{i,j}(s, t) = \Pr(X(t) = j \,|\, X(s) = i) = \Pr(X(t - s) = j \,|\, X(0) = i) = p_{i,j}(0, t - s),$$

then they are referred to as *homogeneous* and are instead denoted by $p_{i,j}(t - s)$. The most natural way to represent the transition probabilities is in a matrix, $\boldsymbol{P}(t)$, known as the *probability transition matrix*. The matrix $\boldsymbol{P}(t)$ is a square matrix of order $|\mathcal{S}_{\mathcal{X}}|$ whose $(i, j)^{\text{th}}$ entry is equal to the transition probability $p_{i,j}(t)$. Since a stochastic process only transitions between states within its state space, it follows that each row of $\boldsymbol{P}(t)$ sums to one.

2.2.2 Generator matrix

Closely related to the transition probabilities are the transition rates, which are often specified when initially defining a stochastic process. Let $q_{i,j}$ denote the

transition rate from state $i \in \mathcal{S}_{\mathcal{X}}$ to state $j \in \mathcal{S}_{\mathcal{X}}$. Provided that the transition probabilities $p_{i,j}(t)$ are continuous and differentiable for $t \geq 0$, and satisfy

$$p_{i,j}(0) = 0, \ \ i \neq j, \qquad p_{i,i}(0) = 1,$$

the transition rates are defined as

$$q_{i,j} = \begin{cases} \lim\limits_{\Delta t \to 0^+} \dfrac{p_{i,j}(\Delta t) - p_{i,j}(0)}{\Delta t} = \lim\limits_{\Delta t \to 0^+} \dfrac{p_{i,j}(\Delta t)}{\Delta t} & \text{for } i \neq j, \\ \lim\limits_{\Delta t \to 0^+} \dfrac{p_{i,i}(\Delta t) - p_{i,i}(0)}{\Delta t} = \lim\limits_{\Delta t \to 0^+} \dfrac{p_{i,i}(\Delta t) - 1}{\Delta t} & \text{for } i = j. \end{cases}$$

Since each row of the transition probability matrix sums to one, it can be shown that

$$q_{i,i} = -\sum_{j=0, j \neq i} q_{i,j}. \tag{2.1}$$

As with transition probabilities, transition rates can be represented by a matrix $\boldsymbol{Q} = (q_{i,j})$, referred to as the *generator matrix*. From (2.1) it follows that the diagonal entries of $\boldsymbol{Q}$ are equal to the negative of the sum of all remaining entries in the corresponding row, and thus the sum of each row of $\boldsymbol{Q}$ is zero.

2.2.3   Classification of states

For the stochastic processes studied in this thesis, states for which the process can enter but cannot leave are incorporated to represent, for example, the fate of an infected cell. These particular states are referred to as absorbing states and are more formally defined as follows:

Definition 6. *A state i in a CTMP is absorbing if the transition rate $q_{ii} = 0$.*

Consider a CTMP, $\mathcal{X} = \{X(t) : t \geq 0\}$, with a single absorbing state. If $X(0) = i$, then since there is a non-zero probability that $\mathcal{X}$ enters and remains in the absorbing state, there is a non-zero probability that $\mathcal{X}$ will never return to state i. In this case, state i is known as *transient* and it therefore follows that all non-absorbing states are transient.

2.2.4 Kolmogorov differential equations

The forward and backward Kolmogorov differential equations describe the rate of change of the transition probabilities. If $\boldsymbol{P}(t) = (p_{i,j}(t))$ is the probability transition matrix, then the forward Kolmogorov differential equation is given in matrix form by

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{Q}\boldsymbol{P}(t) \qquad \boldsymbol{P}(0) = \boldsymbol{I}\,,$$

where $\boldsymbol{I}$ is the identity matrix. The backward Kolmogorov differential equations can also be expressed in matrix form as

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{P}(t)\boldsymbol{Q} \qquad \boldsymbol{P}(0) = \boldsymbol{I}\,.$$

These matrix equations define a linear system of equations, and thus their solution may be written as

$$\boldsymbol{P}(t) = \boldsymbol{P}(0)\exp(\boldsymbol{Q}t)\,,$$

where the matrix exponential is defined as

$$\exp(\boldsymbol{Q}t) = \sum_{k=0}^{+\infty} \frac{(\boldsymbol{Q}t)^k}{k!}\,.$$

2.2.5 Phase-type distributions

Consider a Markov process $\mathcal{X}$ with state space $\mathcal{S}_{\mathcal{X}} = \{0\} \cup \{1, ..., N\}$ where 0 is a single absorbing state and the remaining N states are transient. The generator matrix for this process can be written as

$$\boldsymbol{Q} = \begin{pmatrix} 0 & \boldsymbol{0}^T \\ \boldsymbol{t} & \boldsymbol{T} \end{pmatrix}\,,$$

where the matrix $\boldsymbol{T}$ is a square matrix of order N containing the transition rates between transient states, and $\boldsymbol{t}$ is a column vector of size N containing the transition rates from each transient state to the absorbing state. Suppose that $\boldsymbol{\tau}$ is a vector containing the probability that $\mathcal{X}$ begins in each transient state, and let T^* be a random variable denoting the time until absorption of $\mathcal{X}$ into the absorbing state 0. The random variable T^* follows a phase-type distribution with representation $(\boldsymbol{\tau}, \boldsymbol{T})$.

Definition 7. *A non-negative random variable T^* has a phase-type (PH) distribution if its distribution function is given by*

$$F(t) = \Pr(T^* \leq t) = 1 - \boldsymbol{\tau}^T \exp(\boldsymbol{T}t)\mathbf{1} \qquad t \geq 0\,,$$

where $\mathbf{1}$ is an $(N \times 1)$ column vector of ones.

The similarity between the distribution function for a phase-type and exponential random variable is due to the phase-type distribution being a matrix generalisation of the exponential distribution. For example, consider a Markov process consisting of only two states with a single transition occurring from the first state to the second state. The time until absorption into the second state follows an exponential distribution, and thus the exponential distribution can be seen as the most simple of phase-type distributions. From here it is possible to see how a phase-type representation can be constructed for distributions related to the exponential distribution. However, as the following theorem indicates, it is possible to approximate any non-negative probability distribution using a phase-type distribution.

Theorem 1. *The set of PH-distributions is dense in the set of probability distributions on the non-negative half-line* (He (2014) - Theorem 1.2.1).

This result can be utilised when studying stochastic processes for which the Markov property is not satisfied and so the inter-event times are not exponentially distributed. In order to do this, a method for choosing the phase-type representation is required such that the phase-type distribution closely approximates the chosen distribution well.

2.2.6 Phase-type approximation

In Chapter 3, a PH distribution is used to approximate a log-normal distribution. The method for finding an appropriate PH representation utilises a moment matching algorithm described by Osogami & Harchol-Balter (2006). This algorithm finds a phase-type representation such that the first three moments of the phase-type distribution agree with those of the distribution being approximated.
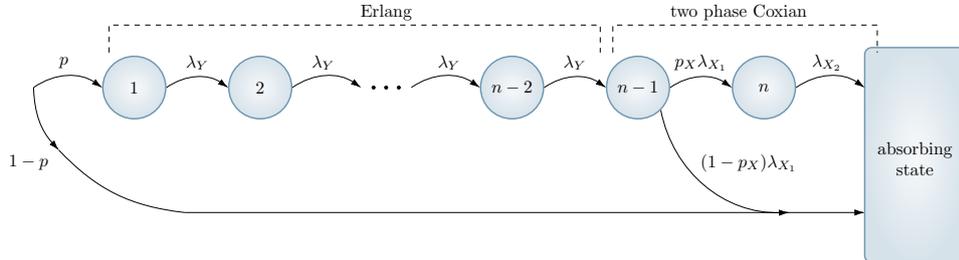
Figure 2.1: A depiction of the underlying Markov process of an EC distribution.

Since phase-type distributions are partly parametrised by a matrix $\boldsymbol{T}$ whose order is equal to the number of transient states, N, the number of parameters to approximate is of order N^2. For this reason the search space is narrowed by considering a subset of phase-type distributions known as Erlang-Coxian (EC) distributions. The structure of the underlying Markov chain of an EC distribution is depicted in Figure 2.1. By restricting the search space in this manner, Osogami & Harchol-Balter (2006) show that closed form expressions for each of the six parameters p, λ_Y, λ_{X_1}, λ_{X_2}, p_X and n can be found such that the first three moments of the chosen distribution agree with those of an EC distribution.

2.2.7 Birth-and-death processes

Birth-and-death processes are specific examples of CTMPs that have particular importance when modelling populations. Let $\mathcal{X} = \{X(t) : t \geq 0\}$ be a birth-and-death process with state space $\mathcal{S}_\mathcal{X} = \{0, 1, 2, ...\}$, the one-step transition probabilities satisfy

$$p_{i,i+j}(\Delta t) = \begin{cases} \lambda_i \Delta t + o(\Delta t) & j = 1, \\ \mu_i \Delta t + o(\Delta t) & j = -1, \\ 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t) & j = 0, \\ o(\Delta t) & \text{otherwise}. \end{cases}$$

Therefore, with each jump the state of the process can either increase by one, indicating a birth event, or decrease by one, indicating a death event. Figure 2.2
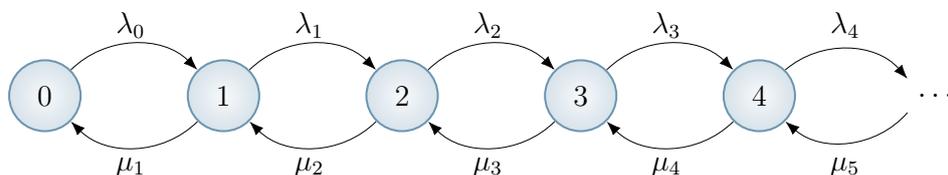
Figure 2.2: A depiction of a birth-and-death process with birth rate λ_n and death rate μ_n.

depicts a birth-and-death process as described here. The generator matrix for a birth-and-death process has a tri-diagonal structure and is given by

$$
\boldsymbol{Q} = \begin{pmatrix}
-\lambda_0 & \lambda_0 & & & \\
\mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\
& \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\
& & \ddots & \ddots & \ddots \\
& & & &
\end{pmatrix},
$$

where only non-zero entries have been reported. Although the birth rate, λ_i, and death rate, μ_i, can take any form, a commonly studied case is when both rates are linear. In this case $\lambda_i = \lambda i$ and $\mu_i = \mu i$, $i \in \mathcal{S}_X$, where λ and μ are the respective per individual birth and death rates. Note that when assuming linear rates, state 0 becomes an absorbing state that represents population extinction, from which there is no recovery. In the following sections, standard results are provided for this process as well as the special case when $\mu = 0$, where the process reduces to a simple birth process.

Linear birth process ($\mu = 0$)

For a simple birth process with linear rates and initial population size $X(0) = n_0$, the transition probabilities satisfy the Kolmogorov differential equations

$$
\frac{dp_{n_0}(t)}{dt} = -\lambda n_0 p_{n_0}(t) \,,
$$

$$
\frac{dp_n(t)}{dt} = \lambda(n-1)p_{n-1}(t) - \lambda n p_n(t) \qquad n > n_0 \,,
$$

where $p_n(t) = \Pr(X(t) = n \,|\, X(0) = n_0)$. Integrating each differential equation sequentially yields the solution

$$
p_n(t) = \binom{n-1}{n_0-1} e^{-\lambda n_0 t} \left(1 - e^{-\lambda t}\right)^{i-n_0} \qquad n \geq n_0 \,,
$$

which can be seen as the probability density function of a negative binomial distribution. Because of this, the probability generating function can immediately be shown to be

$$
G(z;t) = \left(\frac{ze^{-\lambda t}}{1 - z + ze^{-\lambda t}}\right)^{n_0} .
$$

Linear birth-and-death process $(\mu \neq 0)$

Although a birth-and-death process includes just a single additional event compared to a simple birth process, this results in a much more complex structure. For a birth-and-death process, if $p_n(t)$ is defined as above, the Kolmogorov differential equations are given by

$$
\frac{dp_0(t)}{dt} = \mu p_1(t)
$$
$$
\frac{dp_n(t)}{dt} = \lambda(n-1)p_{n-1}(t) + \mu(n+1)p_{n+1}(t) - (\lambda + \mu)n p_n(t) \qquad n > 0 \,.
$$

This system of equations is not as easily solved as the analogous equations for a simple birth process, therefore an approach is instead introduced that utilises the probability generating function. By multiplying the Kolmogorov differential equations by z^n and summing over all possible values of n,

$$
\sum_{n=0}^{+\infty} \frac{dp_n(t)}{dt} z^n = \lambda \sum_{n=0}^{+\infty} n p_{n-1}(t) z^n - \lambda \sum_{n=0}^{+\infty} p_{n-1}(t) z^n + \mu \sum_{n=0}^{+\infty} n p_{n+1}(t) z^n
$$
$$
+ \mu \sum_{n=0}^{+\infty} p_{n+1}(t) z^n - (\lambda + \mu) \sum_{n=0}^{+\infty} n p_n(t) z^n .
$$

The left hand side of this equation can be recognised as the derivative of the probability generating function with respect to time, whilst terms in the right hand side have the form of the derivative with respect to z. Therefore, if $G(z;t)$

is the probability generating function for a birth-and-death process, this satisfies the partial differential equation

$$\frac{\partial G(z;t)}{\partial t} = (\lambda z - \mu)(z - 1)\frac{\partial G(z;t)}{\partial z}, \qquad G(z;0) = z^{n_0}.$$

As shown by Renshaw (2011), the solution to this partial differential equation is

$$G(z;t) = \left(\frac{\mu(1 - z) - (\mu - \lambda z)e^{-(\lambda - \mu)t}}{\lambda(1 - z) - (\mu - \lambda z)e^{-(\lambda - \mu)t}}\right)^{n_0},$$

from which the transition probabilities can be obtained as

$$p_n(t) = \alpha^{n_0}\beta^n \sum_{i=0}^{n}\binom{n_0}{i}\binom{n_0 + n - i - 1}{n - i}\left[\frac{1 - \alpha - \beta}{\alpha\beta}\right]^i,$$

where

$$\alpha(t) = \mu\beta(t)/\lambda \qquad \text{and} \qquad \beta(t) = \frac{\lambda - \lambda e^{-(\lambda - \mu)t}}{\lambda - \mu e^{-(\lambda - \mu)t}}.$$

Therefore, despite introducing just a single additional event to a simple birth process, the transition probabilities have increased in complexity from following a negative binomial distribution to following the convolution of a binomial and negative binomial distribution. It is important to note that $G(0;t) = p_0(t)$ is the probability that the population has become extinct before time t, and is given here by $\alpha(t)$.

In Chapter 4, an extension of a birth-and-death process, known as a birth-and-death process with catastrophe will be considered, where the catastrophe event represents the instantaneous elimination of the entire population. Although introducing a catastrophe event further increases the complexity, it is shown that the transition probabilities still have the same form as described here for a birth-and-death process.

2.3 Quasi-birth-and-death process

Just as the PH distribution is a matrix generalisation of the exponential distribution, quasi-birth-and-death (QBD) processes are a matrix generalisation of the birth-and-death process.

Definition 8. *Consider a bivariate CTMP $\mathfrak{X} = \{(X(t), Y(t) : t \geq 0\}$ with state space $\mathcal{S}_{\mathfrak{X}} = \{(n, m) : n \geq 0, 1 \leq m \leq M\}$. The first coordinate, n, is called the level and the second coordinate, m, is called the phase. The term level can also be used to denote the subset of states*

$$L(n) = \{(n, 1), (n, 2), ..., (n, M)\}, \qquad \mathcal{S}_{\mathfrak{X}} = \bigcup_{n \geq 0} L(n).$$

The Markov process $\mathfrak{X}$ is a quasi-birth-and-death (QBD) process if one-step transitions can only occur between states (n, m) and (n', m') with $n' = n, n + 1$ or $n - 1$.

A QBD process is therefore similar to a birth-and-death process in that the latter may only transition between adjacent states, whereas a QBD process may only transition between states in the same or adjacent levels. Because of this, the generator matrix for a QBD process has the same tri-diagonal structure as that of a birth-and-death process, although the generator matrix is now block tri-diagonal,

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{0,0} & \boldsymbol{Q}_{0,1} & & & \\ \boldsymbol{Q}_{1,0} & \boldsymbol{Q}_{1,1} & \boldsymbol{Q}_{1,2} & & \\ & \boldsymbol{Q}_{2,1} & \boldsymbol{Q}_{2,2} & \boldsymbol{Q}_{2,3} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

The matrices $\boldsymbol{Q}_{i,i}$ contain the transition rates between states in level i, whilst $\boldsymbol{Q}_{i,i+1}$ and $\boldsymbol{Q}_{i,i-1}$ contain the transition rates from states in level i to those in level $i + 1$ and $i - 1$ respectively.

2.4 Stochastic simulation algorithms

Consider a multivariate Markov process $\mathfrak{X} = \{\boldsymbol{X}(t) : t \geq 0\}$. As seen previously for a birth-and-death process, the complexity of $\mathfrak{X}$ does not have to increase greatly before solving the Kolmogorov differential equations becomes too difficult. For this reason, it is often simpler to construct numerical realisations of the process $\mathfrak{X}$, noting that these numerical realisations are not equivalent to solving

the Kolmogorov differential equations numerically (Gillespie (2007)). These numerical realisations can prove to be a valuable tool for verifying analytic results for simpler processes as well. Two approaches for simulating Markov processes are described here, the first is an exact method, known as the Gillespie algorithm, that constructs realisations through determining the next reaction to occur at each time step. The second approach, known as tau-leaping, is an approximative algorithm that can benefit from reduced computational times.

2.4.1 Gillespie algorithm

Suppose that the Markov process $\mathcal{X}$, currently in state $\boldsymbol{x}$, is described by a series of reactions, R_j, $1 \leq j \leq M$, such that $a_j(\boldsymbol{x})\Delta t$ is the probability that reaction R_j occurs in a short time interval $(t, t + \Delta t)$. Let J be the set of indices of these reactions, and I be the set of species for which $\mathcal{X}$ describes the dynamics. The vector $\boldsymbol{\nu}_j$ is the state change vector and describes how each species changes following the occurrence of reaction R_j. The Gillespie algorithm for obtaining a single realisation of $\mathcal{X}$ is then described in Algorithm 2.1. Although exact, this method is not efficient when a single population is large, or the time-scale at which a population evolves is vastly different to the remaining populations. In such a case, $a_0(\boldsymbol{x}) = \sum_{j=1}^{M} a_j(\boldsymbol{x})$, the total rate at which any reaction occurs, will be large. Since the mean time between reactions is $1/a_0(\boldsymbol{x})$, this results in only small increments in time being made at each step of the algorithm. As a result of this, long computational times can be required to construct a single realisation, a problem that can be avoided through implementing a tau-leaping procedure.

2.4.2 Tau-leaping

The tau-leaping algorithm is based on an assumption, often referred to as the *leap condition*, that for a time interval of length τ, the value of $a_j(\boldsymbol{x})$ does not change significantly for each reaction, R_j. As a result of this, the number of times that reaction R_j occurs in this interval is a Poisson random variable with mean $a_j(\boldsymbol{x})\tau$. Therefore, instead of determining at each step the single reaction to next occur, as in the Gillespie algorithm, the number of times that each reaction occurs is considered simultaneously, with the state of $\mathcal{X}$ then updated accordingly.

2. MATHEMATICAL INTRODUCTION

> Select t_{max}, the maximum time point for which a realisation of $\mathfrak{X}$ will be constructed. Set $t = 0$ and $\boldsymbol{x}$, a vector describing the initial state of $\mathfrak{X}$.
>
> **while** $t < t_{max}$ **do**
>
> > 1. Compute $a_j(\boldsymbol{x})$ for $j = 1, ..., M$ and $a_0(\boldsymbol{x}) = \sum_{j=1}^{M} a_j(\boldsymbol{x})$.
> >
> > 2. Sample $u_1, u_2 \sim U(0, 1)$. The next reaction to occur is then the reaction, R_k, that satisfies $\sum_{j=1}^{k-1} \frac{a_j(\boldsymbol{x})}{a_0(\boldsymbol{x})} < u_1 \leq \sum_{j=1}^{k} \frac{a_j(\boldsymbol{x})}{a_0(\boldsymbol{x})}$.
> >
> > 3. Update the state of the process $\mathfrak{X}$ using the state change vector and increment time:
> > $$\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{\nu}_k,$$
> > $$t \to t - \log(u_2)/a_0(\boldsymbol{x}).$$
>
> **end**

Algorithm 2.1: A Gillespie algorithm to obtain a single realisation of a CTMP.

Since the Poisson distribution is unbounded, if the realisation of $\mathfrak{X}$ is near a boundary of its state space, it is possible that the tau-leaping algorithm will attempt to send the realisation to a state outside of the state space of $\mathfrak{X}$. To prevent this from happening, a method described by Cao *et al.* (2006) is used here that involves defining a subset of states for each reaction, referred to as the *critical region*. If the current state of the process, $\boldsymbol{x}$, is in the critical region for a specific reaction, this reaction will occur at most once during the next time step, thus ensuring that the realisation is contained within states in the state space of $\mathfrak{X}$. Furthermore, since selecting a step length, τ, that is too small eliminates the main advantage of tau-leaping, the method described here also optimises the step length whilst still satisfying the leap condition. Algorithm 2.2 details the tau-leaping algorithm that is later applied in sections 3.3.2 and 3.4.2. In this algorithm, $v_{i,j}$ is defined as the change in species i due to reaction j, such that $\boldsymbol{\nu}_j = (v_{1,j}, v_{2,j}, ..., v_{N,j})$.

Select t_{max}, the maximum time point for which a realisation of $\mathfrak{X}$ will be constructed. Set $t = 0$ and $\boldsymbol{x}$, a vector describing the initial state of $\mathfrak{X}$.

while $t < t_{max}$ **do**

1. Identify the critical reactions, that is, reactions that, if occur n_c times, result in the realisation entering a state not in the state space of $\mathfrak{X}$. Let J_c and J_{ncr} be the sets of indices of the critical and non-critical reactions respectively.

2. Compute:
$$\mu_i(\boldsymbol{x}) = \sum_{j \in J_{ncr}} v_{i,j} a_j(\boldsymbol{x}), \ \ \forall \, i \in I$$
$$\sigma_i^2(\boldsymbol{x}) = \sum_{j \in J_{ncr}} (v_{i,j})^2 a_j(\boldsymbol{x}), \ \ \forall \, i \in I$$
$$\tau' = \min_{i \in I} \left\{ \frac{\max(\varepsilon x_i, 1)}{|\mu_i(\boldsymbol{x})|}, \frac{\max(\varepsilon x_i, 1)^2}{\sigma_i^2(\boldsymbol{x})} \right\} .$$

3. If $\tau' < 10/a_0(\boldsymbol{x})$, there is little benefit in performing tau-leaping. Instead perform 100 steps of a Gillespie algorithm in order to reach a different region of the state space and return to 1.

4. Compute $a_0^c(\boldsymbol{x}) = \sum_{j \in J_c} a_j(\boldsymbol{x})$ and sample $\tau'' \sim Exp\left(a_0^c(\boldsymbol{x})\right)$.

5. Set $\tau = \min(\tau', \tau'')$:
   Each non-critical reaction occurs $k_j \sim Poisson(a_j(\boldsymbol{x})\tau)$ times.
   If $\tau = \tau'$: no critical reactions occur.
   If $\tau = \tau''$: a single critical reaction occurs once, with the reaction determined in the same manner as in a Gillespie algorithm.

6. Update the state of the realisation and increment time:
   $\boldsymbol{x} \to \boldsymbol{x} + \sum_{j \in J} k_j \boldsymbol{\nu}_j$
   $t \to t + \tau.$

end

Algorithm 2.2: A tau-leaping algorithm to obtain a single realisation of a Markov process.

2.5 Bayesian inference

In this section I describe an algorithmic approach to infer model parameters within a Bayesian framework. This algorithm, along with experimental results, is then used throughout this thesis. The algorithms and theory described here follow work by Toni *et al.* (2009).

In a Bayesian setting, the process of inference revolves around updating an individual's prior beliefs about a model parameter after observing data, where the updated beliefs take into account both the prior beliefs and the data. Let θ be a model parameter whose true value is unknown, the prior beliefs regarding this parameter are then encoded in the *prior distribution*, denoted $\pi(\theta)$. Prior distributions can be referred to as *informative*, whereby the beliefs of the individual are able to shape the distribution, with greater density assigned to those values that are more likely believed to be the true value. However, in some cases there is not sufficient evidence available to justify an informative prior. In such a case the prior distribution encodes this ignorance by considering that θ can take any value within a certain range with equal probability. Priors of this type are often referred to as *non-informative priors*.

Once the prior distributions are specified, the prior beliefs are updated in the light of the data using Bayes' rule

$$\pi(\theta \mid X) = \frac{\pi(\theta)\pi(X \mid \theta)}{\int_\theta \pi(X \mid \theta)\pi(\theta) \, d\theta} \, , \tag{2.2}$$

where $\pi(\theta \mid X)$ is the *posterior* distribution and $\pi(X \mid \theta)$ is the likelihood of observing the data, X, given θ. Evaluating the posterior beliefs is often made complicated because of the integral in the denominator, however, since this is simply a normalising constant that ensures the posterior probabilities sum to one, (2.2) can be written in the simpler form

$$\pi(\theta \mid X) \propto \pi(\theta)\pi(X \mid \theta) \, .$$

Further complications can also arise when inferring parameters for computational or simulation-based models, as in this case there is usually no mathematical form for the likelihood. The following *likelihood free* method provides an approach to obtain the posterior distribution without evaluating the likelihood function.

2.5.1 Approximate Bayesian computation

Approximate Bayesian computation (ABC) intuitively involves simulating data from a model for multiple parameter values and comparing how 'close' the simulated and observed data are in order to infer the true value of each parameter. Consider a model $\mathcal{M}$ with a parameter θ and observed data $\mathcal{D}$. By sampling θ^* from its respective prior distribution, $\theta^* \sim \pi(\theta)$, simulated data $\mathcal{D}^* \sim \pi(\mathcal{D}|\theta^*)$ can be obtained in order to provide a sample $(\theta^*, \mathcal{D}^*)$ from the joint distribution. The posterior distribution can then be constructed by accepting parameters such that the simulated data matches the observed data. Provided that enough iterations are performed, this algorithm will always provide the exact posterior distribution. However, in practice the simulated and observed data will never match exactly. For this reason, the ABC algorithm requires only that the simulated and observed data are close, with how close they are determined by a chosen measure of distance. The distance function is specific to the inference problem but should be chosen by taking into account the importance of each observation or set of results. An algorithm for obtaining an approximate sample of the posterior distribution is detailed in Algorithm 2.3.

2.6 Global sensitivity analysis

The models that are developed here for host-pathogen interactions will involve multiple parameters representing the key underlying biological mechanisms. These parameters will ultimately be inferred using relevant data sets, however, at this point it is also important to identify the parameters that have the greatest effect on the model output. It is then preferable that during the inference steps, the parameters where the most learning occurs are the parameters that are also the most important. It is not important to confine the estimate of a particular parameter if varying that parameter results in little change in the model output.

Local sensitivity analysis is one method to identify influential parameters by considering the derivative of the model output with respect to individual parameters. The main disadvantage of this approach is that it does not allow for the evaluation of simultaneous changes in all model parameters (Zhang *et al.* (2015)).

Choose the desired posterior sample size, N, the acceptance threshold ε, and set $i = 0$. Let the observed data be stored in the vector $\boldsymbol{x}$.

while $i < N$ **do**

    1. Sample $\boldsymbol{\theta}^*$, from the prior distribution $\pi(\boldsymbol{\theta})$

    2. Simulate a data set $\boldsymbol{x}^*$ from $\pi(\boldsymbol{x} \,|\, \boldsymbol{\theta}^*)$

    3. If $d(\boldsymbol{x}, \boldsymbol{x}^*) \leq \varepsilon$, accept $\boldsymbol{\theta}^*$ into the posterior sample and set $i \to i + 1$.

end

Algorithm 2.3: Approximate Bayesian Computation (ABC) rejection sampling algorithm to obtain an approximate sample of the posterior distribution (Toni *et al.* (2009)).

For this reason a global sensitivity analysis algorithm is applied here, more specifically the Sobol approach to global sensitivity analysis. This method is described in detail in Saltelli *et al.* (2004) and Sobol (1993) with only the most important details provided here.

Consider a model that is parametrised by a vector of parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$, such that the model output is described by $Y = f(\boldsymbol{\theta})$. Each parameter may be thought of as a random variable that can vary over a known range. Since Y is a function of these parameters, it is also a random variable with variance $V(Y)$. The idea behind the global sensitivity analysis algorithm is to ask how the variance of this model output would change if the value of a parameter is known, that is, to consider the conditional variance $V(Y \,|\, \theta_i = \theta_i^*)$. Since the true value, θ_i^*, is not known, the averaged conditional variance is instead considered, $\mathbb{E}\left[V(Y \,|\, \theta_i = \theta_i^*)\right]$, where the expectation is with respect to θ_i, and the variance is taken over all remaining parameters, θ_j, $j \neq i$.

The law of total variance allows the variance of the model output to then be written as

$$V(Y) = \mathbb{E}\left[V(Y \,|\, \theta_i)\right] + V\left(\mathbb{E}\left[Y \,|\, \theta_i\right]\right),$$

from which the first order *sensitivity index* for parameter θ_i may be defined as

$$S_i = \frac{V(\mathbb{E}\left[Y \mid \theta_i\right])}{V(Y)}.$$

In addition to studying the reduction in model output variance obtained when fixing a single parameter, it is also useful to consider the case where the values of multiple parameters are fixed. In particular, let $V(\mathbb{E}\left[Y \mid \boldsymbol{\theta}_{-\theta_j}\right])$ be the expected reduction in variance that results from fixing all parameters excluding θ_j. It then follows that

$$S_{Ti} = \frac{\mathbb{E}\left[V(Y \mid \boldsymbol{\theta}_{-\theta_i})\right]}{V(Y)} = 1 - \frac{V(\mathbb{E}\left[Y \mid \boldsymbol{\theta}_{-\theta_i}\right])}{V(Y)},$$

is the sum of each sensitivity index that involves fixing parameter θ_i. For this reason, S_{Ti} is referred to as the *total effect* term for parameter θ_i.

To obtain these conditional variances, Sobol (1993) showed that a function f, integrable over $[0,1]^n$, may be decomposed as

$$f(\boldsymbol{\theta}) = f_0 + \sum_i f_i(\theta_i) + \sum_{i<j} f_{ij}(\theta_i, \theta_j) + ... + f_{12...n}(\theta_1, \theta_2, ..., \theta_n). \qquad (2.3)$$

Provided that each term in (2.3) has zero mean, squaring both sides and integrating yields

$$\int f^2(\boldsymbol{\theta})\, d\boldsymbol{\theta} - f_0^2 = V(Y) = \sum_i V_i + \sum_{i<j} V_{i,j} + ... + V_{12...n},$$

where $f_0 = \mathbb{E}\left[Y\right]$ and

$V_i = V(\mathbb{E}\left[Y \mid \theta_i\right])$,

$V_{ij} = V(\mathbb{E}\left[Y \mid \theta_i, \theta_j\right]) - V(\mathbb{E}\left[Y \mid \theta_i\right]) - V(\mathbb{E}\left[Y \mid \theta_j\right])$,

$\vdots$

$V_{12...n} = V(\mathbb{E}\left[Y \mid \theta_1, \theta_2, ..., \theta_n\right]) - \left(\sum_i V(\mathbb{E}\left[Y \mid \theta_i\right]) + \sum_{i<j} V(\mathbb{E}\left[Y \mid \theta_i, \theta_j\right]) + ...\right).$

It is generally believed that if S_i and S_{Ti} can be computed for all parameters, then a fairly complete description of the model can be obtained in terms of its global sensitivity analysis properties, where greater values of the sensitivity index indicate that a parameter has greater importance. The Python package SALib is used in this thesis to compute the sensitivity indices when necessary.

2.7 Agent based modelling

An agent based model (ABM) is a stochastic model used to describe populations of interacting agents (Bauer (2009)). The main advantage of ABMs is that each agent, for example a cell or bacterium, has its own set of attributes and obeys its own set of rules. An ABM is therefore able to account for heterogeneity within the population, thus enabling an understanding of how the microscopic rules of the agents affect the macroscopic behaviours of the population (Gorochowski (2012)). This heterogeneity is difficult to incorporate into the CTMPs described previously since each unique cell type would need to be represented by a different random variable. For example, modelling the number of infected and uninfected cells already requires two random variables, one for each population. Suppose now that each cell is also distinguished by its expression of a specific surface marker, this then requires four random variables to model each unique cell type. From here it is simple to see that the CTMP approach is not well-suited to describing populations of interacting agents when each agent has multiple attributes. A second advantage of ABMs is in the way they are defined. Through specifying rules for each agent, this acts as a natural way of constructing a model to mimic a biological system that is both appealing to biologists and mathematicians (Bauer (2009)).

One downside to agent based modelling can be that the complex network of interacting agents results in long simulation times. For this reason it is not always possible to parametrise ABMs from experimental observations, and so ABMs sometimes offer more of an exploratory role in determining the behaviour of populations. In Chapter 4, I describe how an ABM may be constructed to model bacterial infections, in addition to developing approximations that reduce the computation time.

Chapter 3

A multi-scale model for *Francisella tularensis* infection

In this chapter I focus on the dose dependent effects of *Francisella tularensis* infection on a population of susceptible individuals, with the aim of developing a dose and time response model that can quantify the risk of the airborne release of bacteria to such a population. When considering the threat posed by chemical or biological warfare, compared to chemical agents, quantifying the risk to a population from biological agents has the added complexity of time. The initial dose of a chemical agent is ultimately the amount available to cause a response, whereas biological agents have the capability to proliferate once inside the host. Moreover, the process of microbial growth can take days, whereas the toxic effects of chemical warfare agents can be seen within hours or minutes. It is therefore necessary to understand the speed and extent of this replication, and thus the length of time following exposure to the agent, as well as the initial dose received (Huang & Haas (2009)). Since the timescale of infection is often of the order of days, the window of opportunity for effective medical treatment is narrow. A greater understanding of the relationship between this timescale, the probability of response, and the initial dose would therefore provide valuable information to guide optimal treatment, in order to prioritise those most at risk.

The need to understand the effects that the initial dose has on the time and

probability of response has resulted in attempts to incorporate time into existing dose response models, such as the classical exponential and beta-Poisson models. This has previously been performed by allowing for a time dependency of the model parameters (Chen (2007); Huang & Haas (2009)). However, in choosing convenient statistical distributions, such as the exponential distribution, the link between the dose response model and the underlying biological mechanisms that dictate the level of bacterial replication is tenuous. Previous modelling efforts have not made use of parameters with a clear and intuitive interpretation. For example, in the beta-Poisson model described by Haas (1999), α is a shape parameter of the beta distribution used to incorporate variability in the probability that an organism survives to initiate infection. Therefore, the exact biological meaning of this parameter with regards to the infection dynamics is not trivial. An alternative approach is proposed in Pujol *et al.* (2009) whereby a stochastic mechanistic model is used to explicitly account for the interactions between immune effector cells and pathogens, in addition to accounting for the period of time an individual is exposed to the agent. Most recently, a Markov chain model developed by Wood *et al.* (2014) addresses the concerns associated with classical dose response models by considering the key interactions between *F. tularensis* bacteria and host phagocytes within the lung space. By defining a response to correspond with the onset of symptoms, and relating this to the number of extracellular bacteria within the lung space, the probability of response and time to response are studied in a dose dependent manner. In this chapter I extend the original model by Wood *et al.* (2014), that only accounts for within-host dynamics, through the development of a multi-scale model that also addresses two key disadvantages of the existing model.

The first disadvantage is the consideration of a deterministic time for the time until rupture of an infected phagocyte. In assuming this, the variability in the time until rupture between different phagocytes, observed to follow a log-normal distribution, is not explicitly incorporated into the model. Secondly, the number of bacteria released when an infected phagocyte ruptures is found by evaluating a deterministic logistic growth process, representing the intracellular replication

of bacteria, at the deterministic time until rupture. In doing this it is then only possible to obtain a constant value for the number of bacteria released, thereby assuming that all phagocytes release the same number of bacteria when rupturing and not accounting for inter-phagocyte variability. In order to address these limitations, a stochastic within-phagocyte model is developed to incorporate the intracellular dynamics of *F. tularensis* that also accounts for the observed log-normally distributed time until rupture. Using this, the distribution of the number of bacteria released by a rupturing phagocyte can be obtained and used to help parametrise and revise the within-host model, providing a link between within-phagocyte and within-host dynamics.

This chapter is organised in four sections. In Section 3.1, a description is provided of each of the models that form the three levels of the multi-scale model, namely the within-phagocyte, within-host, and population-level models. To analyse the within-phagocyte and within-host models, I make use of matrix analytic methods to determine a number of quantities of interest. These methods, previously employed by Gómez-Corral & López-García (2012) in the context of modelling competition between two species, provide an approach to analytically study the models that does not rely on solving the Kolmogorov equations. Moreover, by exploiting the structure of the underlying Markov process, the methods described here are far more efficient than those previously provided by Wood *et al.* (2014), reducing the overall computational time. In Section 3.2, these matrix analytic methods are further used to perform a local sensitivity analysis on the within-host model. A Bayesian analysis is used in Section 3.3 to parametrise the within-phagocyte and within-host models, with comparisons made between previous parameter estimates obtained by Wood *et al.* (2014). Results for each model are presented in Section 3.4, with the advantages of incorporating inter-phagocyte variability visible through the computation of the rupture distribution. It is also shown how predictions made by the population-level model can be used to influence the design of microbiology facilities, mitigating the risk to individuals inside.

3.1 Mathematical models

In this section I provide a description of the within-phagocyte, within-host and population-level models that contribute to the individual levels of the multi-scale model. Methods to obtain stochastic descriptors for the within-phagocyte and within-host model are presented, along with how these descriptors are used to provide a link between levels.

3.1.1 Within-phagocyte model

The first level of the multi-scale model is a within-phagocyte model, which is used to replicate the dynamics of intracellular *F. tularensis* bacteria following phagocytosis by a host phagocyte. This model should represent the key stages of the intracellular life cycle, including the escape of bacteria from the phagosomes where they initially reside, the replication of bacteria within the cytosol of the phagocyte, and the subsequent rupturing and death of the phagocyte (Cowley & Elkins (2011)). The replication and rupture stages of the intracellular life cycle can be modelled using a continuous-time Markov process $\mathcal{X} = \{X(t) : t \geq 0\}$ that has the structure of a birth-and-death process with total catastrophe (Di Crescenzo *et al.* (2008); Karlin & Tavaré (1982)). At any time $t \geq 0$, $X(t)$ represents the number of bacteria contained within the cytosol of the infected phagocyte. Prior to a rupture event occurring, $\mathcal{X}$ behaves like a birth-and-death process (Allen (2003)) over states $\mathbb{N} = \{1, 2, ...\}$, however, the moment that a rupture event occurs, the process instantaneously transitions to state B. This absorbing state represents the rupturing of the infected phagocyte, releasing its bacterial contents back into the extracellular environment, and is therefore referred to as the rupture state. The state space of $\mathcal{X}$ can then be written as $\mathcal{S}_{\mathcal{X}} = \mathbb{N} \cup \{B\}$.

The birth and death rates for any state $n \in \mathbb{N}$ are not only chosen to reflect the competition between bacteria for intracellular resources, such as iron, but also to keep to the same assumptions made by Wood *et al.* (2014) in their model of intracellular replication. Previously, a deterministic logistic growth process was used to represent the competition between bacteria, therefore the birth and

death rates here should be chosen such that, prior to a rupture event, $\mathcal{X}$ is the stochastic equivalent. From Allen (2003), this means choosing birth rates, λ_n, and death rates, μ_n, such that

$$\lambda_n - \mu_n = \lambda n \left(1 - \frac{n}{C} \right) ,$$

where λ is the per bacterium birth rate and C is the carrying capacity on the size of the population of intracellular bacteria within a single infected phagocyte. If replication of bacteria is proportional to the number of intracellular bacteria then $\lambda_n = \lambda n$, and μ_n can be found accordingly. This results in the following respective birth and death rates for the stochastic logistic growth process

$$\lambda_n = \begin{cases} \frac{\lambda(C-1)}{C} & \text{if } n = 1, \\ \lambda n & \text{otherwise,} \end{cases} \qquad \mu_n = \begin{cases} 0 & \text{if } n = 1, \\ \frac{\lambda n^2}{C} & \text{otherwise.} \end{cases}$$

The rate μ_1 is intentionally set to zero since a positive rate would indicate that phagocytes are capable of clearing their bacterial load. Although this is possible, only phagocytes succumbing to bacterial infection are later considered in the within-host model, as described in Wood *et al.* (2014).

The initial state of $\mathcal{X}$ corresponds to the number of bacteria initially taken up by the phagocyte. Results by Golovliov *et al.* (2003) suggest that phagocytosis of *F. tularensis* bacteria is relatively ineffective in monocytic cells so that, during the early stages of infection, only one or two intracellular bacteria are observed on average. From this, it is assumed here that a single phagocyte will only phagocytose a single bacterium, and thus the process $\mathcal{X}$ will always begin in state $X(0) = 1$. Furthermore, it is assumed that once infected, an increase in the bacterial load of a phagocyte is due solely to intracellular replication, and not further phagocytosis. This assumption is also consistent with Wood *et al.* (2014).

The key reason for developing this within-phagocyte model is to determine the distribution of the number of bacteria released when an infected phagocyte ruptures, herein referred to as the rupture distribution. Provided that state B represents the rupturing of the infected phagocyte, the number of bacteria released through this rupture event is equal to the state that $\mathcal{X}$ enters state B

from. The mechanism of phagocyte rupture is believed to be caspase-3 dependent (Wickstrum *et al.* (2009)), with Santic *et al.* (2010) showing that bacterial escape into the cytosol is both essential and sufficient for triggering caspase-3 activation. This would suggest that extensive bacterial replication is not required for infected phagocyte death to occur, and therefore the rate of rupture in the within-phagocyte model should be independent of bacteria load. Further reason for this assumption is given in a recent study by Brock & Parmely (2017). In this study, macrophages were infected with *F. tularensis* bacteria and after 24 hours were stained to detect cells with compromised membranes, an indicator of cell death. Bacteria were also stained a separate colour such that confocal microscopy could be used to enumerate cells based on their bacterial load. It was observed that over 25% of cells containing 1-5 bacteria had compromised membranes, whilst the same was true for only 50% of cells containing greater than 100 bacteria. This suggests that high levels of intracellular *F. tularensis* alone are neither sufficient nor necessary for the induction of cell death. Due to the assumption that the rate of rupture is independent of bacterial load, $\mathcal{X}$ may be thought of as a stochastic logistic growth process where $t = 0$ marks the start of a 'clock' that counts up towards the time until rupture of the phagocyte. At the moment the clock 'rings', signalling the rupture event, $\mathcal{X}$ immediately transitions into the rupture state B, from whichever state $\mathcal{X}$ is in. Let T^{rupture} be a random variable denoting the time taken for $\mathcal{X}$ to reach the rupture state.

Previously, Lindemann *et al.* (2011) have quantified the release of lactate dehydrogenase (LDH), a cytosolic enzyme that acts as an indicator of membrane disruption, in order to measure the rupturing of human macrophages following infection with *F. tularensis*. Using this data, Wood *et al.* (2014) have then applied maximum likelihood estimation methods to suggest that T^{rupture} is log-normally distributed, $T^{\mathrm{rupture}} \sim logN(3.72, 0.385)$. However, instead of incorporating the full distribution of the time until rupture into their within-phagocyte model, Wood *et al.* (2014) take the number of bacteria released in a rupture event to be equal to their deterministic logistic growth process evaluated at the median time until rupture, $Median[T^{\mathrm{rupture}}]$. This results in a constant number of bacteria
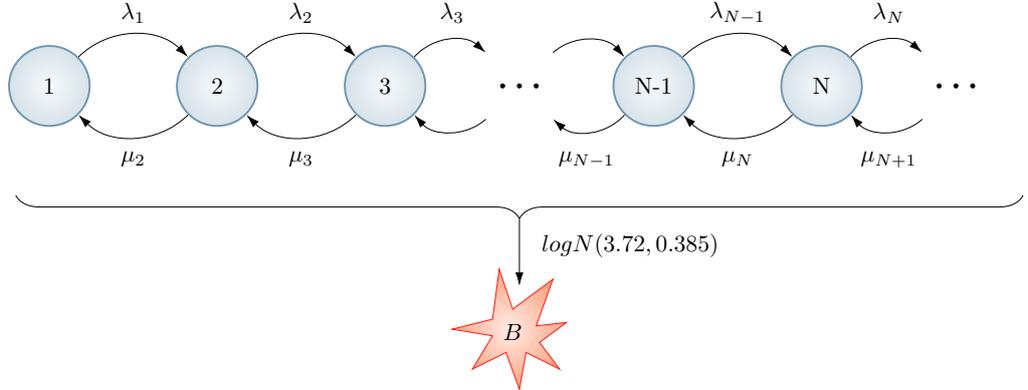
Figure 3.1: A depiction of the within-phagocyte model with log-normally distributed rupture times transitioning the process into the absorbing state B. Mean rupture time is $\mathbb{E}\left[T^{\mathrm{rupture}}\right] = 44.4\,h$.

being released from every rupture event.

In the within-phagocyte model developed here, the focus is to incorporate the full distribution of T^{rupture} in order to obtain the rupture distribution. In doing so it is not true that every inter-event time is exponentially distributed, the time until a rupture event is now log-normally distributed, and because of this the process $\mathcal{X}$ is no longer Markovian. A depiction of the within-phagocyte model with log-normally distributed rupture times is given in Figure 3.1. To resolve this issue, PH distributions are used, in particular, the property that the family of PH distributions is dense within the family of non-negative distributions (He (2014) - Theorem 1.2.1). The Markovian property can therefore be retained by approximating the log-normal distribution, $T^{\mathrm{rupture}} \sim logN(3.72, 0.385)$, using the PH distribution $T^{\mathrm{rupture}} \stackrel{.}{\sim} PH(\boldsymbol{\eta}, \boldsymbol{T})$.

In order to approximate the log-normal distribution, consider an auxiliary Markov process $\mathcal{W} = \{W(t) : t \geq 0\}$ with state space $\mathcal{S}_{\mathcal{W}} = \{I, II, ..., Z\} \cup \{B\}$, where state B is the same absorbing state as in $\mathcal{S}_{\mathcal{X}}$. Furthermore, let $q_{i,j}$ denote the transition rate between states $i, j \in \mathcal{S}_{\mathcal{W}}$. The set of transition rates $\{q_{i,j} : i, j \in \mathcal{S}_{\mathcal{W}}\}$ and the number of transient states, Z, may then be selected such that the time taken to reach state B is approximately $logN(3.72, 0.385)$ distributed. This means that $\mathcal{W}$ may be thought of as the 'clock' on the time until
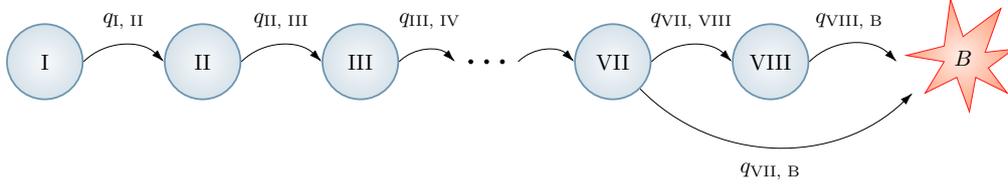
Figure 3.2: A depiction of the one-dimensional Markov process $\mathcal{W}$ associated with the PH($\boldsymbol{\eta}, \boldsymbol{T}$) distribution considered, so that the time to reach state B approximately follows a $logN(3, 72, 0.385)$ distribution. Only non-zero transition rates identified by $\boldsymbol{T}$ have been included.

rupture. As $\mathcal{W}$ transitions across states in $\mathcal{S}_\mathcal{W}$, the rupture event becomes ever nearer.

Using the PH-approximate procedure introduced in Section 2.2.6 and implemented in the statistical software R, an approximative PH distribution is obtained. By specifying the first three moments of $T^{\text{rupture}} \sim logN(3.72, 0.385)$ numerically, the procedure selects the number of transient states that comprise $\mathcal{S}_\mathcal{W}$, in addition to the initial probability vector and the matrix of transition rates. Here, a total of eight transient states are chosen, such that $\mathcal{S}_\mathcal{W} = \{I, II, ..., VIII\} \cup \{B\}$, whilst $\boldsymbol{\eta} = (1, 0, 0, 0, 0, 0, 0, 0)^T$ and

$$
\boldsymbol{T} = \begin{pmatrix}
-0.1447 & 0.1447 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -0.1447 & 0.1447 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -0.1447 & 0.1447 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -0.1447 & 0.1447 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -0.1447 & 0.1447 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -0.1447 & 0.1447 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -0.3396 & 0.0003 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.0127
\end{pmatrix}.
$$

The structure of the process $\mathcal{W}$, along with a plot depicting how accurately the log-normal distribution is being approximated are provided in Figure 3.2 and Figure 3.3 respectively. With this, the log-normal distribution can be replaced by the PH approximative distribution in the within-phagocyte process and $\mathcal{X}$
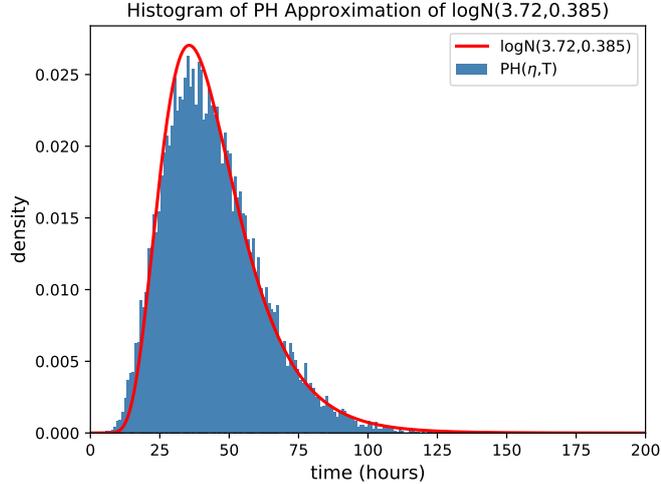
Figure 3.3: Plot showing how accurately the PH($\boldsymbol{\eta}, \boldsymbol{T}$) distribution approximates the desired $logN(3.72, 0.385)$ distribution. The histogram represents 5×10^3 samples from the PH($\boldsymbol{\eta}, \boldsymbol{T}$) distribution, the red curve is the probability density function for a $logN(3.72, 0.385)$ distribution.

can now be represented as in Figure 3.4. The result of this is that $\mathcal{X}$ is now a bivariate Markov process, $\mathcal{X} = \{\boldsymbol{X}(t) = (i,j) : i \in \mathbb{N}, j \in \mathcal{S}_\mathcal{W}\}$. Each state (i,j) of $\mathcal{X}$ represents that at a particular time instance, there are i intracellular bacteria, with the rupture clock in state j. Thus, arrows indicating transitions between rows of states in Figure 3.4 represent the rupture clock moving towards the rupture event, whilst arrows indicating transitions between columns of states represent the stochastic logistic growth process for the number of intracellular bacteria. Once the component j reaches the absorbing state B, the number of bacteria released is given by the corresponding value i at the instant this occurs.

Given the vector $\boldsymbol{\eta}$, the rupture clock will always start in state $I \in \mathcal{S}_\mathcal{W}$ at time $t = 0$. However, when *F. tularensis* bacteria first enter a host phagocyte, there is a period of approximately one hour where the bacteria are contained within a phagosome and are unable to replicate (Golovliov *et al.* (2003)). This phagosomal stage is not currently accounted for in the within-phagocyte model since bacterial replication is allowed to start from $t = 0$. Although bacterial
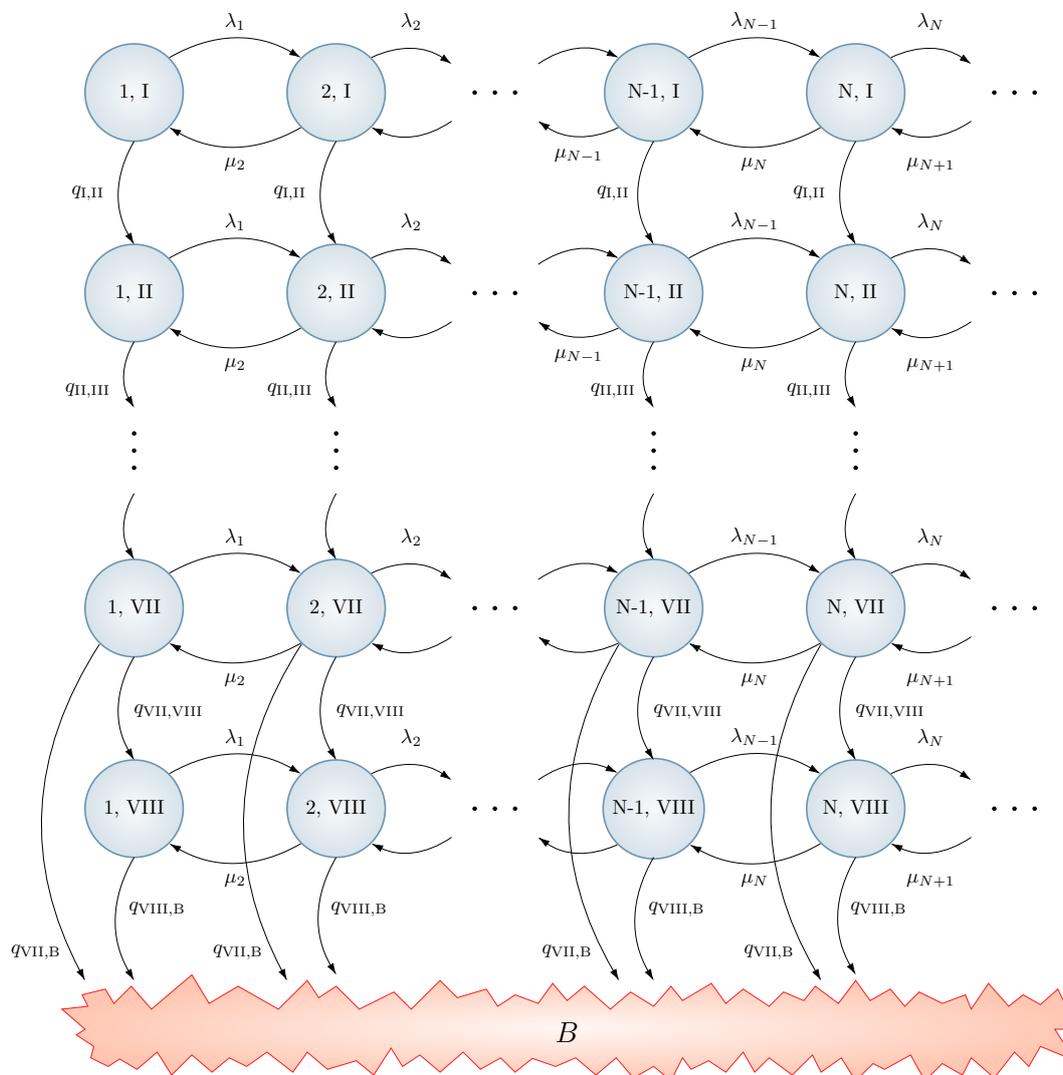
Figure 3.4: Diagram representing the within-phagocyte process $\mathcal{X}$ where the approximate PH distribution can be seen as a *clock* for the rupture event.

replication does not start immediately, the rupture clock does begin counting down as soon as a bacterium enters the phagocyte. Therefore, in order to account for the delay in replication, the assumption is made that during the first hour, the process $\mathcal{X}$ can only transition between states in the 'clock' phase. That is, if

the process depicted in Figure 3.4 starts in state $(i, j) = (1, I)$ at time $t = 0$, then during the first hour $\mathfrak{X}$ is only allowed to transition between states in the first column. By not allowing transitions along rows of states, this represents that replication of bacteria does not occur. To implement this, the probability that the auxiliary process $\mathcal{W}$ is in each of the transient states $I, II, ..., VIII$ at time $t = 1$ must be determined. These probabilities, found by evaluating the first row of the probability transition matrix of $\mathcal{W}$ at $t = 1$, are equal to

$$\boldsymbol{p} = \left(e^{\boldsymbol{T}}\right)_1 = \left(0.8653, 0.1252, 9.0586 \times 10^{-3}, 4.3693 \times 10^{-4}, \right.$$
$$\left. 1.5806 \times 10^{-5}, 4.5742 \times 10^{-7}, 1.0732 \times 10^{-8}, 4.6926 \times 10^{-13}\right).$$

This vector of probabilities now describes the initial probability vector for the component j of the process $\mathfrak{X}$ at time $t = 0$, replacing the existing vector $\boldsymbol{\eta}$. One concern with this approach is that it could be possible for the process $\mathcal{W}$ to enter the rupture state B within one hour, corresponding to a phagocyte rupturing prior to the infecting bacterium escaping the phagosome. However, given the initial probability vector, $\boldsymbol{p}$, the chance of such an event occurring in this case is negligible.

Truncation of the state space

Currently, the within-phagocyte model allows for an infinite number of bacteria to be present in the cytosol of an infected phagocyte, $\mathbb{N} = \{1, 2, ...\}$. If intracellular bacterial replication was instead represented by a deterministic logistic-growth process, the carrying capacity, C, would act as a threshold that the population of intracellular bacteria could not exceed. For a stochastic logistic growth process there is no threshold on the size of the population, instead the population will tend towards the carrying capacity before fluctuating around it. The population is able to exceed the carrying capacity since $\lambda_n = \lambda n > 0$ for any choice of n, in particular $n = C$. However, in order to use matrix analytic methods to compute stochastic descriptors for the within-phagocyte model, it is necessary for there to be a finite number of states, such that the infinitesimal generator matrix has finite dimension. This can be achieved by choosing a threshold N such that the

probability of the population of intracellular bacteria exceeding N is arbitrarily small. The state space of $\mathfrak{X}$ can then be truncated to include only those states that $\mathfrak{X}$ is likely to reach, that is, $\{(i,j) : i \in \{1, 2, ..., N\}, j \in \{I, II, ..., VIII\}\} \cup \{B\}$, with states $\{(i,j) : i \in \{N+1, N+2, ...\}, j \in \{I, II, ..., VIII\}\}$ being discarded (Gómez-Corral & López-García (2015)).

To find this threshold, N, let $\alpha_{(i,j)}^N$ denote the probability that given the initial state $(i,j) \in \{(m,n) : m \in \{1, 2, ..., N\}, n \in \{I, II, ..., VIII\}\}$, component i in process $\mathfrak{X}$ reaches the value $N+1$ prior to being absorbed into state B. This can be interpreted as the probability that $\mathfrak{X}$ enters a state outside of the truncation before reaching the rupture state. The aim is to then choose N such that $\alpha_{(1,j)}^N < \varepsilon$ for $j \in \{I, II, ..., VIII\}$ and a sufficiently small value of ε.

Suppose that the transient states of $\mathfrak{X}$ are organised into levels, such that level i is defined as

$$L(i) = \{(i,j) : j \in \{I, II, ..., VIII\}\}, \qquad 1 \leq i \leq N,$$

and the levels are ordered according to $L(1) \prec L(2) \prec \ldots \prec L(N)$. Using this particular grouping of states, levels can then be thought of as columns of states as depicted in Figure 3.4. If $\boldsymbol{\alpha}_i^N = \left(\alpha_{(i,I)}^N, ..., \alpha_{(i,VIII)}^N\right)^T$, that is, $\boldsymbol{\alpha}_i^N$ is a vector containing the probabilities that $\mathfrak{X}$ transitions to a state outside of the truncation provided it begins in each state of level i, then first step arguments can be applied to give the matrix equation

$$\begin{pmatrix} \boldsymbol{\alpha}_1^N \\ \boldsymbol{\alpha}_2^N \\ \vdots \\ \boldsymbol{\alpha}_{N-1}^N \\ \boldsymbol{\alpha}_N^N \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & & & \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{A}_{N-1,N-2} & \boldsymbol{A}_{N-1,N-1} & \boldsymbol{A}_{N-1,N} \\ & & & \boldsymbol{A}_{N,N-1} & \boldsymbol{A}_{N,N} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1^N \\ \boldsymbol{\alpha}_2^N \\ \vdots \\ \boldsymbol{\alpha}_{N-1}^N \\ \boldsymbol{\alpha}_N^N \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{A}_{N,N+1}\boldsymbol{1}_8 \end{pmatrix},$$

where $\boldsymbol{1}_8$ is a (8×1) column vector containing only ones. The matrices $\boldsymbol{A}_{i,i-1}$, $\boldsymbol{A}_{i,i}$, $\boldsymbol{A}_{i,i+1}$ each have dimension (8×8), since this is the number of states in each level, and contain the respective probabilities that the process $\mathfrak{X}$ transitions from states in level $L(i)$ to states in levels $L(i-1), L(i)$ and $L(i+1)$ in one

3. A MULTI-SCALE MODEL FOR *F. TULARENSIS* INFECTION

jump. Furthermore, since one-jump transitions can only occur between states in the same or adjacent levels, $\mathcal{X}$ has the structure of a QBD process. The non-zero entries of $\boldsymbol{A}_{i,i-1}$, $\boldsymbol{A}_{i,i}$, $\boldsymbol{A}_{i,i+1}$ are given respectively by

$$(\boldsymbol{A}_{i,i-1})_{j,j} = \mu_i \left(\lambda_i + \mu_i + \sum_{\substack{j'=I \\ j'\neq j}} q_{j,j'} + q_{j,B} \right)^{-1} \qquad i = 2, ..., N, \ j = I, II, ..., VIII \, ,$$

$$(\boldsymbol{A}_{i,i})_{j,l} = q_{j,l} \left(\lambda_i + \mu_i + \sum_{\substack{j'=I \\ j'\neq j}} q_{j,j'} + q_{j,B} \right)^{-1} \qquad i = 2, ..., N, \ j,l = I, II, ..., VIII \, ,$$

$$(\boldsymbol{A}_{i,i+1})_{j,j} = \lambda_i \left(\lambda_i + \mu_i + \sum_{\substack{j'=I \\ j'\neq j}} q_{j,j'} + q_{j,B} \right)^{-1} \qquad i = 2, ..., N, \ j = I, II, ..., VIII \, .$$

Given the block tridiagonal structure of the matrix in the previous matrix equation, the solution to this can be found in a recursive manner using a block-Gaussian elimination process. This results in Algorithm 3.1 which can be used to find the truncation value, N. Since $N \geq C$, an initial prediction of the truncation value is $N = C$, which is then incremented until $\alpha_{(1,j)}^N < \varepsilon$ for all $j \in \{I, II, ..., VIII\}$. A value of $\varepsilon = 10^{-5}$ has been used here to determine N.

Rupture size distribution

Following the truncation of the state space of $\mathcal{X}$, it is possible to find the rupture distribution, the key result from the within-phagocyte model that will provide a link between the intracellular and within-host dynamics. Given the previous grouping of states into levels, the probability that an infected phagocyte releases k bacteria when it ruptures is equivalent to the probability that $\mathcal{X}$ transitions to state B from a state in level $L(k)$. Therefore, let $R_{(i,j)}^{(k)}$ denote the probability that, given $X(0) = (i, j)$, $\mathcal{X}$ enters state B from level $L(k)$. First step arguments can be applied to obtain the following matrix equation for $\boldsymbol{R}_i^{(k)} = \left(R_{(i,I)}^{(k)}, ..., R_{(i,VIII)}^{(k)} \right)^T$

$$
\begin{aligned}
&N = C \\
&\textbf{while } \alpha_{1,j}^N \geq \epsilon \textit{ for any } j \in \{I, II, ..., VIII\} \textbf{ do} \\
&\quad\left|\begin{aligned}
&\boldsymbol{H}_1 = \boldsymbol{I}_8 - \boldsymbol{A}_{1,1} \\
&\textbf{for } n = 2, ..., N \textbf{ do} \\
&\quad\left|\ \boldsymbol{H}_n = \boldsymbol{I}_8 - \boldsymbol{A}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{A}_{n-1,n} - \boldsymbol{A}_{n,n}\right. \\
&\textbf{end} \\
&\boldsymbol{\alpha}_N^N = \boldsymbol{H}_N^{-1}\boldsymbol{A}_{N,N+1}\boldsymbol{1}_8 \textbf{ for } n = 2, ..., N \textbf{ do} \\
&\quad\left|\ \boldsymbol{\alpha}_n^N = \boldsymbol{H}_n^{-1}\boldsymbol{A}_{n,n+1}\boldsymbol{\alpha}_{n+1}^N\right. \\
&\textbf{end} \\
&N \to N + 1
\end{aligned}\right. \\
&\textbf{end}
\end{aligned}
$$

Algorithm 3.1: To obtain the truncation value, N, for the within-phagocyte model.

$$
\begin{pmatrix}
\boldsymbol{R}_1^{(k)} \\
\boldsymbol{R}_2^{(k)} \\
\vdots \\
\boldsymbol{R}_k^{(k)} \\
\vdots \\
\boldsymbol{R}_{N-1}^{(k)} \\
\boldsymbol{R}_N^{(k)}
\end{pmatrix}
=
\begin{pmatrix}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & & & & & \\
\boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} & & & & \\
& \ddots & \ddots & \ddots & & & \\
& & \boldsymbol{A}_{k,k-1} & \boldsymbol{A}_{k,k} & \boldsymbol{A}_{k,k+1} & & \\
& & & \ddots & \ddots & \ddots & \\
& & & & \boldsymbol{A}_{N-1,N-2} & \boldsymbol{A}_{N-1,N-1} & \boldsymbol{A}_{N-1,N} \\
& & & & & \boldsymbol{A}_{N,N-1} & \boldsymbol{A}_{N,N}
\end{pmatrix}
\begin{pmatrix}
\boldsymbol{R}_1^{(k)} \\
\boldsymbol{R}_2^{(k)} \\
\vdots \\
\boldsymbol{R}_k^{(k)} \\
\vdots \\
\boldsymbol{R}_{N-1}^{(k)} \\
\boldsymbol{R}_N^{(k)}
\end{pmatrix}
+
\begin{pmatrix}
\boldsymbol{0} \\
\boldsymbol{0} \\
\vdots \\
\boldsymbol{b}_k \\
\vdots \\
\boldsymbol{0} \\
\boldsymbol{0}
\end{pmatrix},
$$

where the matrices $\boldsymbol{A}_{i,i-1}$, $\boldsymbol{A}_{i,i}$ and $\boldsymbol{A}_{i,i+1}$ are defined as previously for the truncation of the state space. The (8×1) column vector $\boldsymbol{b}_k$ contains the probabilities that $\mathcal{X}$ enters state B from each state in level $L(k)$, that is, each of the states $\{(k, I), (k, II), ..., (k, VIII)\}$ and has entries given by

$$
(b_k)_j = q_{j,B}\left(\lambda_k + \mu_k + \sum_{\substack{j'=I \\ j' \neq j}}^{VIII} q_{j,j'} + q_{j,B}\right)^{-1}, \qquad j \in \{I, ..., VIII\}.
$$

Solving this matrix equation in the same recursive manner as before leads to Algorithm 3.2 for computing the rupture distribution. This algorithm takes into

$$\mathbf{H}_1 = \mathbf{I}_8 - \mathbf{A}_{1,1}$$

for $n = 2, ..., N$ **do**

$\quad \mid \quad \mathbf{H}_n = \mathbf{I}_8 - \mathbf{A}_{n,n-1}\mathbf{H}_{n-1}^{-1}\mathbf{A}_{n-1,n} - \mathbf{A}_{n,n}$

end

for $k = 1, ..., N$ **do**

$\quad \mathbf{s}_k = \mathbf{b}_k$

$\quad$ **for** $n = k + 1, ..., N$ **do**

$\quad\quad \mid \quad \mathbf{s}_n = \mathbf{A}_{n,n-1}\mathbf{H}_{n-1}^{-1}\mathbf{s}_{n-1}$

$\quad$ **end**

$\quad \mathbf{R}_N^{(k)} = \mathbf{H}_N^{-1}\mathbf{s}_N$ **for** $n = N - 1, ..., 1$ **do**

$\quad\quad \mid \quad \mathbf{R}_n^{(k)} = \mathbf{H}_n^{-1}\left(\mathbf{A}_{n,n+1}\mathbf{R}_{n+1}^{(k)} + \mathbf{s}_n\right)$

$\quad$ **end**

$\quad R_k = \boldsymbol{p}\mathbf{R}_1^{(k)}$

end

Algorithm 3.2: To compute the distribution of the number of bacteria released when an infected phagocyte ruptures.

account the one hour delay for the time the infecting bacterium spends in the phagosome by computing an averaged rupture distribution, weighted over the possible states that the 'rupture clock' can be in after one hour. Recall that the probability that $\mathcal{X}$ is in each of the states in the first level after one hour is stored in the vector $\boldsymbol{p}$. Let R_k denote the probability that a phagocyte releases k bacteria on rupture, provided it was initially infected by a single bacterium. After taking into account the effects of the phagosomal delay, it then follows that $R_k = \boldsymbol{p}\mathbf{R}_1^{(k)}$.

3.1.2 Within-host model

The within-host model proposed here is a birth-death-rupture model that replicates the dynamics of *F. tularensis* infection within the lung of an infected individual after inhaling an initial dose of bacteria. The model is largely an extension of

an original model by Wood *et al.* (2014) that utilises the within-phagocyte model to address previous limitations regarding inter-phagocyte variability. Within the lung, bacteria can be killed by host immune responses or through phagocytosis, provided that the phagocyte is activated. If the phagocyte is not capable of killing the ingested bacterium, the bacterium can escape the phagosome and enter the cytosol, resulting in proliferation and rupture of the host phagocyte, as described by the within-phagocyte model. These three key events of extracellular bacterial death, phagocytosis resulting in bacterial survival, and phagocyte rupture are incorporated into the within-host model, along with their effect on the population of bacteria and host phagocytes. The events are detailed as follows

- **Phagocytosis and bacterial survival** (rate $\alpha > 0$ (h^{-1})): This event refers to the phagocytosis of bacteria by an uninfected phagocyte that results in the bacteria escaping the phagosome, replicating, and ultimately leading to the rupture of the phagocyte.

- **Extracellular bacterial death** (rate $\mu > 0$ (h^{-1})): This event encompasses the effects of multiple host defence mechanisms that result in bacterial death, including the complement system, antibodies, antimicrobial peptides and activated phagocytes (Jones *et al.* (2012)). It is important to note that, although phagocytosis is also incorporated into this event, here it refers to phagocytosis followed by effective killing of the ingested bacterium.

- **Rupture of infected phagocytes** (rate $\delta = Median(T^{rupture})^{-1}$ (h^{-1})): Following phagocytosis that results in bacteria escaping the phagosome and proliferating, infected phagocytes will ultimately rupture and die. The number of bacteria released when they rupture is determined from the within-phagocyte model by means of the rupture distribution.

In Wood *et al.* (2014), these three events are referred to respectively as survival, death and birth. The term survival relates to bacteria entering an intracellular replicative niche where they escape the dangers of immune defence mechanisms, whilst the term birth indicates the instantaneous increase in the population of extracellular bacteria that results from the rupture of an infected phagocyte.

3. A MULTI-SCALE MODEL FOR *F. TULARENSIS* INFECTION

The within-host model can be described using a continuous-time two-dimensional Markov process $\mathcal{Y} = \{\boldsymbol{Y}(t) = (B(t), P(t)) : t \geq 0\}$, where $B(t)$ denotes the number of extracellular bacteria and bacteria-containing phagocytes at time $t \geq 0$, and $P(t)$ represents the number of bacteria-containing phagocytes at time $t \geq 0$, so that $B(t) \geq P(t)$ for all times $t \geq 0$. The variable $B(t)$ is defined in this manner since the within-host model does not explicitly account for the intracellular infection dynamics, however, this also highlights one of the advantages of a multi-scale framework. Incorporating the intracellular life cycle directly into the within-host model would add another level of complexity, resulting in a model that is both more difficult to study analytically and more time consuming to simulate. Instead, through developing the within-phagocyte model, the rupture distribution can be used to characterise these intracellular stages of infection and then applied to the within-host model to allow for variability between phagocytes.

An initial state of $\mathcal{Y}$, given by $\boldsymbol{Y}(0) = (k, 0)$, represents that k bacteria have reached the lung space following inhalation, whilst initially there are zero infected phagocytes. When the total population of bacteria reaches a threshold $M \in \mathbb{N}$, a response is said to occur, representing the onset of symptoms in the infected individual (Wood *et al.* (2014)). This state M, herein referred to as the response state, is one of two absorbing states of $\mathcal{Y}$; the other is state 0 and represents the elimination of bacteria from the lung before a response can occur; see Figure 3.5. Although state 0 cannot be reached if $\mathcal{Y}$ is already absorbed into the response state, M, this is not to say that individuals cannot recover once they display symptoms of tularemia. The purpose of this model is to determine how likely it is, and how long it takes, for an individual to show symptoms following exposure to a specific dose and does not currently consider recovery from infection.

A depiction of the within-host model is provided in Figure 3.5. The rate of rupture of an infected phagocyte is set as $\delta = Median\,[T^{\mathrm{rupture}}]^{-1}$ hours and the distribution of the number of bacteria released provides an extension to the original model by Wood *et al.* (2014), as indicated by dashed arrows. The rate at which an infected phagocyte ruptures, releasing k bacteria is given here by δR_k, and since $\sum_{k=1}^{+\infty} R_k = 1$, δ is analogous to how it is defined by Wood *et al.* (2014).
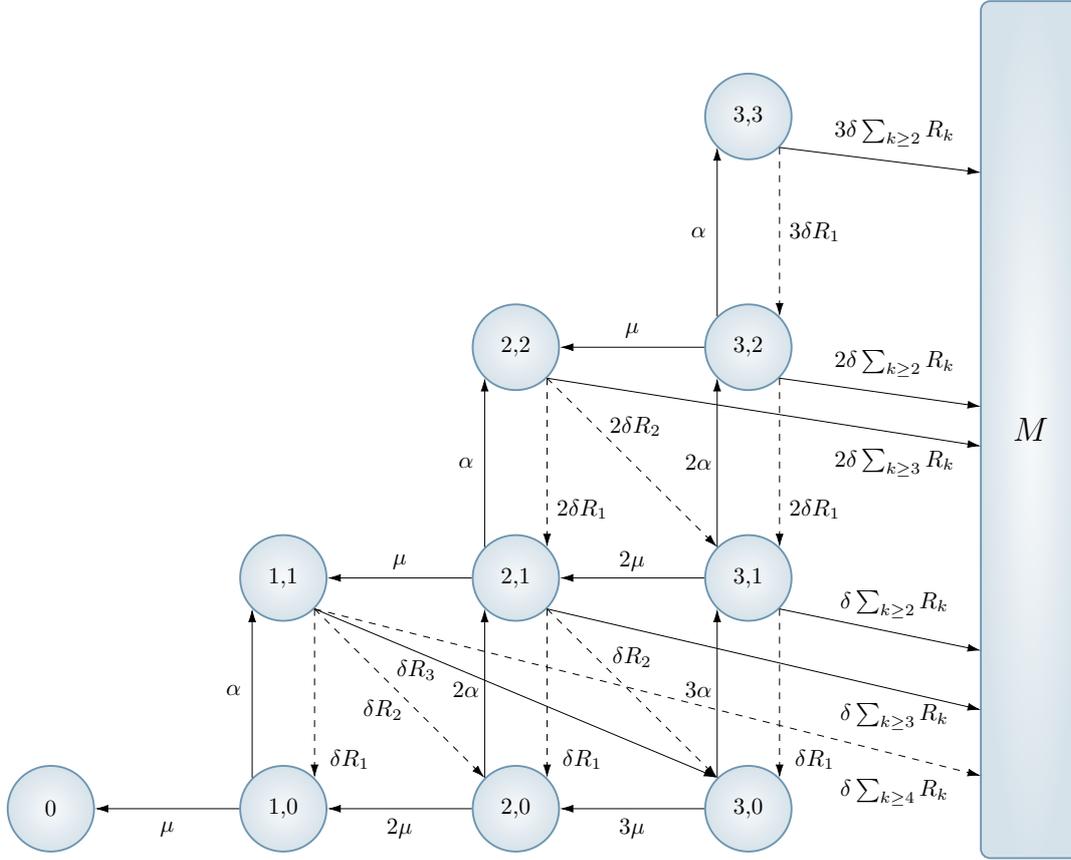
Figure 3.5: A depiction of the extended two-dimensional Markov chain with $M = 4$. State (i, j) represents i extracellular bacteria and bacteria-containing phagocytes, and j bacteria-containing phagocytes. The rates of rupture, phagocytosis, and death of extracellular bacteria are $\delta > 0$, $\alpha > 0$, and $\mu > 0$ respectively. Rupturing phagocytes release k bacteria with probability R_k. Solid arrows represent reactions included in the original model by Wood *et al.* (2014), where, for illustrative purposes, each phagocyte always releases $G = 3$ bacteria. Dashed arrows indicate reactions that are new to the extended within-host model described here.

For the within-host model, two summary statistics of interest are the probability of response and the mean response time. These quantities are also computed

analytically in Wood *et al.* (2014), however, the approach used is inefficient as it requires the inversion of the entire generator matrix associated with $\mathcal{Y}$. Given that the generator matrix associated with $\mathcal{Y}$ is a square matrix with order $2 + (M+2)(M-1)/2$, the value of M does not have to be large before inverting the matrix becomes computationally expensive. An alternative approach is outlined here that exploits the structure of the process $\mathcal{Y}$ and is therefore more efficient than the method used in Wood *et al.* (2014). Following this approach, the largest matrix that must be inverted only has order $(M-1)$. As for the within-phagocyte model, the state space of the within-host model, $\mathcal{S}_{\mathcal{Y}} = \{0\} \cup \{(i,j) : 1 \leq i \leq M-1, 0 \leq j \leq i\} \cup \{M\}$, can be organised into levels

$$
\begin{aligned}
L(0) &= \{(i,0) : 1 \leq i \leq M-1\}, \\
L(j) &= \{(i,j) : j \leq i \leq M-1\}, 1 \leq j \leq M-1,
\end{aligned}
$$

such that $\mathcal{S}_{\mathcal{Y}} = \{0\} \cup \bigcup_{j=0}^{M-1} L(j) \cup \{M\}$. With this grouping of states, $\mathcal{Y}$ has the form of a QBD process, and thus the tridiagonal block structure of the generator matrix can be exploited to yield efficient algorithms for the probability of response and mean response time.

Let $\pi_{(i,j)}$ be defined as the probability of response given that the initial state is $\boldsymbol{Y}(0) = (i,j)$, that is,

$$
\pi_{(i,j)} = \lim_{t \to +\infty} \Pr\left(\boldsymbol{Y}(t) = M | \boldsymbol{Y}(0) = (i,j)\right), \qquad 0 \leq j \leq i \leq M-1.
$$

If $\pi_{(i,j)}$ is stored in column vectors $\boldsymbol{\pi}_0 = \left(\pi_{(1,0)}, \pi_{(2,0)}, ..., \pi_{(M-1,0)}\right)^T$, and $\boldsymbol{\pi}_j = \left(\pi_{(j,j)}, \pi_{(j+1,j)}, ..., \pi_{(M-1,j)}\right)^T, 1 \leq j \leq M-1$, then first step arguments can be applied to give the matrix equation

$$
\begin{pmatrix} \boldsymbol{\pi}_0 \\ \boldsymbol{\pi}_1 \\ \vdots \\ \boldsymbol{\pi}_{M-2} \\ \boldsymbol{\pi}_{M-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{G}_{00} & \boldsymbol{G}_{01} & & & \\ \boldsymbol{G}_{10} & \boldsymbol{G}_{11} & \boldsymbol{G}_{12} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{G}_{M-2,M-1} & \boldsymbol{G}_{M-2,M-2} & \boldsymbol{G}_{M-2,M-1} \\ & & & \boldsymbol{G}_{M-1,M-2} & \boldsymbol{G}_{M-1,M-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\pi}_0 \\ \boldsymbol{\pi}_1 \\ \vdots \\ \boldsymbol{\pi}_{M-2} \\ \boldsymbol{\pi}_{M-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{c}_0 \\ \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_{M-2} \\ \boldsymbol{c}_{M-1} \end{pmatrix}.
$$

The non-zero entries of matrices $\boldsymbol{G}_{j,j-1}$ are given by

$$(\boldsymbol{G}_{1,0})_{i,i+k-1} = \frac{\delta R_k}{(\mu+\alpha)(i-1)+\delta}, \quad i=1,...,M-1, k=1,...,M-i$$

$$(\boldsymbol{G}_{j,j-1})_{i-j+1,i-j+1+k} = \frac{\delta j R_k}{(\mu+\alpha)(i-j)+\delta j}, \quad i=j,...,M-1, \, j=2,...,M-1,$$

$$k=1,...,M-i,$$

whilst the non-zero entries of $\boldsymbol{G}_{j,j}$ satisfy

$$(\boldsymbol{G}_{0,0})_{i,i-1} = \frac{\mu}{(\mu+\alpha)}, \quad i=2,...,M-1,$$

$$(\boldsymbol{G}_{j,j})_{i-j+1,i-j} = \frac{\mu(i-j)}{(\mu+\alpha)(i-j)+\delta j}, \quad i=j+1,...,M-1, \, j=1,...,M-2.$$

If $j=M-1$, the corresponding level consists of only a single state and hence $\boldsymbol{G}_{M-1,M-1}=0$. The non-zero entries of the matrices $\boldsymbol{G}_{j,j+1}$ are

$$(\boldsymbol{G}_{0,1})_{i,i} = \frac{\alpha}{(\mu+\alpha)}, \quad i=1,...,M-1,$$

$$(\boldsymbol{G}_{j,j+1})_{i-j+1,i-j} = \frac{\alpha(i-j)}{(\mu+\alpha)(i-j)+\delta j}, \quad i=j+1,...,M-1, \, j=1,...,M-2.$$

Finally, the vectors $\boldsymbol{c}_j$ contain the probabilities that $\mathcal{Y}$ transitions into the response state, M, from states in level j. Since it is not possible to enter the response state from level 0, $\boldsymbol{c}_0 = \boldsymbol{0}$, whilst for the remaining levels, the non-zero entries of $\boldsymbol{c}_j$ are defined as

$$(\boldsymbol{c}_j)_i = \frac{\delta j}{(\mu+\alpha)(i-j)+\delta j}\left(\sum_{k\geq M-i+1} R_k\right), \quad i=j,...,M-1, \, j=1,...,M-1.$$

A solution to the previous matrix equation is provided in Algorithm 3.3. Since states representing initial doses of bacteria are contained in level 0, $\boldsymbol{\pi}_0$ is the most informative quantity found using this algorithm.

To consider the time taken for an infected individual to respond, let this be defined as the first time at which the number of extracellular bacteria and infected phagocytes is equal to M, that is, $T_{(i,j)} = inf\{t \geq 0 : B(t) = M \,|\, \boldsymbol{Y}(0) = (i,j)\}$.

$$\boldsymbol{H}_0 = \boldsymbol{I}_{M-1} - \boldsymbol{G}_{0,0} \,, \boldsymbol{s}_0 = \boldsymbol{c}_0$$

for $n = 1, ..., M-1$ **do**

$\quad\Big|\quad \boldsymbol{H}_n = \boldsymbol{I}_{M-n} - \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{G}_{n-1,n} - \boldsymbol{G}_{n,n}$

$\quad\Big|\quad \boldsymbol{s}_n = \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{s}_{n-1} + \boldsymbol{c}_n$

end

$$\boldsymbol{\pi}_{M-1} = \boldsymbol{H}_{M-1}^{-1}\boldsymbol{s}_{M-1}$$

for $n = M-2, ..., 0$ **do**

$\quad\Big|\quad \boldsymbol{\pi}_n = \boldsymbol{H}_n^{-1}\left(\boldsymbol{G}_{n,n+1}\boldsymbol{\pi}_{n+1} + \boldsymbol{s}_n\right)$

end

Algorithm 3.3: To compute the probability of response for states $(i,j) \in \mathcal{S}_y$.

Since absorption into the response state is not certain, there is no guarantee that the time to response will be finite; the time to response will be infinite if $\mathcal{Y}$ is instead absorbed into state 0. Due to this, the *restricted* mean time may first be considered, after which the *conditioned* mean time can be obtained by dividing by the probability of response, $\pi_{(i,j)}$. The restricted and conditioned mean times are given respectively by

$$r_{(i,j)} = \mathbb{E}\left[T_{(i,j)} 1_{T_{(i,j)} < +\infty}\right], \qquad 0 \le j \le i \le M-1\,,$$

$$m_{(i,j)} = \mathbb{E}\left[T_{(i,j)} \,|\, T_{(i,j)} < +\infty\right] = \frac{r_{(i,j)}}{\pi_{(i,j)}}\,, \qquad 0 \le j \le i \le M-1\,,$$

where $1_{T_{(i,j)} < +\infty}$ is an indicator variable that is equal to one if $T_{(i,j)} < +\infty$ and zero otherwise. If the restricted mean times $r_{(i,j)}$ are stored in column vectors $\boldsymbol{r}_0 = \left(r_{(1,0)}, r_{(2,0)}, ..., r_{(M-1,0)}\right)^T$, and $\boldsymbol{r}_j = \left(r_{(j,j)}, r_{(j+1,j)}, ..., r_{(M-1,j)}\right)^T$, $1 \le j \le M-1$, then first step arguments can be used to obtain the matrix equation

$$\begin{pmatrix} \boldsymbol{r}_0 \\ \boldsymbol{r}_1 \\ \vdots \\ \boldsymbol{r}_{M-2} \\ \boldsymbol{r}_{M-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{G}_{00} & \boldsymbol{G}_{01} & & & \\ \boldsymbol{G}_{10} & \boldsymbol{G}_{11} & \boldsymbol{G}_{12} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{G}_{M-2,M-1} & \boldsymbol{G}_{M-2,M-2} & \boldsymbol{G}_{M-2,M-1} \\ & & & \boldsymbol{G}_{M-1,M-2} & \boldsymbol{G}_{M-1,M-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_0 \\ \boldsymbol{r}_1 \\ \vdots \\ \boldsymbol{r}_{M-2} \\ \boldsymbol{r}_{M-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{d}_0 \\ \boldsymbol{d}_1 \\ \vdots \\ \boldsymbol{d}_{M-2} \\ \boldsymbol{d}_{M-1} \end{pmatrix}\,,$$

where the vectors $\boldsymbol{d}_j$ are given by:

$$(\boldsymbol{d}_0)_i = \frac{i\left(\alpha\pi_{i,1} + \mu\pi_{i-1,0}\right)}{(i(\mu+\alpha))^2} \qquad 1 \le i \le M-1\,,$$

$$(\boldsymbol{d}_j)_i = \frac{(i-j)\left(\alpha\pi_{(i,j+1)} + \mu\pi_{(i-1,j)}\right) + \delta j \left(\sum_{k=1}^{M-i} R_k \pi_{(i+k-1,j-1)} + \sum_{k=M-i+1}^{N} R_k\right)}{((i-j)(\alpha+\mu) + \delta j)^2}\,,$$

for $j = 1, ..., M-1$, $i = j, ..., M-1$. Solving this matrix equation yields Algorithm 3.4 for computing the conditioned mean response time, $m_{(i,j)}$. The algorithm makes use of $\boldsymbol{m}_n = \mathcal{D}(\boldsymbol{\pi}_n)^{-1}\boldsymbol{r}_n$, where $\mathcal{D}(\boldsymbol{\pi}_n)$ is an $(n \times n)$ diagonal matrix whose diagonal entries are constructed from the vector $\boldsymbol{\pi}_n$.

3.1.3 Population-level model

With two models of *F. tularensis* infection that describe the dynamics at both the within-phagocyte and within-host levels, it is possible to use results from these to construct a population scale model, the final level of the multi-scale model. Outbreaks of tularemia have previously been reported at a population level by microbiology laboratories as a result of infection by *F. tularensis* (Shapiro & Schwartz (2002)). Since *F. tularensis* can cause a potentially lethal disease, clinicians are required to notify the laboratory when tularemia is suspected in a given patient. However, due to the nonspecific symptoms associated with tularemia, these cases can often be undetected and result in outbreaks within the laboratory through interactions with contaminated samples, as in the outbreak reported by Shapiro & Schwartz (2002). If laboratory personnel are identified as having tularemia, manipulation of the corresponding samples must be carried out under strict control measures, such as Biosafety Level 2 (BSL-2) or BSL-3 conditions. Should these control measures not be taken, further release of *F. tularensis* into the air could occur, facilitating its dispersal and spread. Particular activities within a laboratory that are believed to promote the release of bacteria when not performed with care include centrifuging and vigorous shaking (SFDPH (2008)).

Recently, work has been carried out to study the airborne spread of pathogens

$$\boldsymbol{H}_0 = \boldsymbol{I}_{M-1} - \boldsymbol{G}_{0,0} \, , \boldsymbol{s}_0 = \boldsymbol{d}_0$$

for $n = 1, ..., M - 1$ **do**

$\quad \boldsymbol{H}_n = \boldsymbol{I}_{M-n} - \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{G}_{n-1,n} - \boldsymbol{G}_{n,n}$

$\quad \boldsymbol{s}_n = \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{s}_{n-1} + \boldsymbol{d}_n$

end

$$\boldsymbol{r}_{M-1} = \boldsymbol{H}_{M-1}^{-1}\boldsymbol{s}_{M-1}$$

for $n = M - 2, ..., 0$ **do**

$\quad \boldsymbol{r}_n = \boldsymbol{H}_n^{-1}\left(\boldsymbol{G}_{n,n+1}\boldsymbol{r}_{n+1} + \boldsymbol{s}_n\right)$

$\quad \boldsymbol{m}_n = \mathcal{D}(\boldsymbol{\pi}_n)^{-1}\boldsymbol{r}_n$

end

Algorithm 3.4: To compute the conditioned mean response time for states $(i, j) \in \mathcal{S}_y$.

within a facility by taking into consideration the ventilation regime in place. One particular example of this looks at the airborne spread of bacteria in health care facilities (Liao *et al.* (2005)). In these situations, zonal ventilation models prove an extremely useful approach to link the dynamics of airflow around the facility with epidemic dynamics at the population level (Noakes & Sleigh (2009)). Here, the scenario is considered of a laboratory consisting of two rooms linked by a corridor. Suppose that at $t = 0$, a fixed amount of *F. tularensis* bacteria are released at a source location in a given room. This could be the result of manipulation of contaminated samples or an accident that causes airborne release of bacteria. The aim of the population-level model is to estimate the total number of laboratory personnel who would develop symptoms of tularemia in the near future, under different ventilation regimes. A ventilation regime is defined here to be a description of the dynamics of airflow between the two rooms and the corridor, as well as the extraction of air away from the laboratory facility.

The approach used to model the release of bacteria within the laboratory follows that by Noakes & Sleigh (2009), recently extended by López-García *et al.* (2019), where a system of ODEs is used to model the airborne concentration of *F. tularensis* bacteria in different spatial compartments of the laboratory. A ven-

tilation regime is first described by splitting the laboratory into ventilation zones, where the assumption is made that the air is well mixed within each zone. Air can then flow between different ventilation zones, resulting in different pathogen concentrations in the air for each zone (Noakes & Sleigh (2009)). Since the air is well mixed, individuals in the same ventilation zones are assumed to have equal probability of inhaling *F. tularensis* bacteria. This approach can be refined by partitioning the laboratory into many more ventilation zones, as considered in Parker & Williamson (2016). A depiction of the scenario studied here, whereby a microbiology facility is divided into six zones, is provided in Figure 3.6.

Let $C_i(t)$ denote the concentration of bacteria in the air in zone i at time t, and $p_i(t)$ be the cumulative amount of bacteria inhaled by each individual in zone i up to and including time t. The concentration of bacteria in zone j then increases due to air flowing into the zone from neighbouring zones, i, with rate $\beta_{i,j}^{(v)}$ (m^3/minute), and decreases due to air flowing from zone j into neighbouring zones with rate $\beta_{j,i}^{(v)}$, air being extracted from the facility with rate $Q_j^{(v)}$ (m^3/minute), and individuals within the facility inhaling bacteria with rate ρ (m^3/minute). Since tularemia is currently not known to spread from person to person, the population-level model does not allow for forward infection and individuals can only inhale bacteria released from the initial source. In each of the rates introduced here, the index v denotes the ventilation setting under consideration, which is also used in specifying the initial concentrations $C_i(0)$. By considering in this way the factors that affect the concentration of bacteria in each zone, the following system of ODEs for $C_j^{(v)}(t)$ and $p_j^{(v)}(t)$ may be constructed

$$V_1 \dot{C}_1^{(v)} = - \left(Q_1^{(v)} + \beta_{12}^{(v)} + \rho n_1 \right) C_1^{(v)}(t) + \beta_{21}^{(v)} C_2^{(v)}(t) \,,$$

$$V_2 \dot{C}_2^{(v)} = - \left(Q_2^{(v)} + \beta_{21}^{(v)} + \beta_{23}^{(v)} + \rho n_2 \right) C_2^{(v)}(t) + \beta_{12}^{(v)} C_1^{(v)}(t) + \beta_{32}^{(v)} C_3^{(v)}(t) \,,$$

$$V_3 \dot{C}_3^{(v)} = - \left(Q_3^{(v)} + \beta_{32}^{(v)} + \beta_{36}^{(v)} \right) C_3^{(v)}(t) + \beta_{23}^{(v)} C_2^{(v)}(t) + \beta_{63}^{(v)} C_6^{(v)}(t) \,,$$

$$V_4 \dot{C}_4^{(v)} = - \left(Q_4^{(v)} + \beta_{45}^{(v)} + \rho n_4 \right) C_4^{(v)}(t) + \beta_{54}^{(v)} C_5^{(v)}(t) \,,$$

$$V_5 \dot{C}_5^{(v)} = - \left(Q_5^{(v)} + \beta_{54}^{(v)} + \beta_{56}^{(v)} + \rho n_5 \right) C_5^{(v)}(t) + \beta_{54}^{(v)}(t) + \beta_{65}^{(v)} C_6^{(v)}(t) \,,$$
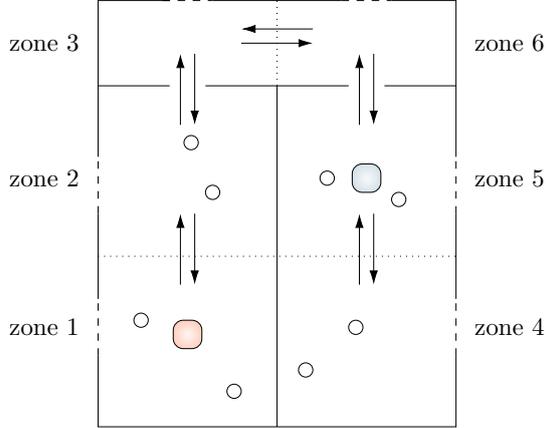
Figure 3.6: A diagram showing the setup of two rooms and an adjoining corridor within a laboratory, split into six ventilation zones. Dotted lines represent the partitioning of each room and corridor into zones, arrows between zones show potential airflow, and dashed lines represent potential ventilation systems that extract air from within each zone. Individuals are represented by circles whilst red and blue squares indicate two possible release locations of bacteria.

$$V_6 \dot{C}_6^{(v)}(t) = -\left(Q_6^{(v)} + \beta_{65}^{(v)} + \beta_{63}^{(v)}\right) C_6^{(v)}(t) + \beta_{36}^{(v)} C_3^{(v)}(t) + \beta_{56}^{(v)} C_5^{(v)}(t) \,,$$

$$\dot{p}_k^{(v)}(t) = \rho C_k^{(v)}(t) \qquad k = 1, 2, 4, 5 \,,$$

where V_j and n_j denote the respective volume (m^3), and number of individuals in zone j. In the example described here, $n_i = 2$ for $i \in \{1, 2, 4, 5\}$, representing two individuals in each zone, aside from the corridor, during the release of bacteria. If the initial number of *F. tularensis* bacteria released in source room i is B_0, then the initial conditions for each equation are $C_i^{(v)}(0) = B_0/V_i$ and $C_j^{(v)}(0) = 0$ for $j \neq i$. It is assumed that to begin with, individuals are not infected with *F. tularensis* bacteria, that is, $p_i^{(v)}(0) = 0$, $i \in \{1, 2, 4, 5\}$.

To link the within-host and population-level models, it is assumed that the cumulative number of bacteria inhaled by each individual reaches steady state at a shorter timescale than the timescale of the within-host infection dynamics.

Under this assumption, the steady state values may then be treated as the initial doses for the within-host model. The output from the within-host model may then provide the probability of response and mean response time for each individual in the facility as a function of the number of bacteria they inhale.

3.2 Local sensitivity analysis

3.2.1 Within-host model

Local sensitivity analysis can be used to determine the effects that perturbations in the within-host model parameters, α, μ and δ, have on the probability of response. The method used here follows a matrix calculus approach originally described in Caswell (2011) and more recently extended by Gómez-Corral & López-García (2018) for QBD processes. This method involves finding partial derivatives of each stochastic descriptor with respect to each parameter. If $\boldsymbol{\theta} = (\alpha, \mu, \delta)$ is the vector of parameters, then Algorithm 3.5 provides a method for computing the sensitivities of the probability of response. This algorithm is obtained by differentiating the expressions for $\boldsymbol{H}_n$, $\boldsymbol{s}_n$ and $\boldsymbol{\pi}_n$ given in Algorithm 3.3 with respect to $\boldsymbol{\theta}$, where the elementwise derivatives of $\boldsymbol{G}_{n,n}$, $\boldsymbol{G}_{n-1,n}$ and $\boldsymbol{G}_{n,n-1}$ are also required. Here, the vec operator transforms a $(m \times n)$ matrix into a $(mn \times 1)$ vector by stacking the columns of the matrix. Due to computational resources available, these elasticities have been computed using $M = 500$ which is significantly lower than the median value, $M_{\mathrm{med}} = 2.4 \times 10^{11}$, suggested by Wood et al. (2014). Despite this, the elasticities are still informative since it is not necessary to consider such large values of M. It is noted by Wood et al. (2014) that once $B(t)$, the number of extracellular bacteria and infected phagocytes, reaches a specific value, a response is almost certain to occur. This value, referred to as the 'point of no return' was previously predicted to be approximately equal to 500 bacteria.

The elasticities of the probability of response evaluated for selected initial doses are provided in Table 3.1. From here it is possible to see that δ does not appear to be an important parameter, with equally low elasticities reported at

$$\frac{\text{dvec}\boldsymbol{H}_0}{d\boldsymbol{\theta}^T} = -\frac{\text{dvec}\boldsymbol{G}_{00}}{d\boldsymbol{\theta}^T} \ , \ \frac{d\boldsymbol{s}_0}{d\boldsymbol{\theta}^T} = \frac{d\boldsymbol{b}_0}{d\boldsymbol{\theta}^T}$$

for $n = 1, ..., M - 1$ **do**

$$\frac{\text{dvec}\boldsymbol{H}_n}{d\boldsymbol{\theta}^T} = -\left(\left(\boldsymbol{H}_{n-1}^{-1}\boldsymbol{G}_{n-1,n}\right)^T \otimes \boldsymbol{I}_{M-n}\right) \frac{\text{dvec}\boldsymbol{G}_{n,n-1}}{d\boldsymbol{\theta}^T}$$

$$+ \left(\left(\boldsymbol{H}_{n-1}^{-1}\boldsymbol{G}_{n-1,n}\right)^T \otimes \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\right) \frac{\text{dvec}\boldsymbol{H}_{n-1}}{d\boldsymbol{\theta}^T}$$

$$- \left(\boldsymbol{I}_{M-n} \otimes \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\right) \frac{\text{dvec}\boldsymbol{G}_{n-1,n}}{d\boldsymbol{\theta}^T} - \frac{\text{dvec}\boldsymbol{G}_{n,n}}{d\boldsymbol{\theta}^T}$$

$$\frac{d\boldsymbol{s}_n}{d\boldsymbol{\theta}^T} = \left(\left(\boldsymbol{H}_{n-1}^{-1}\boldsymbol{s}_{n-1}\right)^T \otimes \boldsymbol{I}_{M-n}\right) \frac{\text{dvec}\boldsymbol{G}_{n,n-1}}{d\boldsymbol{\theta}^T}$$

$$- \left(\left(\boldsymbol{H}_{n-1}^{-1}\boldsymbol{s}_{n-1}\right)^T \otimes \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\right) \frac{\text{dvec}\boldsymbol{H}_{n-1}}{d\boldsymbol{\theta}^T}$$

$$+ \boldsymbol{G}_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\frac{d\boldsymbol{s}_{n-1}}{d\boldsymbol{\theta}^T} + \frac{d\boldsymbol{b}_n}{d\boldsymbol{\theta}^T}$$

end

$$\frac{d\boldsymbol{\pi}_{M-1}}{d\boldsymbol{\theta}^T} = -\left(\left(\boldsymbol{H}_{M-1}^{-1}\boldsymbol{s}_{M-1}\right)^T \otimes \boldsymbol{H}_{M-1}^{-1}\right) \frac{\text{dvec}\boldsymbol{H}_{M-1}}{d\boldsymbol{\theta}^T} + \boldsymbol{H}_{M-1}^{-1}\frac{d\boldsymbol{s}_{M-1}}{d\boldsymbol{\theta}^T}$$

for $n = M - 2, ..., 0$ **do**

$$\frac{d\boldsymbol{\pi}_n}{d\boldsymbol{\theta}^T} = -\left(\left(\boldsymbol{H}_n^{-1}\left(\boldsymbol{G}_{n,n+1}\boldsymbol{\pi}_{n+1} + \boldsymbol{s}_n\right)\right)^T \otimes \boldsymbol{H}_n^{-1}\right) \frac{\text{dvec}\boldsymbol{H}_n}{d\boldsymbol{\theta}^T}$$

$$+ \left(\boldsymbol{\pi}_{n+1}^T \otimes \boldsymbol{H}_n^{-1}\right) \frac{\text{dvec}\boldsymbol{G}_{n,n+1}}{d\boldsymbol{\theta}^T}$$

$$+ \boldsymbol{H}_n^{-1}\left(\frac{d\boldsymbol{s}_n}{d\boldsymbol{\theta}} + \boldsymbol{G}_{n,n+1}\frac{d\boldsymbol{\pi}_{n+1}}{d\boldsymbol{\theta}^T}\right)$$

end

Algorithm 3.5: To compute the sensitivities of the probability of response with respect to the parameters $\boldsymbol{\theta} = (\alpha, \mu, \delta)$.

each of the three doses. When computing the elasticity for δ, the relationship it has with the within-phagocyte model has been ignored. The reciprocal of δ is the median time until rupture, which is determined by the median of the log-normal distribution on the time until rupture. Therefore, perturbations in δ would affect this log-normal distribution, although it is not believed that this effect would be noticeable once the log-normal distribution has been approximated by a PH distribution.

More interestingly, the elasticities with respect to α and μ are the negative of each other, suggesting that the probability of response depends on the ratio

	Initial dose	α	μ	δ
$\dfrac{\dfrac{\partial \pi_{(i,j)}}{\partial \theta}}{\dfrac{\pi_{(i,j)}}{\theta}}$	1	1.01	-1.01	4.38×10^{-6}
	10	8.53×10^{-1}	-8.53×10^{-1}	4.38×10^{-6}
	100	1.06×10^{-1}	-1.06×10^{-1}	4.38×10^{-6}

Table 3.1: Elasticities for the probability of response, with respect to the within-host parameters α, μ and δ, for states (1,0), (10,0) and (100,0).

α/μ instead of the individual parameters. Furthermore, the magnitude of these elasticities decays as the initial dose increases, which can be seen more clearly in Figure 3.7. This is linked to the idea of a 'point of no return', since, if the initial dose is large enough then a response is almost certain to occur and small perturbations in either α or μ will not effect this. A plot showing the elasticities with respect to δ has not been included since the elasticities in Table 3.1 have already concluded that this parameter is not important.

Although it is informative to compute the elasticities of the probability of response for a reduced value of M, the same is not true for the mean response time. The mean response time will naturally be lower for smaller values of M and it is difficult to conclude whether the elasticities computed in this way would be reflective of those for a larger, more realistic value of M.

3.3 Parameter values

In this section, I will use a Bayesian approach to estimate the parameters in the within-phagocyte and within-host model. The parameter values for the population-level model are also reported according to those provided by Noakes & Sleigh (2009) for the airborne spread of bacteria within a health care facility.

3.3.1 Within-phagocyte model

In order to use the within-phagocyte model to compute the rupture distribution, the parameters λ and C must first be estimated. These parameters are
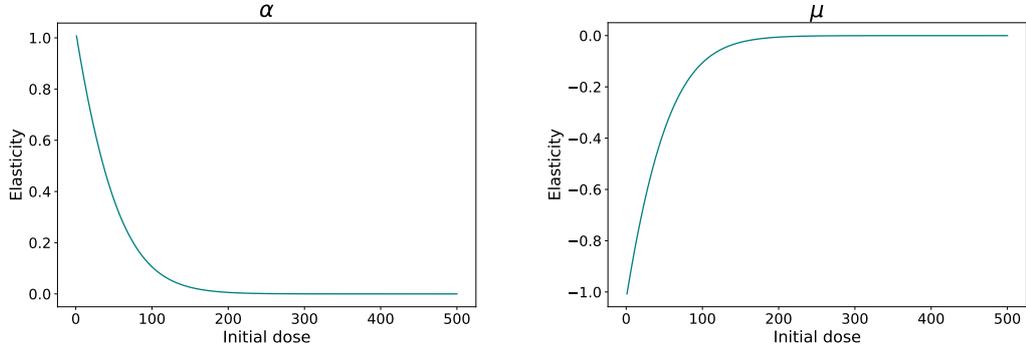
Figure 3.7: Elasticities for the probability of response with respect to α (left) and μ (right) as a function of the initial dose.

estimated using *in vitro* data for the number of intracellular bacteria contained within a phagocyte. The corresponding experiments involved infecting human macrophages with *F. tularensis* bacteria and lysing the cells at 1, 8, 16, 24 and 48 hours post infection to enumerate viable bacteria (Lindemann *et al.* (2011)). The results used here in the inference are representative of the average bacterial load of a single cell. In the experiment, the number of intracellular bacteria is only recorded for viable phagocytes, that is, phagocytes that have not ruptured at the time of measurement. Therefore, when fitting the within-phagocyte model to this data, rupture events should be neglected. A stochastic logistic growth process without catastrophe is instead fitted to the data in order to obtain λ and C, a depiction of this process is given in Figure 3.8.

A sequential Approximate Bayesian Computation (ABC) method is used to find estimates for these parameters. Prior distributions for each parameter are selected to be $\lambda \sim U(0.01, 1) \, h^{-1}$ and $C \sim U(100, 1500)$ bacteria based on previous estimates of $\lambda = 0.212 \, h^{-1}$ and $C = 384$ bacteria obtained by Wood *et al.* (2014). For each pair (λ, C) sampled from their respective prior distributions, the number of intracellular bacteria is obtained at time points corresponding to those at which experimental measurements were recorded. The number of intracellular bacteria is determined at each time point by simulating the stochastic logistic growth process using the Gillespie algorithm, as introduced in Section 2.4.1. With
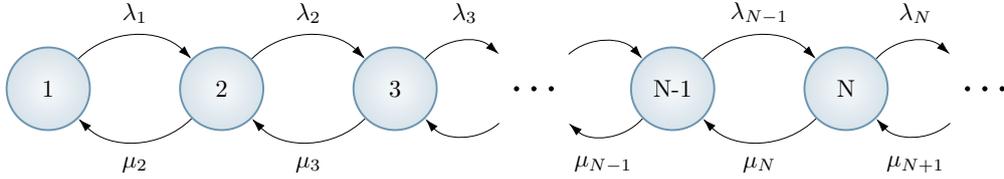
Figure 3.8: A depiction of the stochastic logistic growth process without catastrophe used in the estimation of within-phagocyte model parameters λ and C.

this, the model is compared to the data by Lindemann *et al.* (2011) using the Euclidean distance. If T denotes the set of time points at which experimental measurements were recorded, $x^{(M)}(t)$ denotes the value of the stochastic logistic growth process at time $t \in T$, and $x^{(E)}(t)$ represents the experimentally observed number of intracellular bacteria at time $t \in T$, then the distance between the model predictions and experimental data is given by

$$d(M, E) = \left(\sum_{t \in T} \left(x^{(E)}(t) - x^{(M)}(t) \right)^2 \right)^{\frac{1}{2}} .$$

Following the ABC rejection sampling algorithm (see Section 2.5.1), parameter pairs are then only accepted if $d(M, E) < \varepsilon$. Here, since the size of the prior parameter space is large, and parameter pairs in extreme regions of this space can result in longer simulation times for the Gillespie algorithm, the ABC algorithm is implemented sequentially for successively smaller tolerances, ε. Initially, the tolerance is chosen to be large, $\varepsilon = 100$, in order to refine the parameter space. Following this, the prior distributions are adjusted accordingly and the ABC algorithm is repeated for a smaller tolerance, $\varepsilon = 50$. Continuing in this manner for tolerances $\varepsilon = 25$ and $\varepsilon = 15$, an approximate posterior distribution is obtained. It is important that after each implementation of the ABC algorithm, the prior distributions are adjusted conservatively as not to inadvertently exclude points that ultimately have high density in the theoretical posterior distribution. The tolerances, ε, should also be chosen after a preliminary exploration of the sample space and should be influenced by the computational resources available along with the desired size of the posterior sample.
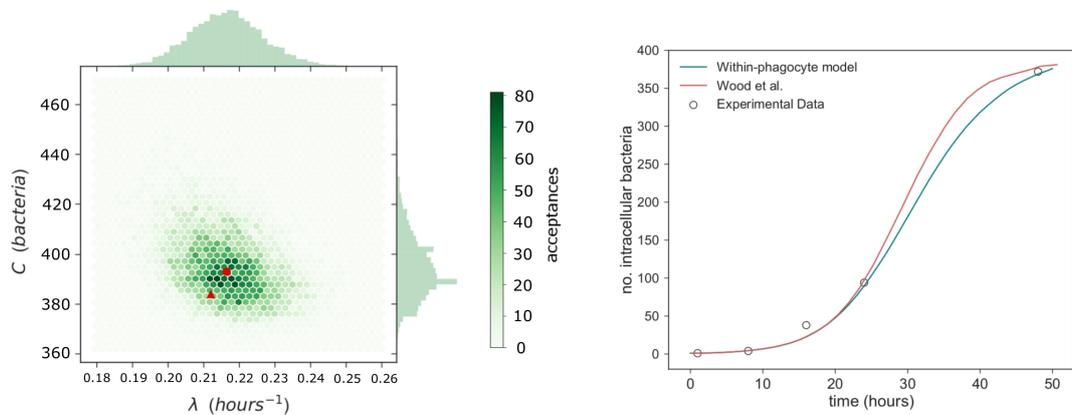
Figure 3.9: Left: a bivariate histogram of λ and C accepted values as a result of the ABC method, with median values marked with a red circle. Corresponding estimates by Wood *et al.* (2014), ($\lambda = 0.212\,h^{-1}$, $C = 384$ bacteria), are indicated with a red triangle. Right: Mean number of bacteria within an infected phagocyte, predicted using the stochastic logistic growth process with posterior median values of λ and C (blue), and compared to the theoretical predictions by Wood *et al.* (2014) (red) and experimental data by Lindemann *et al.* (2011) (circles).

A bivariate histogram of the sample posterior distribution obtained using this approach is provided in Figure 3.9, with the median of the posterior sample indicated alongside the corresponding estimates obtained by Wood *et al.* (2014). The posterior sample found here contains 10^4 parameter pairs, with a summary of this sample provided in Table 3.2. Figure 3.9 also shows a comparison between the experimental data and the average of 10^4 simulations of the stochastic logistic growth process for posterior median values of λ and C.

3.3.2 Within-host model

To determine estimates for α and μ, Wood *et al.* (2014) use a nonlinear least squares approach to fit their within-host model to *in vivo* experimental data for the number of extracellular bacteria during the initial 48 hours post infection. Since their model assumes that a fixed number of bacteria are released on rup-

	Min.	1$^{\text{st}}$ Qu.	Median	Mean	3$^{\text{rd}}$ Qu.	Max.
λ	1.8×10^{-1}	2.1×10^{-1}	2.2×10^{-1}	2.2×10^{-1}	2.2×10^{-1}	2.6×10^{-1}
C	3.6×10^2	3.9×10^2	3.9×10^2	4.0×10^2	4.0×10^2	4.7×10^2

Table 3.2: Summary statistics for the approximate posterior sample of the within-phagocyte model parameters ($[\lambda] = h^{-1}$, $[C] = $ bacteria).

ture of an infected phagocyte but the within-host model described here depends on a rupture distribution obtained from the within-phagocyte model, the parameters α and μ must also be estimated here. The ABC rejection sampling algorithm is again used to estimate α and μ using the same within-host infection data used by Wood *et al.* (2014). In the two studies considered, monkeys were infected with aerosolised *F. tularensis* bacteria and pairs of monkeys were subsequently sacrificed at selected times during the first 48 hours of infection. The number of viable bacteria in the lungs was then quantified, with the data used here representing the average bacterial load (Eigelsbach *et al.* (1962); White *et al.* (1964)). Before implementing the fitting procedure, the rupture distribution is obtained using the approach outlined in Section 3.1.1 and Algorithm 3.2 for posterior median values of λ and C. Due to the computational time required to calculate the rupture distribution, this same distribution is used in each iteration of the ABC algorithm. However, in keeping with Wood *et al.* (2014), and to represent heterogeneity of the susceptibility of infection between individuals, the number of extracellular bacteria required for a response to occur is a random variable $M \sim logN(26.2, 6.05)$ (Saslaw *et al.* (1961); Sawyer *et al.* (1966); Wood *et al.* (2014)). Therefore, when evaluating the within-host model, a new value of M is sampled for each iteration from a $logN(26.2, 6.05)$ distribution. Due to the potentially large values of M that can be sampled, a tau-leaping procedure with adaptive step size, as described in Section 2.4.2, has been implemented in order to simulate the model. The Python code for simulating the within-host model using a tau-leaping procedure has been included in Section A.1 of Appendix A.

The two data sets used to estimate α and μ concern the number of extracellular bacteria observed in the lungs of infected monkeys over the first 48 hours
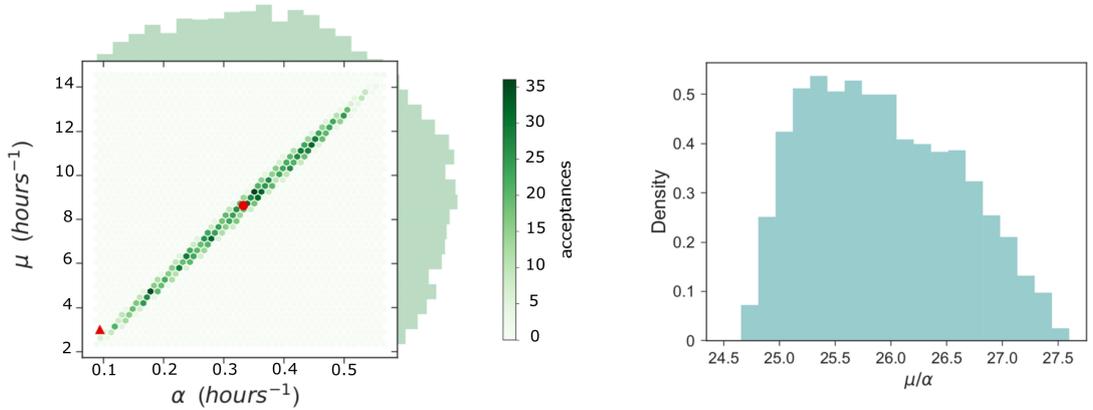
Figure 3.10: Left: A bivariate histogram for the parameters α and μ obtained as a result of the ABC procedure for the within-host model. The posterior median values are indicated with a red circle and the corresponding estimates ($\alpha = 0.939\,h^{-1}$, $\mu = 3\,h^{-1}$) found by Wood *et al.* (2014) are marked with a red triangle. Right: A posterior histogram for the ratio μ/α.

post infection (Eigelsbach *et al.* (1962); White *et al.* (1964)). Prior distributions assumed for each parameter are $\alpha \sim U(0,1)\,h^{-1}$ and $\mu \sim U(0,25)\,h^{-1}$. Because of the smaller ranges considered for these prior distributions compared to those initially used when inferring λ and C in the within-phagocyte model, the ABC rejection sampling algorithm is applied once, as opposed to sequentially. A total of 2×10^5 iterations of the ABC algorithm have been performed, with an acceptance rate of 1% resulting in an approximate posterior sample size of 2×10^3. Due to the large orders of magnitude for the number of extracellular bacteria experimentally observed within the host, the log-ratio of the model predictions and experimental data is considered in the distance measure. Let T_1 and T_2 denote the sets of times at which measurements were recorded by Eigelsbach *et al.* (1962) and White *et al.* (1964) respectively. The distance between the within-host model predictions, $x^{(M)}(t)$, and experimental data, $x^{(E)}(t)$, is then given by

$$d^2(M,E) = \sum_{t \in T_1} \left(\log(x^{(M)}(t)) - \log(x_1^{(E)}(t)) \right)^2 + \sum_{t \in T_2} \left(\log(x^{(M)}(t)) - \log(x_2^{(E)}(t)) \right)^2 .$$

The results of the ABC lead to the posterior bivariate histogram in Figure 3.10,

	Min.	1ˢᵗ Qu.	Median	Mean	3ʳᵈ Qu.	Max.
α	8.9×10^{-2}	2.4×10^{-1}	3.3×10^{-1}	3.3×10^{-1}	4.1×10^{-1}	5.7×10^{-1}
μ	2.3	6.3	8.6	8.4	10.7	14.6

Table 3.3: Summary statistics for the approximate posterior sample of the within-host model parameters ($[\alpha] = [\mu] = h^{-1}$).

with the approximate posterior distributions for each parameter also summarised in Table 3.3. The bivariate histogram shows a clear positive correlation between parameters α and μ. This is perhaps as expected since the rates α and μ relate to opposite events in the within-host model. The rate α corresponds to phagocytosis and survival of bacteria, thereby facilitating the disease, whilst μ corresponds to the prevention of disease through extracellular bacterial death. Therefore, increasing the rate at which bacteria are taken up by susceptible phagocytes increases the number of bacteria being released by rupturing phagocytes, but this can be counteracted by increasing the rate at which the bacteria are killed. This suggests that it is possible to increase, or similarly decrease α and μ but still attain a similar total number of extracellular bacteria. Further evidence for this correlation is seen in the elasticities computed in Section 3.2.1, with the elasticities with respect to α equal to the negative of those with respect to μ.

Despite this, it is clear from the histogram in Figure 3.10 that the accepted values are not distributed homogeneously around the elliptic shape, instead being focused more towards the centre. Therefore the posterior median estimates for α and μ have a larger posterior probability than that of the estimated values obtained by Wood *et al.* (2014). Figure 3.11 shows predictions of the within-host model using posterior median estimates of α and μ, along with the data by Eigelsbach *et al.* (1962); White *et al.* (1964) used to obtain these estimates. Similar to the results by Wood *et al.* (2014), the within-host model is better at predicting the data by White *et al.* (1964). This may be because the extracellular bacteria counts from this data set are larger, and therefore when minimising the distance between the model and data during the ABC rejection sampling algorithm, there is a bias towards minimising the distance between these points.
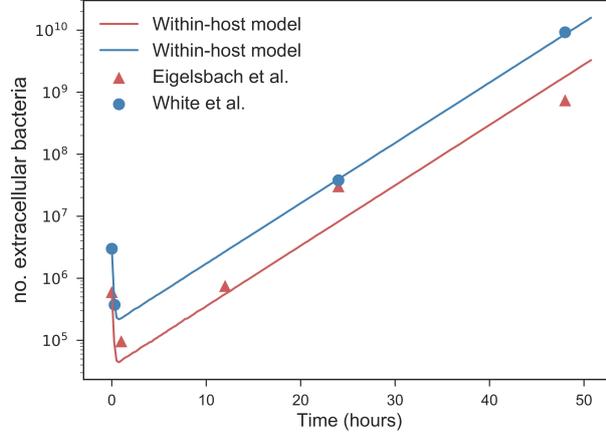
Figure 3.11: A comparison between within-host model predictions (curves) obtained as mean values throughout time from tau-leaping simulations for different initial bacterial loads (blue and red) corresponding to the initial values measured by Eigelsbach *et al.* (1962); White *et al.* (1964), and the data points by Eigelsbach *et al.* (1962); White *et al.* (1964). Posterior median values of α and μ have been used to obtain the predictions.

However, the data by Eigelsbach *et al.* (1962) suggests a slowing in the growth of the bacterial population, a feature that would be difficult to capture in the current description of the within-host model.

A full list of parameter values, along with their interpretation, is reported for both the within-phagocyte and within-host models in Table 3.4.

3.3.3 Population-level model

Four scenarios, are considered for the release of bacteria inside a microbiology facility, $A1$, $A2$, $C1$ and $C2$. These four scenarios can be considered as two separate ventilation regimes, regime A and C, both of which correspond to regimes considered by Noakes & Sleigh (2009) and López-García *et al.* (2019). Regardless of the source zone, it is assumed that 10^5 bacterial counts of *F. tularensis* are released at time $t = 0$. In all scenarios the zone sizes are equal, $V_i = 36m^3$ for $i \in \{1, 2, 4, 5\}$ and $V_i = 12m^3$ for $i \in \{3, 6\}$, and the pulmonary rate is set to $\rho = 0.01m^3/\text{minute}$

Parameter	Event	Parameter value
λ	per intracellular bacterium growth rate	$0.2164\,h^{-1}$
C	intracellular carrying capacity	393 bacteria
μ	extracellular bacterial death	$8.63\,h^{-1}$
α	phagocytosis with bacterial survival	$0.3325\,h^{-1}$
M	threshold for symptoms onset	$M \sim logN(26.2, 6.05)$
δ	phagocyte rupture rate	$0.0241\,h^{-1}$
R_k	probability of rupture with k bacteria	(see Figure 3.14)

Table 3.4: A description of the parameters for the within-phagocyte and within-host models.

(Noakes & Sleigh (2009)). The remaining parameters differ between regimes and are provided in Table 3.5. Regime A corresponds to equal rates of air extraction for each zone of the facility, and hence the rate of airflow between any pair of adjacent zones is equal. For Regime C, air is only extracted from zones 1 and 4, and therefore the airflow is greater in the direction of these two zones. Steady state values for the total number of bacteria an individual in each zone inhales are also provided in Table 3.5, $\boldsymbol{p}^{(v)} = \lim_{t \to +\infty} \left(p_1^{(v)}(t), p_2^{(v)}(t), p_4^{(v)}(t), p_5^{(v)}(t) \right)$, $v \in \{A1, A2, C1, C2\}$. These steady states then act as the initial doses used as inputs in the within-host model. A graphical representation of scenarios $A1$, $A2$, $C1$, and $C2$ is given in Figure 3.12 and the time courses of the variables $C_i(t)$, $1 \le i \le 6$, and $p_j(t)$, $j \in \{1, 2, 4, 5\}$ are plotted for scenario A1 in Figure 3.13 for illustrative purposes. The python function `scipy.integrate.odeint` is used to solve the population-level model and obtain steady state values for the amount of bacteria inhaled by each individual.

3.4 Results

In this section, I present results for each level of the multi-scale model. For the within-phagocyte model, the rupture distribution is computed using the analyt-

Scenario	$\beta_{i,j}$ (m^3/min)	Q_i (m^3/min)	Source Zone	Steady State
A1	$\beta_{12} = \beta_{23} = \beta_{36} = \beta_{63}$ $= \beta_{56} = \beta_{45} = \beta_{21}$ $= \beta_{32} = \beta_{65} = \beta_{54} = 9$	$Q_i = 3, i = 1,..,6$	1	$\boldsymbol{p}^{(A1)} = (145, 82, 13, 17)$
A2	$\beta_{12} = \beta_{23} = \beta_{36} = \beta_{63}$ $= \beta_{56} = \beta_{45} = \beta_{21}$ $= \beta_{32} = \beta_{65} = \beta_{54} = 9$	$Q_i = 3, i = 1,..,6$	5	$\boldsymbol{p}^{(A2)} = (17, 23, 82, 110)$
C1	$\beta_{12} = \beta_{23} = \beta_{36}$ $= \beta_{63} = \beta_{56} = \beta_{45} = 9$ $= \beta_{21} = \beta_{32}$ $= \beta_{65} = \beta_{54} = 18$	$Q_1 = Q_4 = 9$ $Q_2 = Q_3$ $= Q_5 = Q_6 = 0$	1	$\boldsymbol{p}^{(C1)} = (102, 46, 9, 9)$
C2	$\beta_{12} = \beta_{23} = \beta_{36}$ $= \beta_{63} = \beta_{56} = \beta_{45} = 9$ $= \beta_{21} = \beta_{32}$ $= \beta_{65} = \beta_{54} = 18$	$Q_1 = Q_4 = 9$ $Q_2 = Q_3$ $= Q_5 = Q_6 = 0$	5	$\boldsymbol{p}^{(C2)} = (18, 18, 92, 92)$

Table 3.5: Airflow parameters for the four scenarios considered, and the steady state bacteria intake values representing the initial dose for individuals in each zone. Parameters have been chosen according to ventilation regimes A and C considered by López-García *et al.* (2019); Noakes & Sleigh (2009).

ical methods presented in Section 3.1.1. This distribution is then applied to the within-host model to study the relationship between the initial dose, the probability of response, and the mean time until response. Finally, the within-host model is used to obtain the probability of response for each individual in the microbiology facility. From this, the distribution of the number of individuals who respond can be found, thereby indicating the optimum scenario for minimising the number of responses.

3.4.1 Within-phagocyte model

The distribution of the number of bacteria released on rupture of an infected phagocyte is provided in Figure 3.14 for posterior median values of λ and C. In order to compare with results by Wood *et al.* (2014), recall that the approach they use to compute the number of bacteria released on rupture involves
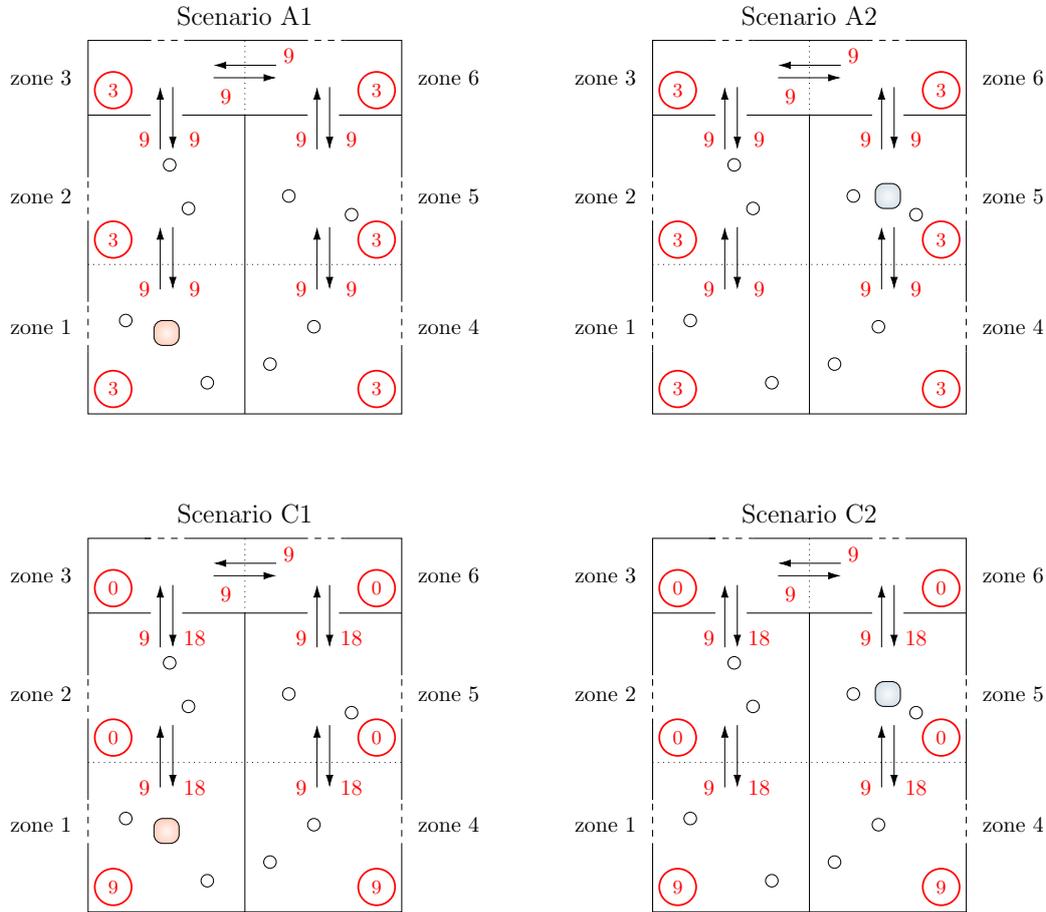
Figure 3.12: Four scenarios $A1$, $A2$, $C1$ and $C2$ corresponding to two potential release locations (zone 1 for scenarios $A1$ and $C1$ and zone 5 for scenarios $A2$ and $C2$). The ventilation regime in scenarios $A1$ and $A2$ represents a well-mixed ventilation, where airflow is balanced across zones and there are equal levels of air extraction (circled values). The regime in $C1$ and $C2$ represents airflow occurring from the corridor areas to the opposed sides of the rooms where extraction of air is in place.

evaluating a deterministic logistic growth process at the median rupture time, $Median[T^{rupture}]$. The method here may instead be interpreted as evaluating an analogous stochastic logistic growth process that incorporates the actual log-
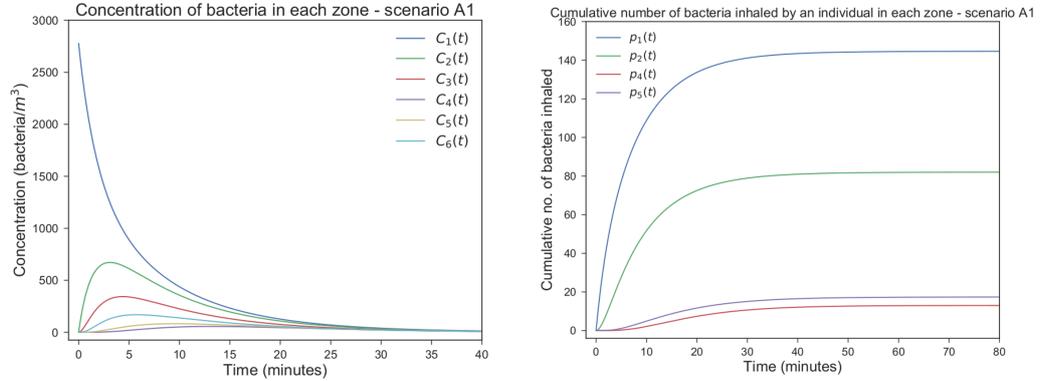
Figure 3.13: Time course of the variables $C_i(t)$, $1 \leq i \leq 6$ for the concentration of *F. tularensis* bacteria in each zone of the population-level model, and $p_j(t)$, $j \in \{1, 2, 4, 5\}$ for the cumulative number of bacteria inhaled by each individual in each populated zone. Both plots are created using parameters values specified in scenario $A1$.

normally distributed rupture time. Since the deterministic and stochastic processes have both been parametrised using the same data set, they are comparable, and the median number of bacteria released according to the rupture distribution is approximately equal to the fixed value of 358 bacteria obtained by Wood *et al.* (2014). This suggests that the median value had previously been calculated correctly. Despite this, the method outlined here is far more general and allows for a more comprehensive analysis of the number of bacteria released from rupture events. For example, from the rupture distribution it is possible to calculate that the mean number of bacteria released is 288, significantly fewer than the median. This observation is directly related to the bimodal shape of the distribution, suggesting that phagocytes are likely to either release a few bacteria or a few hundred bacteria on rupture. However, both of these features would go unnoticed when applying the approach used by Wood *et al.* (2014). Although the rupture distribution has not been experimentally measured and the distribution reported here cannot therefore be verified, there is reason to believe that the bimodal shape

is appropriate. Recent single-cell analysis by Brock & Parmely (2017) showed that a significant number of infected phagocytes release very few bacteria when rupturing, suggesting that the first peak in the rupture distribution is not an artifact of the model but that phagocytes rupturing soon would not have enough time for substantial bacterial proliferation. This limited proliferation then results in small rupture sizes, as predicted by the first peak in the rupture distribution given in Figure 3.14.

One of the assumptions made in developing the within-phagocyte model is that bacteria are only phagocytosed by uninfected phagocytes, that is, once a phagocyte has taken up a bacterium, it will not do so again. As bacterial counts can increase rapidly during infection, the validity of this assumption could be questioned. Because of this, the effect that multiple phagocytosis events has on the rupture distribution has been investigated. A numerical approach was used to explore this by using a Gillespie algorithm to simulate the within-phagocyte model. The times at which additional phagocytosis events would occur were sampled according to a $U(0, \mathbb{E}[T^{rupture}])$ distribution and at each of these times, if the phagocyte had not ruptured, the intracellular bacterial load was increased by one. Figure 3.15 shows the rupture distribution for different numbers of additional phagocytosis events. As the number of additional phagocytosis events increases, the peak around zero becomes less pronounced and the rupture distribution is more uni-modal. However, this does not agree with the recent findings of Brock & Parmely (2017) that suggest that some phagocytes release only a few bacteria on rupture. Therefore, if additional phagocytosis events are to occur, it is thought to only be a small number, in which case there is little impact on the rupture distribution compared to having no additional phagocytosis events. It should also be noted that for the rupture distributions shown in Figure 3.15, the larger peak always remains centred around the same value. Although additional phagocytosis events will lead to faster growth of the intracellular population, the carrying capacity on the size of the population remains the same. Since the position of this peak is determined by the carrying capacity, the number of additional bacteria that enter the phagocyte will have no effect.
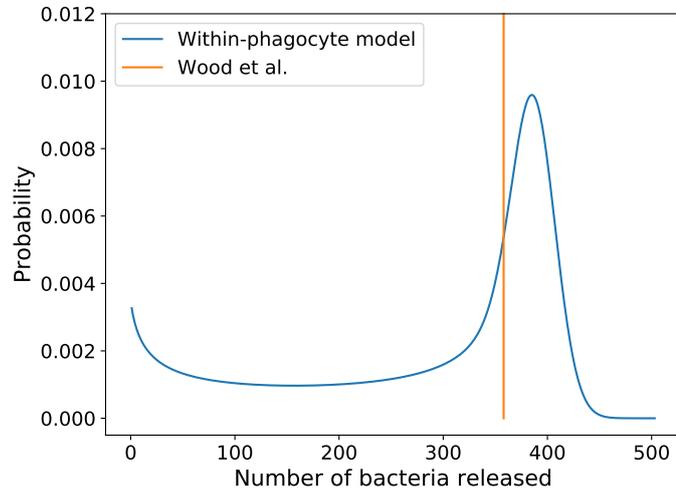
Figure 3.14: The distribution of the number of bacteria released by an infected phagocyte on rupture, as predicted by the within-phagocyte model, compared to the fixed value assumed by Wood *et al.* (2014). Posterior median values of λ and C have been used to compute this distribution.

3.4.2 Within-host model

By using the rupture distribution in Figure 3.14, and the within-host model as defined in Section 3.1.2, the probability of response and the mean time until response can be computed for varying initial doses. Figure 3.16 shows a comparison between the probability of response computed using the multi-scale model and the within-host model by Wood *et al.* (2014). Both models give similar predictions, with a dose of approximately 100 bacteria or greater always resulting in the onset of symptoms. A response is also still possible at lower doses, with around 30% of individuals responding given an initial dose of 10 bacteria. This is consistent with the belief that *F. tularensis* is capable of causing tularemia following the inhalation of as few as 10 organisms (Oyston *et al.* (2004)). It is likely that these similarities observed between the predictions of the two models are due to both models being parametrised using the same data sets.
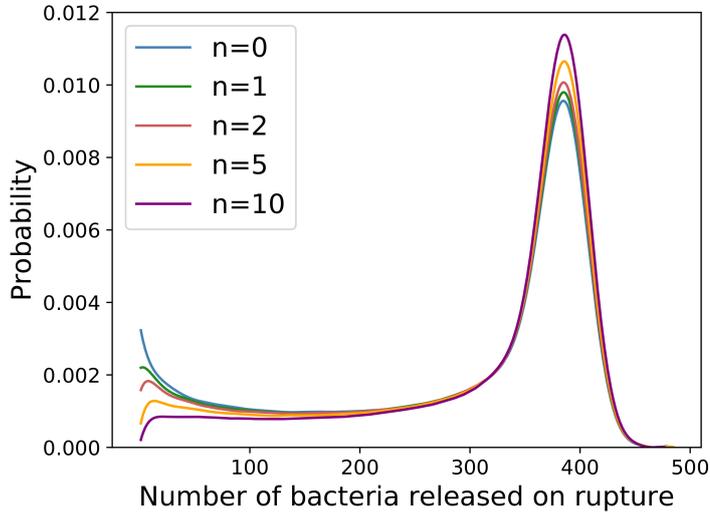
Figure 3.15: The rupture distribution when n additional phagocytosis events are allowed to occur uniformly between $t = 0$ and $t = \mathbb{E}\left[T^{rupture}\right]$. The case where $n = 0$ corresponds to the rupture distribution in Figure 3.14.

The left hand plot of Figure 3.17 shows the cumulative probability of response for different initial doses, that is, the probability that an individual responds prior to time t. The asymptotic values in this plot therefore correspond to the probability that an individual ultimately responds for each initial dose. The advantage of computing the cumulative probability of response is that it gives an idea of the timescale of response. For example, it could be used to quickly estimate the time at which a response would occur in a certain proportion of a population, all of whom receive a similar initial dose.

The right hand plot of Figure 3.17 shows the mean response time, conditioned on symptom onset, for different initial doses, along with equivalent predictions made by Wood *et al.* (2014) and experimental data for the time until symptom onset observed in infected individuals (Saslaw *et al.* (1961); Sawyer *et al.* (1966)). As with the probability of response, the multi-scale model predictions for the mean response time agree as well with the experimental data as the predictions made by Wood *et al.* (2014). However, it is worth noting that the posterior
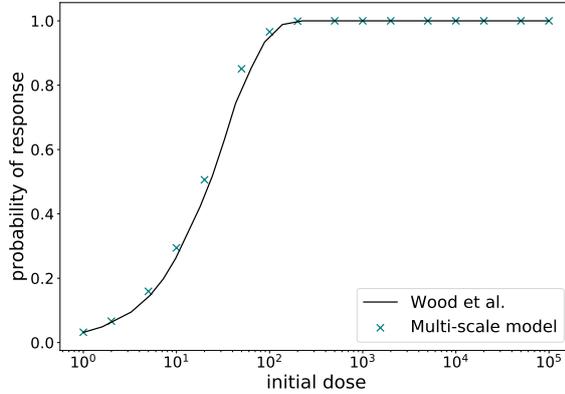
Figure 3.16: Predictions of the probability of response using the multi-scale model for different initial doses.

median estimates of α and μ used here are significantly larger than the values estimated by Wood *et al.* (2014). Furthermore, although the parameters α and μ are highly correlated, and thus determining their true values is difficult, the histogram for the ratio α/μ (Figure 3.10, right) indicates that this ratio ranges from 24.69 to 27.54, lower than the ratio of 31.95 implied by the estimates of Wood *et al.* (2014). Together this suggests that this pair of parameters had previously been underestimated.

3.4.3 Population-level model

To consider the population-level effects that a release of *F. tularensis* bacteria would have on individuals working in a microbiology facility, the probability of response for each individual can first be computed using the within-host model, where the initial dose is given by the steady state values in Table 3.5. Suppose that Z_i is a random variable denoting the number of individuals who develop symptoms in zone i, then Z_i follows a binomial distribution, $Z_i \sim \text{Bin}(n_i, \pi_i)$, where n_i is the number of individuals in zone i and π_i is the probability of response for an individual in zone i. If Z is the total number of individuals in the microbiology facility who respond, then in the scenarios described in Figure 3.12, $Z = Z_1 +$
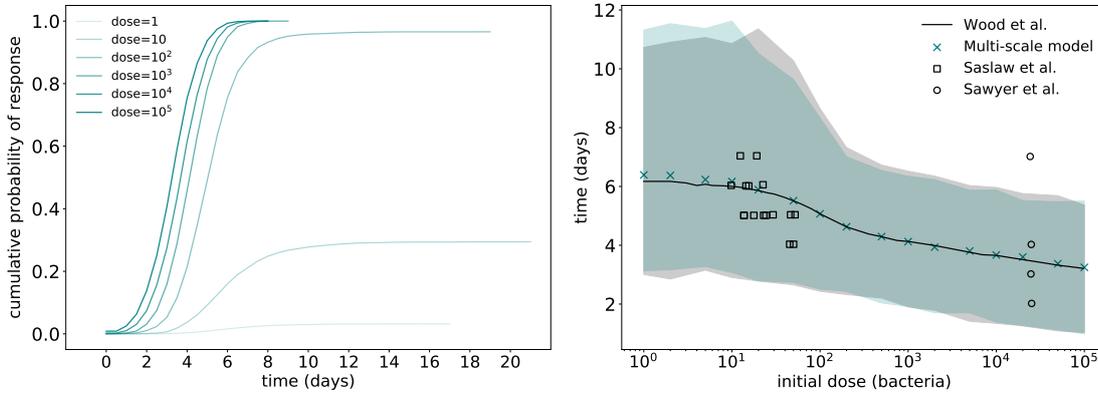
Figure 3.17: Left: Predicted cumulative probability of response up to time t, from the multi-scale model for different initial doses. Right: comparison between the conditioned mean time until response predicted by Wood *et al.* (2014) and by the multi-scale model developed here. Shaded regions represent 95% quantiles.

$Z_2 + Z_4 + Z_5$ is the sum of four independent binomial distributions. Although simple in this case, for scenarios where the facility is divided into many zones, the probability mass function for Z can be efficiently evaluated by successively computing the convolution of two binomial random variables. (Butler & Stephens (2017)).

For each of the four scenarios considered here, $A1$, $A2$, $C1$ and $C2$, the distribution of Z is plotted in Figure 3.18. From this, it can be seen that the optimum scenarios for reducing the number of individuals developing symptoms are scenarios $A1$ and $C1$, corresponding to the initial release of *F. tularensis* bacteria occurring in zone 1. This is perhaps intuitive given the layout of the facility, where zone 1 can be thought of as being at the end of the facility, whilst zone 5, the other zone considered for bacterial release, is located in the middle. Therefore, if the bacteria are released in zone 1, they must travel further in a single direction in order for all individuals in the facility to become exposed. The different ventilation regimes between scenarios $A1$ and $C1$ also have an effect on the distribution of the number of individuals who respond. Scenario $C1$, corresponding to a greater rate of air extraction that only occurs in zones 1 and 4, appears to

74
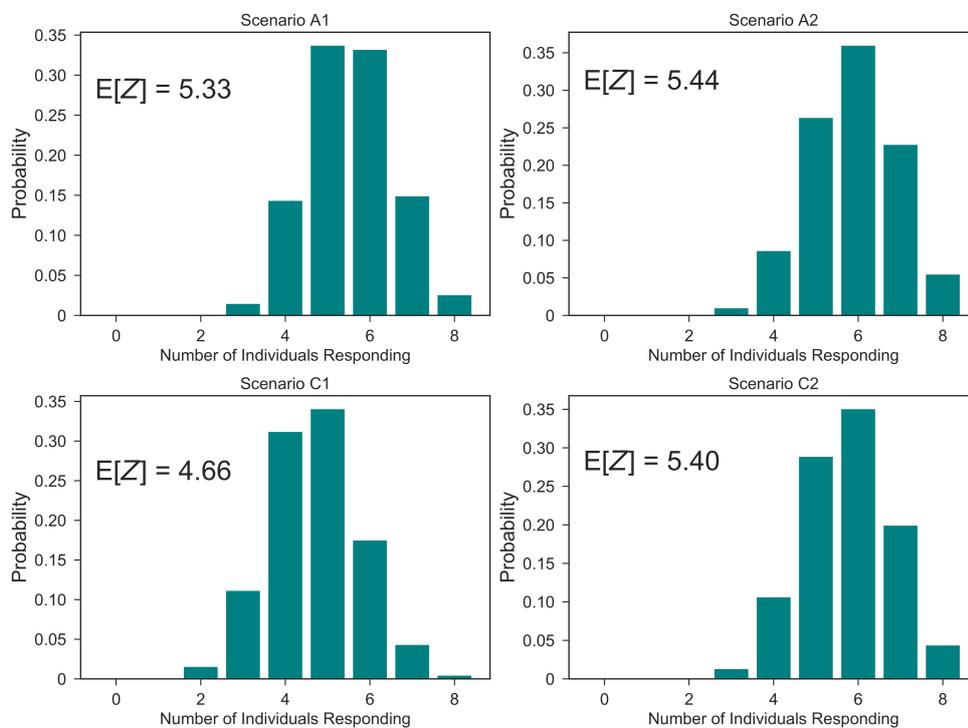
Figure 3.18: Distribution of the number of individuals showing symptoms of infection following bacterial release for scenarios $A1, A2, C1$ and $C2$.

be the preferred option compared to scenario $A1$, where extraction of air occurs equally across all zones. This links back to the source zone of the initial release of bacteria, with a greater volume of air being extracted from zone 1 in scenario $C1$, it is more likely that the bacteria will be removed from the facility before being carried through to the remaining zones. This idea is also observed for when the initial release of bacteria occurs in zone 5, with scenario $C2$ preferred to scenario $A2$.

3.5   Discussion

In this chapter, I have shown how a multi-scale model can be used to link the dynamics of *F. tularensis* infection at a single cell, within-host and population level.

The model can be seen as an extension of that in Wood *et al.* (2014) through the consideration of inter-phagocyte variability by computing the probability mass function for the number of bacteria released on rupture of an infected phagocyte. In addition to this, it has been shown how the experimentally observed log-normally distributed times until rupture can be accounted for by approximating the log-normal distribution using a PH distribution, thus maintaining the Markovian property of the process.

At the within-phagocyte level, the rupture distribution provides insights otherwise unnoticed in the previous analysis by Wood *et al.* (2014), such as the mean number of bacteria released being significantly lower than the median. Furthermore, the bimodal shape of the distribution, indicating that it is possible for phagocytes to release just a few bacteria on rupture, is in line with the experimental results of Brock & Parmely (2017). The within-host model has been parametrised using infection data to provide predictions that agree with dose and time response data, whilst suggesting that the parameters α and μ had previously been underestimated. The population-scale model provides a practical application of the within-phagocyte and within-host models within a laboratory setting, predicting the number of individuals who would develop symptoms following the release of *F. tularensis* bacteria. The practicality of this multi-scale model is also not only restricted to an indoor setting since the within-host model only requires that an initial dose is specified. Therefore the within-phagocyte and within-host models could be used in conjunction with outdoor dispersion models that use a known source of release to predict the amount of bacteria in specific locations across a much larger area than is currently being considered with a microbiology facility.

When developing mathematical models that explicitly account for biological mechanisms, it is sometimes desirable to include many reactions in order for a high level of detail to be achieved. However, more reactions results in a greater number of rate constants that will have to be estimated, and this is not always possible. For example, in the within-host model, the rate μ accounts for multiple mechanisms of extracellular bacterial death, such as by antimicrobial peptides or

phagocytosis by activated phagocytes. Experimental data are not currently available for the contribution to extracellular death from each of these mechanisms and so it would not be possible to obtain reliable estimates for parameters that correspond to them. If more data are made available, a new within-host model could be proposed such that μ is replaced by multiple parameters, each representing the contribution to bacterial death from a different biological mechanism. It would then be interesting to determine how each of these individual mechanisms affect the probability of response and mean response time.

A second limitation of the current multi-scale model is that, in the within-phagocyte model, the rupture time is modelled as a log-normally distributed time that is independent of intracellular bacterial proliferation. It would instead be preferable to in some way relate the rate of rupture, δ_n, to the number of intracellular bacteria. Although current experimental knowledge provides little insight into how these are related, this idea is developed further in Chapter 4, whereby the rate of rupture is assumed to depend linearly on the number of intracellular bacteria. For the multi-scale model at least, the log-normally distributed time estimated by Wood *et al.* (2014) offers a compromise between current knowledge and model complexity.

Chapter 4

An agent based model for *Francisella tularensis* infection

In this chapter, I describe how an agent based model (ABM) can be constructed to represent the early stages of respiratory *F. tularensis* infection, as well as the spreading of infection across multiple organs. Whereas the multi-scale model, developed in Chapter 3, ultimately describes the effects of the release of *F. tularensis* on a population, the ABM incorporates a greater level of detail and is thus used to study infection dynamics for a single infected individual. The ABM is designed to represent what are believed to be the key biological mechanisms involved during *F. tularensis* infection, however, greater detail comes at a cost of increased computational intensity. As a result of this, a system of ODEs is derived in order to offer a fast but accurate approach to obtain the total number of bacteria present in each infected organ.

When describing infections involving intracellular pathogens, the method in which the pathogen leaves an infected cell following replication is often classified as either continuous release or bursting. For viruses such as Ebola virus, studied in Chapter 5, and influenza virus, intracellular replication of the viral genome is followed by the continual release of new virus particles until the death of the infected host cell (Pornillos *et al.* (2002); Rossman & Lamb (2011)). On the other hand, *F. tularensis* and *S. enterica* bacteria are released in a single burst or rup-

ture event, coinciding with the death of the infected cell (Mastroeni *et al.* (2009)). When creating a mathematical model to describe continuous release, it is common to assume that this occurs at a rate proportional to the number of infected cells. Despite describing a very different mechanism, this same approach has also been used by Day *et al.* (2011) and Chen *et al.* (2007) to model the bursting of infected cells during anthrax and SIV infection respectively, where an adjustment of the rate is made to account for the average bacterial load of a cell when it bursts. Although this adjustment incorporates the size of the burst, it does not take into account that the release of pathogen occurs at a single point in time, instead allowing for the release to occur continuously. The number of bacteria released from the bursting events therefore increases immediately and not after a delay, during which replication of the pathogen is occurring within the cell. In this sense there is no distinction between deterministic models of bursting and models of continuous release. This assumption is perhaps valid when considering a large number of cells, each infected at separate times, since the bursting of cells would occur frequently enough that the release of pathogen may be thought of as occurring continuously. However, the interest here is in modelling the initial stages of *F. tularensis* infection, particularly focusing on low initial doses since *F. tularensis* is able to cause a lethal infection in humans with fewer than 10 colony forming units (CFUs), and in mice with just 1 CFU (Oyston *et al.* (2004)). In such a case, a small number of cells initially become infected at approximately the same time and so the discrete times at which rupture events occur must be explicitly accounted for.

A stochastic approach that considers the bursting of infected cells during human immunodeficiency virus (HIV) infection is used by Pearson *et al.* (2011), although only under the assumption that the size of the burst is a Poisson random variable. This has a similar disadvantage to the log-normally distributed rupture times used in the within-phagocyte model in Chapter 3, whereby the distribution is imposed instead of derived. Variable burst sizes are addressed by Brown *et al.* (2006) in a stochastic model of a single *S. enterica* infected cell that is used to develop a system of ODEs for describing the fraction of infected cells containing

n bacteria. This model then allows forms of transition rates to be identified that best characterise the burst size for virulent and attenuated strains.

In the same way in which Brown *et al.* (2006) utilise a model of a single cell to describe a population, the approximations derived here for the ABM use a stochastic intracellular model of a single macrophage to determine the average number of bacteria released from a bursting macrophage as a function of time. Although the result of this is a function that describes the continuous release of bacteria, the discrete nature of the rupture event is taken into consideration by incorporating at each time point the probability that the macrophage ruptures. Furthermore, this is achieved without any assumptions regarding the distribution of the burst size or time until rupture, other than those that are naturally imposed through the choice of rates at the single cell level. The average number of bacteria released can then be used to describe the progression of infection in a population of cells throughout multiple organs. This therefore provides a second approach of linking intracellular and within-host dynamics, alongside the multi-scale approach already described in Sections 3.1.1 and 3.1.2.

In order to develop the intracellular model of *F. tularensis* infection, this chapter begins with an overview of linear birth-and-death processes with catastrophe, building on results for birth-and-death processes given in Section 2.2.7. A description of the ABM follows in Section 4.2, along with the derivation of the system of ODEs used to approximate the total bacterial load of different infected organs in Section 4.3. Since the ABM is computationally demanding to simulate, Section 4.4 details how parameter inference can be performed using *in vivo* infection data alongside the approximations in order to determine parameter values.

4.1 Birth-and-death processes with catastrophe

The idea of a birth-and-death process that accounts for catastrophe events was introduced in Section 3.1.1 for the within-phagocyte model in order to describe the intracellular life cycle of *F. tularensis* following its escape from the phago-some. One downside of this previously used approach is that the log-normally

distributed time until phagocyte rupture is not derived, but rather assumed. A preferred approach would have been to instead construct the birth, death and rupture rates using existing knowledge of intracellular mechanisms, such that the time until rupture follows the desired log-normal distribution. In this section, a step is taken in this direction by deriving the distribution for the time until rupture under the assumption that the birth, death and rupture rates are each proportional to the number of intracellular bacteria.

For the within-phagocyte model described in Section 3.1.1, the analogy of a 'rupture clock' is used to describe how a rupture event and the birth-and-death process act independently of each other. When the rupture clock 'rings', a rupture event immediately occurs, regardless of the number of intracellular bacteria. However, when the rate of rupture depends on the number of intracellular bacteria, the rupture clock is integrated into the birth-and-death process. In this case the distribution of the time until rupture is no longer as trivial. From herein, the focus is on deriving key results regarding linear birth-and-death processes with catastrophe, and how these link to equivalent expressions for linear birth-and-death processes outlined in Section 2.2.7. Two scenarios are considered here, firstly when the intracellular death rate of bacteria is zero and the process reduces to a birth process with catastrophe, and also when the death rate is non-zero. Since a birth process with catastrophe is equivalent to the intracellular dynamics being described by the ABM, these results will then provide a basis for the analytic approximations that follow in Section 4.3.

4.1.1 Linear birth process with catastrophe

Let $\mathcal{X} = \{X(t) : t \geq 0\}$ be a Markov process defined by the one step transition probabilities

$$
p_{i,j}(\Delta t) = \begin{cases} \beta i \Delta t + o(\Delta t) & j = i + 1 \,, \\ \delta i \Delta t + o(\Delta t) & j = B \,, \\ 1 - (\beta + \delta) i \Delta t + o(\Delta t) & j = i \,, \\ o(\Delta t) & \text{otherwise} \,, \end{cases}
$$

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION
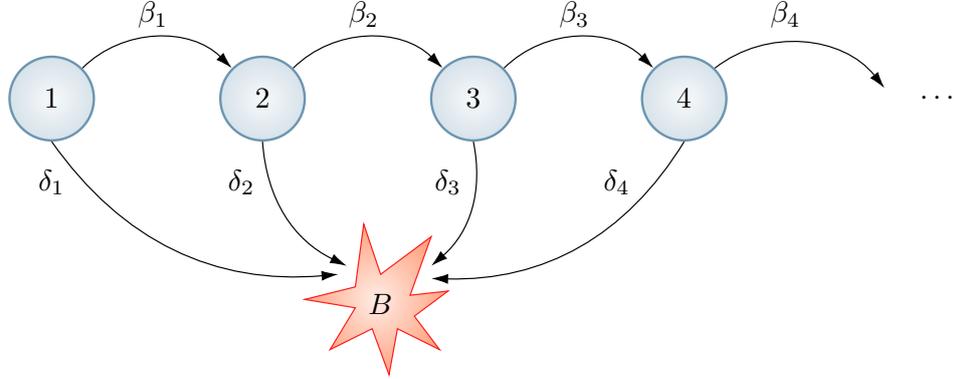
Figure 4.1: A depiction of a birth process with catastrophe, with the catastrophe event represented by transitions into state B.

with state space $\mathcal{S}_{\mathcal{X}} = \{B\} \cup \mathbb{N}$, such that a transition into state B represents the occurrence of a catastrophe event. The process $\mathcal{X}$ may be referred to as a linear birth process with catastrophe, a depiction of which is provided in Figure 4.1. Suppose that $p_n^{(k)}(t) = \Pr(X(t) = n \mid X(0) = k)$, the forward Kolmogorov equations are given by

$$\frac{d}{dt}p_n^{(k)}(t) = \beta(n-1)p_{n-1}^{(k)}(t) - (\beta + \delta)np_n^{(k)}(t) \qquad n \geq k, \ k \geq 1.$$

Since the equation for $n = k$ depends only on $p_k^{(k)}(t)$, it is possible to solve this system of equations recursively to obtain the transition probabilities

$$p_n^{(k)}(t) = \binom{n-1}{k-1}e^{-k(\beta+\delta)t}\left(\left(\frac{\beta}{\beta+\delta}\right)\left(1 - e^{-(\beta+\delta)t}\right)\right)^{n-k}, \qquad n \geq k. \quad (4.1)$$

For the approximations of the ABM, it will be important to determine the time at which an infected macrophage ruptures, that is, the time at which a catastrophe event occurs. Let $S^{(k)}(t)$, also referred to here as the *survival function*, be the probability that a catastrophe event has not occurred before time t. In the context of *F. tularensis* infection, this represents the probability that a macrophage has not ruptured before time t. Since this can be thought of as the probability that a macrophage contains at least one bacterium, it is equivalent to the probability

that X resides in any transient state of $\mathcal{S}_X$ at time t, that is,

$$S^{(k)}(t) = 1 - p_B^{(k)}(t) = \sum_{n=k}^{+\infty} p_n^{(k)}(t) = \left(\frac{\beta + \delta}{\beta + \delta e^{(\beta+\delta)t}} \right)^k . \qquad (4.2)$$

Interestingly, this indicates that the survival probability of a macrophage that is initially infected with k bacteria may be thought of as the product of the survival probabilities for a separate population of k macrophages, each infected with a single bacterium. Since the birth and rupture rates are all linear, the k macrophages are independent prior to any rupture events occurring. However, at the very instance that a single rupture event occurs, the k macrophages are no longer independent, with this resulting in the rupture of all macrophages. Therefore, since a single rupture event is sufficient to cause all k macrophages to rupture, a single macrophage containing k bacteria only survives if all k macrophages in the separate population survive. More formally, consider the case $k = 2$ where a macrophage initially infected with two bacteria may be thought of as two macrophages each infected by one bacterium. These macrophages each have probability $S(t) \equiv S^{(1)}(t)$ of surviving to time t. The probability that a rupture event occurs between t and $t + \Delta t$ may then be thought of as the probability that one macrophage ruptures whilst the second remains alive. Since there are two ways of choosing the macrophage that ruptures, this probability is given by

$$-\frac{dS^{(2)}(t)}{dt} = -2\frac{dS(t)}{dt} S(t) ,$$

indicating that $S^{(2)}(t) = (S(t))^2$. By this reasoning, an alternative approach can be described for finding the survival function which involves formulating an ODE for $S^{(k)}(t)$.

Suppose that $X(0) = k$, the one step transition probabilities indicate that either

- $X(\Delta t) = k$ with probability $1 - (\beta + \delta)k\Delta t$,

- $X(\Delta t) = k + 1$ with probability $\beta k \Delta t$,

- $X(\Delta t) = B$ with probability $\delta k \Delta t$.

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

Since the probability that a macrophage, initially infected with k bacteria, survives from t to $t + \Delta t$ is $S^{(k)}(t)$,

$$S^{(k)}(t + \Delta t) = S^{(k+1)}(t)\beta k \Delta t + S^{(k)}(t)(1 - (\beta + \delta)k\Delta t) \,,$$

and thus in the limit as $\Delta t \to 0$

$$\frac{d}{dt}S^{(k)}(t) = \beta k S^{(k+1)}(t) - (\beta + \delta)k S^{(k)}(t) \,. \tag{4.3}$$

Given that $S^{(k)}(t) = (S^{(1)}(t))^k$, it is sufficient to solve (4.3) for $k = 1$, which can be shown to yield (4.2).

With an expression for the survival function, it is interesting to note that the form of the transition probabilities provided in (4.1) closely resembles that of a negative binomial distribution. Furthermore, in Section 2.2.7, it was described how the transition probabilities for a linear birth process also follow a negative binomial distribution. The similarity between a birth process and a birth process with catastrophe is not surprising since the two are equivalent prior to a catastrophe event occurring. By conditioning on a catastrophe event not occurring before time t, the transition probabilities can be written in a more attractive form that clearly depicts this relationship. This idea is also used by Gani & Swift (2007) to obtain the probability generating function (p.g.f.) for a simple death process with catastrophe where the catastrophe rate is constant. In their case, because the catastrophe is independent of the decay of the population, the p.g.f. for the death process can be used directly in the formulation of the p.g.f. for the process with catastrophe. This is not the case here, where the transition probabilities may be written in the form

$$p_B^{(k)}(t) = 1 - S^{(k)}(t) \,,$$
$$p_n^{(k)}(t) = S^{(k)}(t)\binom{n-1}{k-1}(1 - g(t))^k (g(t))^{n-k} \qquad n \geq k \,, \tag{4.4}$$

where

$$g(t) = \left(\frac{\beta}{\beta + \delta}\right)\left(1 - e^{-(\beta+\delta)t}\right) \,.$$

When comparing $p_n^{(k)}(t)$ here with the expression given in Section 2.2.7, they have the same form but with $1 - g(t)$ in place of $e^{-\beta t}$, with this difference due to the

probability of catastrophe depending on the current state of $\mathfrak{X}$. An advantage of writing the transition probabilities in this form, is that it is simple to obtain results regarding $\mathfrak{X}$ conditioned on a catastrophe event not occurring before time t, since standard results for birth processes can be applied. Although informative, these conditioned quantities do not describe the full behaviour of $\mathfrak{X}$. For a more complete description, let $G^{(k)}(z;t)$ be the p.g.f. of a birth process with catastrophe, by utilising the p.g.f. for a negative binomial distribution

$$G^{(k)}(z;t) = p_B^{(k)}(t) + \sum_{n=1}^{+\infty} z^n p_n^{(k)}(t) = 1 - S^{(k)}(t) + S^{(k)}(t) \left(\frac{z(1 - g(t))}{1 - zg(t)} \right)^k . \quad (4.5)$$

The expected value of $\mathfrak{X}$ is therefore given by

$$\mathbb{E}\left[X(t) \mid X(0) = k \right] = \left. \frac{\partial G^{(k)}(z;t)}{\partial z} \right|_{z=1} = \frac{k S^{(k)}(t)}{1 - g(t)} = \frac{f^{(k)}(t)}{\delta}, \quad (4.6)$$

where $f^{(k)}(t)$ is the density of the time until catastrophe provided that $X(0) = k$, given for the case $k = 1$ by

$$f(t) = \frac{\delta(\beta + \delta)^2 e^{(\beta+\delta)t}}{\left(\beta + \delta e^{(\beta+\delta)t} \right)^2} . \quad (4.7)$$

This relationship between the expected value and the density can also be found by considering a population of N independent macrophages, each initially infected by the same number of bacteria. Suppose that $X^{(i)}(t)$ denotes the number of bacteria inside macrophage i at time t. Given that $\delta X^{(i)}(t)\Delta t$ is the probability that macrophage i ruptures in the short interval $(t, t+\Delta t)$, the number of macrophages that are lost on average during this interval is

$$N \left(S^{(k)}(t + \Delta t) - S^{(k)}(t) \right) = -\sum_{i=1}^{N} \delta X^{(i)}(t)\Delta t, \quad (4.8)$$

and in the limit as $\Delta t \to 0$,

$$\frac{d}{dt} S^{(k)}(t) = -\delta \left(\frac{1}{N} \sum_{i=1}^{N} X^{(i)}(t) \right) = -\delta \mathbb{E}\left[X(t) \mid X(0) = k \right] .$$

When considering the progression of infection within a population of cells, it is necessary to determine the average number of bacteria released from rupture

events in order to study beyond the point at which the initial set of infected
cells rupture. The rupture distribution obtained in Chapter 3 provides a time
independent description of the number of bacteria released when an infected
phagocyte ruptures. The approach described in this chapter instead considers
the average number of bacteria released from an infected cell as a function of
time that also takes into account how likely it is for the cell to rupture.

For a birth process with catastrophe, the probability that a rupture event
occurs between times t and $t + \Delta t$ is $\delta X(t)$. Since $X(t)$ is also the state from
which the rupture state, B, is entered, this will then correspond to the number of
bacteria being released. Suppose that $\tilde{f}(t)$ denotes the mean number of bacteria
released by a macrophage at time t

$$\tilde{f}(t) = \delta \sum_{n=1}^{+\infty} n^2 p_n^{(k)}(t) = \delta \mathbb{E}\left[(X(t))^2 \mid X(0) = k\right].$$

The p.g.f. defined in (4.5) can be used to show that the second moment satisfies

$$\mathbb{E}\left[(X(t))^2 \mid X(0) = k\right] = \frac{f^{(k)}(t)}{\delta}\frac{1+g(t)}{1-g(t)} + k(k-1)\left(\frac{f(t)}{\delta}\right)^2 S^{(k-2)}(t), \quad (4.9)$$

and thus in the case where $k = 1$,

$$\tilde{f}(t) = f(t)\bar{n}(t) = f(t)\frac{1+g(t)}{1-g(t)}.$$

Therefore, whilst $\tilde{f}(t)$ incorporates the probability that a macrophage ruptures
at time t, $\bar{n}(t)$ may be thought of as the average number of bacteria that a
macrophage releases when rupturing at time t. Suppose that R is a random
variable representing the number of bacteria released from a single rupture event,

$$\mathbb{E}\left[R \mid X(0) = 1\right] = \int_0^{+\infty} \tilde{f}(t)\, dt = \frac{\beta + \delta}{\delta}. \qquad (4.10)$$

This result has also previously been obtained by Gillard *et al.* (2014), who show
that the rupture size of a birth process with catastrophe is geometrically dis-
tributed.

4.1.2   Linear birth-and-death process with catastrophe

Let $\mathcal{Y} = \{Y(t) : t \geq 0\}$ be a linear birth-and-death process with catastrophe, with state space $\mathcal{S}_\mathcal{Y} = \{0\} \cup \{B\} \cup \mathbb{N}$ and one-step transition probabilities

$$
p_{i,j}(\Delta t) = \begin{cases}
\beta i \Delta t + o(\Delta t) & j = i + 1\,, \\
\mu i \Delta t + o(\Delta t) & j = i - 1\,, \\
\delta i \Delta t + o(\Delta t) & j = B\,, \\
1 - (\beta + \mu + \delta) i \Delta t + o(\Delta t) & j = i\,, \\
o(\Delta t) & \text{otherwise}\,.
\end{cases}
$$

Since death events are now possible, this leads to the inclusion of a second absorbing state, 0, representing in the case of *F. tularensis* infection, the elimination of bacteria from the infected macrophage. This particular process has previously been studied by Karlin & Tavaré (1982), who scale the instantaneous transition rates in order to separately study the cases where the process is ultimately absorbed into state 0 and state B. In doing so, the p.g.f. and distribution of states from which B is entered are reported, the latter having an equivalent interpretation to the rupture distribution defined in Section 3.1.1. An alternative approach to study this process is described here that follows the same reasoning as previously used for the birth process with catastrophe. Let $S^{(k)}(t)$ again denote the survival function, then since all rates remain linear, a macrophage containing k bacteria can be thought of as k independent macrophages each containing a single bacterium. From the one-step transition probabilities, the following ODE can be constructed for $S^{(k)}(t)$

$$
\frac{d}{dt}S^{(k)}(t) = \beta k S^{(k+1)}(t) + \mu k S^{(k-1)}(t) - (\beta + \mu + \delta) k S^{(k)}(t)\,, \quad S^{(k)}(0) = 1\,. \quad (4.11)
$$

Since the survival function denotes the probability that a catastrophe event has not occurred before time t, it therefore includes the possibility that $\mathcal{Y}$ has instead been absorbed into state 0. This is consistent with *F. tularensis* infection, since the infected macrophage survives if it is capable of eliminating all intracellular
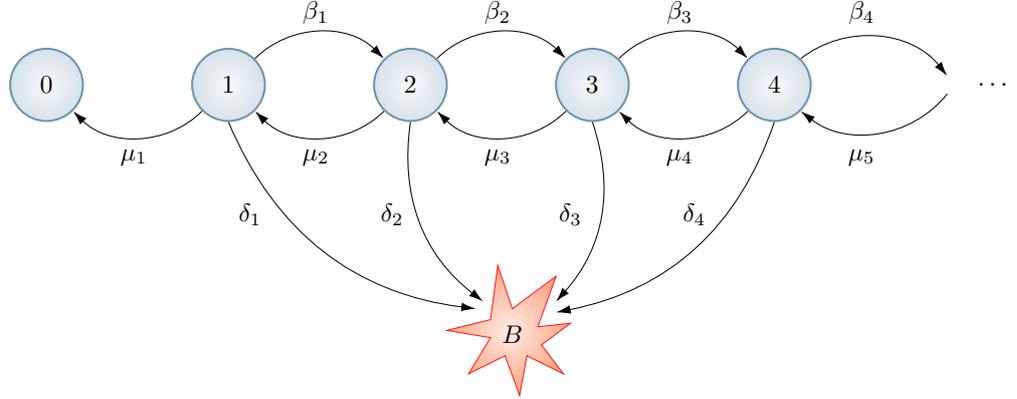
Figure 4.2: A depiction of a birth-and-death process with catastrophe. In addition to the catastrophe state, a second absorbing state at 0 is now included.

bacteria. Because of this, $S^{(0)}(t) = 1$ and the solution to (4.11) is

$$S^{(k)}(t) = \left[\frac{a(1-b) - b(1-a)e^{-\beta(b-a)t}}{(1-b) - (1-a)e^{-\beta(b-a)t}} \right]^k ,$$

where a and b are the roots of the polynomial $h(S) = \beta S^2 - (\beta + \mu + \delta)S + \mu$. For consistency, let a be defined as the smaller of the two roots.

In the limit as t tends to infinity, this survival function tends to a, and as identified by Karlin & Tavaré (1982), a therefore represents the probability that $\mathcal{Y}$ is ultimately absorbed into state 0, instead of state B. The density of the time until catastrophe is then given by

$$f(t) = \frac{\delta(b-a)^2 e^{-\beta(b-a)t}}{(b-1 + (1-a)e^{-\beta(b-a)t})^2} ,$$

and the expectation of $\mathcal{Y}$ is again related to this density through (4.6).

To determine the transition probabilities, the same approach used for the birth process with catastrophe can be applied, that is, to condition the process on a catastrophe event not occurring before time t. This differs from the approach used by Karlin & Tavaré (1982), who choose to condition the process on whether or not a catastrophe event ultimately occurs. Provided that a catastrophe event has not occurred before time t, the process $\mathcal{Y}$ takes the form of a birth-and-death

process. Assuming, for now, that $Y(0) = 1$, Renshaw (2011) suggests that an appropriate form for the transition probabilities of a birth-and-death process is

$$p_n(t) \equiv p_n^{(1)}(t) = [1 - \alpha(t)][1 - \gamma(t)][\gamma(t)]^{n-1} \qquad n \geq 1.$$

To obtain the transition probabilities, the unknown functions $\alpha(t)$ and $\gamma(t)$ must be determined. Knowing that this trial solution must satisfy the forward Kolmogorov equations

$$\frac{d}{dt}p_n(t) = \beta(n-1)p_{n-1}(t) + \mu(n+1)p_{n+1}(t) - (\beta + \mu + \delta)np_n(t) \qquad n \geq 1,$$

the following ODE for $\gamma(t)$ can be formulated,

$$\frac{d}{dt}\gamma(t) = \mu\gamma(t)^2 - (\beta + \mu + \delta)\gamma(t) + \beta. \tag{4.12}$$

To ensure that $p_1(0) = 1$ and $p_n(0) = 0$ for all $n \neq 1$, the initial condition $\gamma(0) = 0$ is imposed. Using this to solve (4.12) yields the solution

$$\gamma(t) = \frac{1 - e^{-\beta(b-a)t}}{b - ae^{-\beta(b-a)t}}.$$

To obtain $\alpha(t)$, an ODE can be found in a similar way to that for $\gamma(t)$, however, a simpler approach is to consider the probability that the process Y remains in a transient state at time t, that is

$$\sum_{n=1}^{+\infty} p_n(t) = [1 - \alpha(t)][1 - \gamma(t)] \sum_{n=1}^{+\infty} [\gamma(t)]^{n-1} = 1 - \alpha(t).$$

Therefore, $\alpha(t)$ represents the probability that Y has been absorbed into either absorbing state at time t. The probability that Y has been absorbed into B at time t is $1 - S(t)$, whilst the probability that Y has been absorbed into state 0 at time t is $\mu\gamma(t)/\beta$. An appropriate form for $\alpha(t)$ is therefore given by

$$\alpha(t) = 1 - S(t) + \frac{\mu}{\beta}\gamma(t).$$

The general form of the transition probabilities for a birth-and-death process may now be used to show that the transition probabilities for a birth-and-death

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

process with catastrophe are

$$p_B(t) = 1 - S^{(k)}(t) \,,$$

$$p_0^{(k)}(t) = S^{(k)}(t)[\alpha(t)]^k \,,$$

$$p_n^{(k)}(t) = S^{(k)}(t)[\alpha(t)]^k[\gamma(t)]^n \sum_{i=0}^{n} \binom{k}{i}\binom{k+n-i-1}{n-i}\left[\frac{1-\alpha(t)-\gamma(t)}{\alpha(t)\gamma(t)}\right]^i \,.$$

Although it is possible to obtain expressions for the transition probabilities using the p.g.f. defined by Karlin & Tavaré (1982), the same resemblance between those of a birth-and-death process would not be so easily achieved.

The forward Kolmogorov equations can also be used to determine the p.g.f. of $\mathcal{Y}$. Through applying the approach described in Section 2.2.7, $G^{(k)}(z;t)$ satisfies the partial differential equation

$$\frac{\partial G}{\partial t} - \left[\beta z^2 - (\beta + \mu + \delta)z + \mu\right]\frac{\partial G}{\partial z} = 0\,, \qquad G(z;0) = z^k \,.$$

The method of characteristics may then be used to obtain the solution

$$G^{(k)}(z;t) = \left[\frac{ab\left(1 - e^{-\beta(b-a)t}\right) + z\left(be^{-\beta(b-a)t} - a\right)}{b - ae^{-\beta(b-a)t} - z\left(1 - e^{-\beta(b-a)t}\right)}\right]^k \,.$$

This p.g.f. can be used to show that the second moment of a birth-and-death process with catastrophe also satisfies (4.9) with $g(t)$ replaced by $\gamma(t)$. The average number of bacteria released due to cell rupture at time t is therefore given by

$$\tilde{f}(t) = f(t)\frac{1 + \gamma(t)}{1 - \gamma(t)} \,.$$

Previously, Karlin & Tavaré (1982) show that when considering paths that ultimately end with catastrophe, the rupture size for a birth-and-death process with catastrophe follows a geometric distribution with mean $b/(b-1)$. This mean does not, however, take into account the scenario when $\mathcal{X}$ is absorbed into state 0 and no bacteria are released. For this, the average rupture size is given by

$$\mathbb{E}\left[R \mid X(0) = 1\right] = \int_0^{+\infty} \tilde{f}(t)\,dt = \frac{\delta b}{\beta(b-1)^2} = (1-a)\frac{b}{b-1} \,.$$

In either case, it is demonstrated how both roots of the polynomial $h(S) = \beta S^2 - (\beta + \mu + \delta)S + \mu$ have a meaningful interpretation, with $0 < a < 1$ previously shown to represent the probability that absorption into state 0 occurs before absorption into state B.

As a comparison of the density and survival functions for a birth process and birth-and-death process with catastrophe, Figure 4.3 depicts $f(t)$ and $S(t)$ for each process. Also included here is a previous effort by Gillard *et al.* (2014) to approximate the survival function of a birth process with catastrophe, who use it to help model the early dynamics of *F. tularensis* infection. This approximation satisfies the ODE

$$\frac{d}{dt}S(t) = -\delta e^{\beta t}S(t)\,, \qquad S(0) = 1\,,$$

where $e^{\beta t}$ is used to approximate the mean number of intracellular bacteria, provided that a rupture event has not occurred. Given the transition probabilities in (4.4), it is known that this conditioned mean is given exactly by $1/(1 - g(t))$ and thus the survival function can be obtained exactly by instead solving

$$\frac{d}{dt}S(t) = -\frac{\delta S(t)}{1 - g(t)}\,, \qquad S(0) = 1\,.$$

This is also evident from the expression for the mean of a birth process with catastrophe given in (4.6). Since $e^{\beta t}$ grows faster than $1/(1 - g(t))$, the mean number of intracellular bacteria is quickly overestimated, therefore resulting in the faster decay of the approximate survival function that can be seen in Figure 4.3.

4.1.3   Populations of cells

The results so far described in this chapter regard single cells, for example, the probability that a single infected cell survives to time t. Now, consider the scenario where a population of cells are infected and let $M(t)$ be a random variable denoting the number of cells that survive to time t. If each cell is infected by k bacteria, then a cell is either alive at time t with probability $S^{(k)}(t)$ or has ruptured with probability $1 - S^{(k)}(t)$. If all cells are independent, $M(t)$ follows

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION
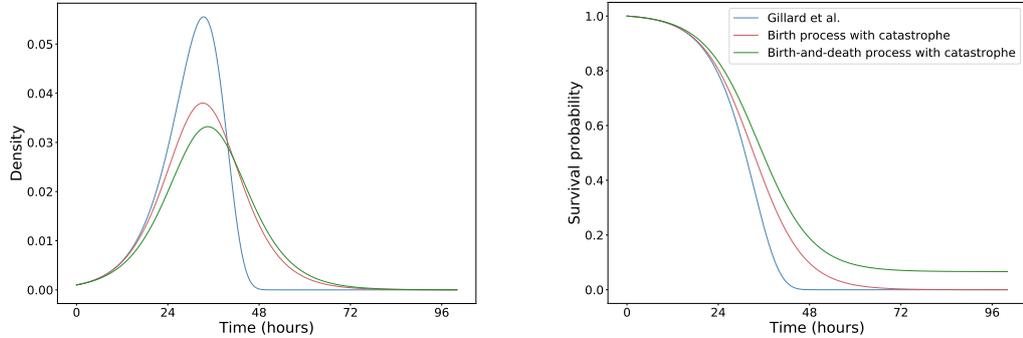
Figure 4.3: A comparison between the density of the time until rupture and survival function determined by Gillard *et al.* (2014) and exact expressions for a birth process and birth-and-death process with catastrophe ($\beta = 0.15\,h^{-1}$, $\mu = 10^{-2}\,h^{-1}$, $\delta = 10^{-3}\,h^{-1}$).

a binomial distribution, $M(t) \sim Bin(M, S^{(k)}(t))$, where M is the initial number of infected cells. When each infected cell is modelled as a birth process with catastrophe, the probability of ultimately rupturing is one and hence both the mean and variance of $M(t)$ tend to zero as t tends to infinity. If each cell is modelled as a birth-and-death process with catastrophe, the average number of cells that successfully clear their bacterial load and survive is $\mathbb{E}\left[M(t)\right] = Ma$, with variance $\mathrm{Var}[M(t)] = Ma(1 - a)$.

In an experimental setting, it is unlikely that each cell is infected by the exact same number of bacteria. Instead, a multiplicity of infection (MOI) is defined beforehand to represent the initial ratio of bacteria and cells. It is often then assumed that the number of bacteria a cell is infected by follows a Poisson distribution with mean equal to the MOI (Dixit & Perelson (2004); Shabram & Aguilar-Cordova (2000)). When modelling each cell, if $X(0)$ is the initial amount of pathogen a cell is infected by and λ is the MOI, this suggests that

$$\Pr(X(0) = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k \geq 0.$$

Suppose now that $\pi(t)$ denotes the probability that a cell survives to time t. In order to take into account the variable number of bacteria that a cell can be

infected by, the law of total probability can be used to show that

$$\pi(t) = \sum_{k=0}^{+\infty} S^{(k)}(t)\Pr(X(0) = k) = \exp\left[-\lambda(1 - S(t))\right], \qquad t \geq 0\,,$$

where the property $S^{(k)}(t) = (S(t))^k$ of the survival function has been applied. As in the previous case, provided that the cells act independently of each other, it is still true that $M(t)$ follows a binomial distribution, $M(t) \sim Bin(M, \pi(t))$. However, for both the birth process and birth-and-death process with catastrophe, there is now a non-zero probability that the cell survives. For a birth process with catastrophe, this is given by the probability that the cell never gets infected, whilst for a birth-and-death process with catastrophe, it also incorporates the probability that the cell recovers from infection by eliminating its bacterial load.

Figure 4.4 shows the mean of $M(t)$ for both types of process following infection at different MOI. In the comparisons that follow, it is assumed that β, δ and λ are equal in both processes with the only difference being that $\mu = 0$ for a birth process with catastrophe and $\mu \neq 0$ for a birth-and-death process with catastrophe.

As expected, a larger MOI causes the average population size to decrease faster, since cells are on average infected with a greater number of bacteria. Furthermore, the average number of surviving cells is always greater for a birth-and-death process with catastrophe compared to a birth process with catastrophe. Despite this, when considering the average number of cells that ultimately survive, the difference between the two processes is not as intuitive. For a birth-and-death process with catastrophe, as the MOI increases it is less likely that the cell will be able to clear its bacterial load since the process starts in a state further away from 0. This suggests that for high MOI, the cells that ultimately survive are those that are never infected, the same as for a birth process with catastrophe, and thus the ratio of $\pi(t)$ for the two processes is approximately one. From Figure 4.4, this same conclusion could also be reached, with the solid and dashed lines appearing closer for higher MOI. However when evaluating the ratio,

$$\lim_{t \to +\infty} \frac{\pi(t \mid \mu = 0)}{\pi(t \mid \mu \neq 0)} = e^{-\lambda a}\,,$$
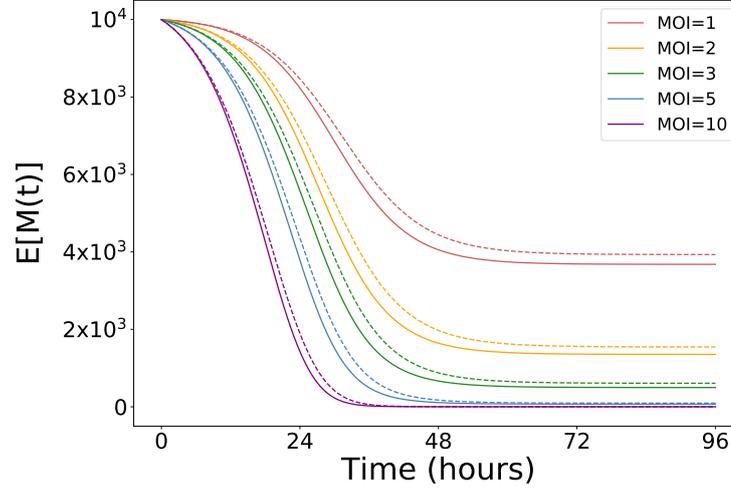
Figure 4.4: The expected number of surviving cells at time t for different MOI, assuming an initial population size of 10^4 cells. The cases where each cell is modelled as a birth process with catastrophe (solid) and birth-and-death process with catastrophe (dashed) are both considered ($\beta = 0.15\,h^{-1}$, $\mu = 10^{-2}\,h^{-1}$, $\delta = 10^{-3}\,h^{-1}$).

and so the relative difference between the two processes increases for larger MOI. This feature is only not visible in Figure 4.4 due to the scale that $\mathbb{E}\left[M(t)\right]$ is plotted on.

When considering different MOI, the coefficient of variation provides a better measure of dispersion than the standard deviation since the mean of $M(t)$ spans multiple orders of magnitude. The coefficient of variation of $M(t)$ is given here by

$$c_v(t) = \sqrt{\frac{1 - \pi(t)}{M\pi(t)}}\,,$$

and is depicted for multiple MOI in Figure 4.5, suggesting that there is greater relative variability when infecting at a high MOI. Furthermore, since the survival function is greater for a birth-and-death process with catastrophe compared to a birth process with catastrophe, it follows that $c_v(t)$ is greater for a birth process with catastrophe. Although this is true for the coefficient of variation, it is not
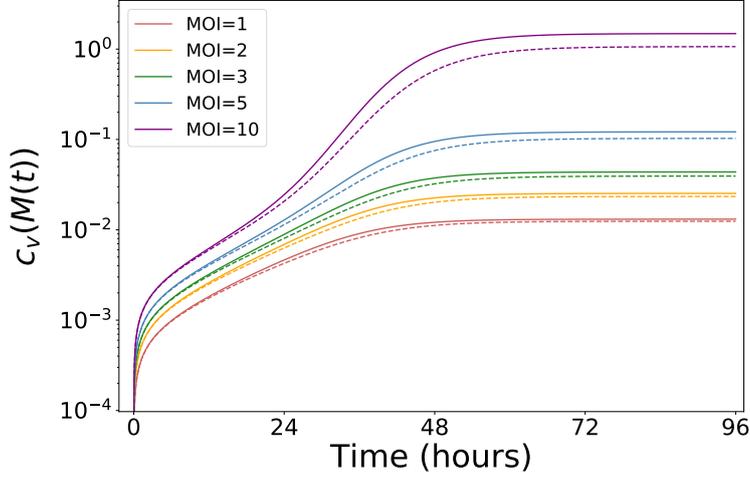
Figure 4.5: The coefficient of variation of the number of surviving cells at time t for different MOI, assuming an initial population size of 10^4 cells. The cases where each cell is modelled as a birth process with catastrophe (solid) and birth-and-death process with catastrophe (dashed) are both considered ($\beta = 0.15\,h^{-1}$, $\mu = 10^{-2}\,h^{-1}$, $\delta = 10^{-3}\,h^{-1}$).

always true for the variance. Given the shape of $\pi(t)(1 - \pi(t))$, and that the variance is initially greater for a birth process with catastrophe, it can be shown that if

$$a < \frac{1}{\lambda}\log\left(e^\lambda - 1\right),$$

then there exists a time t^* such that for times $t > t^*$, the variance of the birth-and-death process with catastrophe is larger than that of the birth process with catastrophe.

4.1.4 Linear birth-and-death processes with general catastrophes

This chapter has so far considered birth-and-death processes with catastrophe where all three rates are linearly related to the state of the process. As a generalisation, the case is now considered where the birth and death rates remain linear,

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

but the rate of catastrophe is instead given by $\delta c(X(t))$. The survival function now satisfies the ODE

$$\frac{d}{dt}S^{(k)}(t) = -\delta \mathbb{E}\left[c(X(t) \,|\, X(0) = k\right], \qquad S^{(k)}(0) = 1\,,$$

and the average number of bacteria released at time t is

$$\tilde{f}(t) = \delta \mathbb{E}\left[X(t)\,c(X(t)) \,|\, X(0) = k\right]\,.$$

For example, consider the case where the catastrophe rate is independent of the state of $\mathcal{X}$. A similar process has been studied by Van Doorn & Zeifman (2005), where catastrophes occur at a constant rate but are also allowed to occur from state 0. In this case, state B would remain the only absorbing state and the survival function is given by $e^{-\delta t}$. Here, however, given the different interpretations of the states 0 and B, it is important to consider the case where both states remain absorbing. As a result, the catastrophe rate is not constant, but instead defined as

$$c(X(t)) = \begin{cases} 1 & X(t) > 0 \text{ and } X(t) \neq B\,, \\ 0 & X(t) = 0 \text{ or } X(t) = B\,. \end{cases}$$

If $\tilde{\mathcal{X}}$ is a linear birth-and-death process without catastrophe, the survival function for the process with catastrophe therefore satisfies

$$\frac{d}{dt}S^{(k)}(t) = -\delta e^{-\delta t}\tilde{S}^{(k)}(t)\,, \qquad S^{(k)}(0) = 1\,,$$

and the average number of bacteria released from a cell at time t is

$$\tilde{f}(t) = \delta \mathbb{E}\left[X(t) \,|\, X(0) = k\right] = \delta k e^{(\beta-\mu-\delta)t}\,.$$

Given the previous definition in 4.10 of the average rupture size, $\mathbb{E}\left[R\right]$, it can be seen here that this is only finite provided that $\mu + \delta > \beta$. This restriction on when the mean is finite suggests that it is more appropriate from a biological perspective to consider a process where the birth, death and rupture rates are all linear, compared to the process described above where the rate of rupture is approximately constant.

4.2   Agent based model

In this section I provide a detailed description of the agent based model (ABM) used to study the early stages of *F. tularensis* infection. The purpose of this model is to examine populations of bacteria and macrophages in the organs of a mouse where infection is most prevalent, particularly focusing on the 48 hour period following inhalational exposure to *F. tularensis*. The ABM is ultimately used to help infer the value of specific parameters, such as the intracellular growth rate, with the help of *in vivo* infection data.

The two main entities, or agents, included in the model are macrophages and *F. tularensis* bacteria. Each bacterium is characterised by a set of attributes, described by:

Bacterium:
- **location**: the name of the organ where the bacterium is currently located, P: lung, L: liver, S: spleen, MLN: mediastinal lymph node, K: kidney.

- **cohort number**: a non-negative integer representing the number of macrophages a single bacterium has infected.

- **intracellular location**: an indicator of whether the bacterium is located extracellularly, 0, within a phagosome, P, or within the cytosol of a macrophage, C.

Individual macrophages are described according to the following attributes:

Macrophage:
- **location**: the name of the organ where the bacterium is currently located, location definitions remain the same as those for bacteria.

- **phagosomal load**: a non-negative integer representing the number of bacteria contained within phagosomes inside of the macrophage.

- **cytosolic load**: a non-negative integer representing the number of bacteria contained within the cytosol of the macrophage.

- **cohort number**: a non-negative integer equal to the cohort number of the first bacterium to infect the macrophage. Once set, this cohort number is fixed until the death of the macrophage where it is used to classify rupture events.

- **bacteria list**: a list comprising the objects that represent the bacteria contained within the phagosome or cytosol of the macrophage.

- **activation state**: the current activation state of the macrophage that is used to dictate whether the macrophage controls or facilitates infection, -1: suppressed (anti-inflammatory), 0: resting, $+1$: activated (pro-inflammatory). Only activated macrophages are capable of clearing their intracellular bacterial load.

A further set of entities included in the model are the cytokines IFN-γ and TGF-β. However, instead of being represented as objects, as is the case for bacteria and macrophages, these are represented by continuous variables governed by ODEs. Due to the different scales that cytokines and cells operate on, including molecules of cytokines as individual agents would result in such a large number of interacting agents that the model becomes inefficient. This hybrid approach of defining an ABM through both rules and ODEs has previously been used by Athale & Desiboeck (2006) and Wakeland *et al.* (2007).

Let IFN(t) and TGF(t) denote the concentration of IFN-γ and TGF-β at time t. Since IFN-γ is secreted by activated macrophages and TGF-β is secreted by suppressed macrophages, if $A(t)$ and $I(t)$ denote the respective number of activated and suppressed macrophages at time t, then

$$\frac{d}{dt}\text{IFN}(t) = \alpha_{\text{IFN}}A(t) - d_{\text{IFN}}\text{IFN}(t)\,,$$

$$\frac{d}{dt}\text{TGF}(t) = \alpha_{\text{TGF}}I(t) - d_{\text{TGF}}\text{TGF}(t)\,,$$

where α_{IFN} and α_{TGF} are the rates at which the cytokines are secreted, and the loss terms represent natural decay and consumption of cytokines through
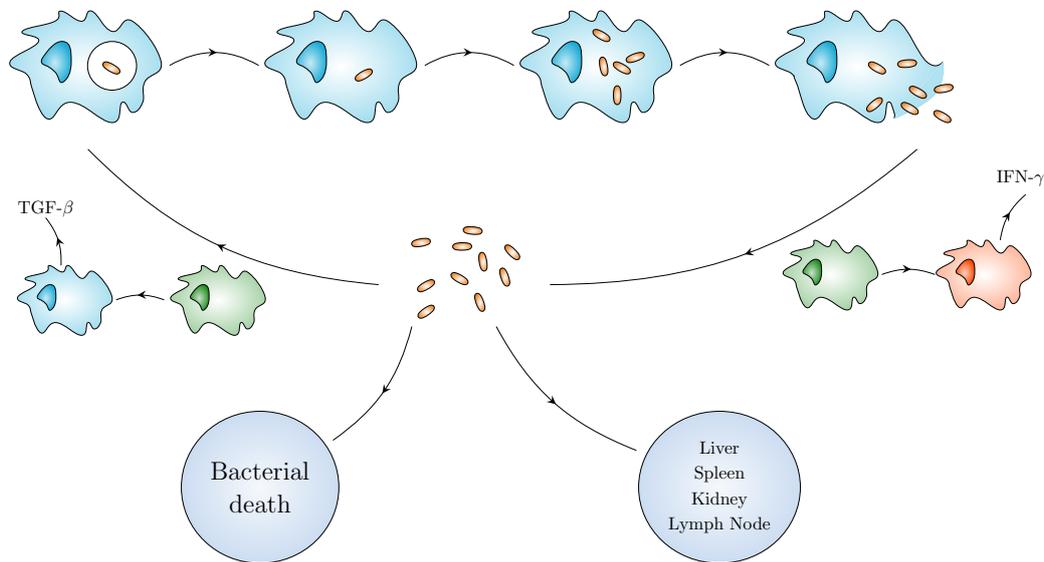
Figure 4.6: A depiction of the processes included in the ABM of *F. tularensis* infection. An initial source of bacteria enters the lung and can infect macrophages, suppressing them and resulting in the production of TGF-β. Following intracellular bacterial growth, these macrophages rupture, releasing their bacterial load and activating neighbouring macrophages. Activated macrophages contribute to clearing infection by eliminating any bacteria they phagocytose. In addition to infecting macrophages, death of extracellular bacteria can occur, as well as the migration of bacteria to different organs (resting: green, suppressed: blue, activated: red).

the process of activating and suppressing neighbouring macrophages (Mosser & Edwards (2008)).

In the ABM, the agents interact with each other according to a set of reactions, where the reactions are the same within each of the five locations. A depiction of the reactions included in the ABM is presented in Figure 4.6. Across the initial 48 hours of infection, the processes described below are simulated in the following biologically meaningful order, with each population updated after each process occurs.

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

- **Phagocytosis** (ρ (cell^{-1} bacterium^{-1} h^{-1})): a sample of extracellular bacteria is selected to enter macrophages. Both the bacteria and macrophages are chosen at random with no preference between bacteria entering uninfected or already infected macrophages. If the activation state of the chosen macrophage is resting (0), this is changed to suppressed (-1), whilst the intracellular location of each bacterium is set to P and their cohort number is increased by one. The objects representing the bacteria are appended to the bacteria list of the corresponding macrophage. If the activation state of the chosen macrophage is activated ($+1$), no bacteria are appended to the bacteria list of the macrophage. This represents that only activated macrophages are capable of clearing their bacterial load. The phagocytosed bacteria are removed from the population of extracellular bacteria.

- **Extracellular bacterial death** (rate μ (h^{-1})): a random sample of extracellular bacteria is killed by host-defence mechanisms such as anti-microbial peptides (Jones *et al.* (2012)). The objects representing each killed bacteria are removed from the population of extracellular bacteria.

- **Phagosomal escape** (rate ϕ (h^{-1})): a sample of bacteria currently residing within phagosomes escapes into the cytosol of the infected macrophage. The number of bacteria contained within phagosomes in each macrophage acts as a weight to determine how likely an escape event is to occur in the given macrophage. The intracellular location of each escaping bacterium is changed to C.

- **Intracellular bacterial replication** (rate β (h^{-1})): a sample of cytosolic bacteria replicate within infected macrophages. The number of bacteria in the cytosol of each macrophage acts as a weight to determine which macrophages these replication events occur within. An exact copy of each replicating bacterium is created such that the value of all attributes for the offspring bacterium are inherited from the parent bacterium, including the cohort number.

- **Macrophage rupture** (rate δ (h^{-1})): a sample of infected macrophages rupture and die. Macrophages are sampled such that those with a greater cytosolic bacterial load are more likely to rupture than those with a smaller bacterial load. The released bacteria are returned to an extracellular environment and so their intracellular location is set to 0. For each rupturing macrophage a second macrophage is sampled from the population, if the activation state of this macrophage is resting (0) then this is changed to activated (+1) to represent the effects of PAMPs/DAMPs that are released when a cell ruptures (D'Elia *et al.* (2013)). The objects representing the rupturing macrophages are removed from the macrophage population and the contents of each bacteria list is appended to the list of extracellular bacteria.

- **Bacterial migration** (rate γ (h^{-1})): a sample of extracellular bacteria migrate from their current location to another location. The number of bacteria that migrate to each location is a binomial random variable where the probability of migration is equal to the relative weight assigned to each organ. The choice of these weights reflects the preference of bacteria to migrate towards specific organs, as observed experimentally, and is later discussed in more detail. The location attribute of each migrating bacteria is set equal to the organ it migrates towards.

- **Suppression of neighbouring macrophages** (rate ν_{TGF} (h^{-1})): a sample of macrophages whose current activation state is resting are suppressed due to the effects of TGF-β. The activation state of each macrophage is changed to suppressed (-1). Macrophages are only suppressed in this manner after the concentration of TGF-β exceeds a specified threshold.

- **Activation of neighbouring macrophages** (rate ν_{IFN} (h^{-1})): a sample of macrophages whose current activation state is resting are activated due to the effects of IFN-γ. The activation state of each macrophage is changed to activated (+1). Macrophages are only activated in this manner after the concentration of IFN-γ exceeds a specified threshold.

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

When simulating an ABM, the order in which the events occur is important as this can impact the dynamics. Suppose here that phagocytosis of bacteria by macrophages is always the first event to occur. It is then likely that due to the large number of macrophages present in the lung, the entire initial dose would be taken up immediately. These intracellular bacteria are then able to replicate to high numbers until the macrophage ruptures. At this point, bacterial counts are then large enough that extinction of the population due to bacterial death is unlikely, and thus the population will almost certainly continue to grow. Now suppose that extracellular bacterial death is the first event to occur. Depending on the rate μ, there is a much larger probability that the population of bacteria immediately becomes extinct, particularly at very low initial doses. If this happens, the simulations will continue as if infection never occurred. One way to avoid having to choose the order of events is to allow it to vary during the simulation, that is, to randomise the order of events after each time step (Railsback & Volker (2012)). Here, however, the initial dose is large enough, and the choice of time step small enough, that the order of these first two events has little effect.

To initialise the agent based model, an initial dose of *F. tularensis* bacteria, N, is specified. The starting number of bacterial agents is then sampled from a Poisson distribution with mean N to represent possible uncertainty in this initial dose. All bacteria are initially extracellular and are present only in the lung, the cohort number of this initial dose is zero. The initial number of macrophages within the lung is given by M, whilst the initial number of macrophages in the remaining locations is specified as a fraction of M. The interpretation of M should not be that it represents the total number of macrophages present in the lung of a mouse. Since the topology of the lung is highly complex, infection is likely to only occur throughout small pockets of lung space, particularly during the first 48 hours. Therefore, M instead represents the total number of macrophages in these areas of localised infection. In this sense, the idea that rupturing of infected macrophages results in the activation of neighbouring macrophages is a plausible assumption. All macrophages are initially uninfected and begin in a resting activation state. The ABM is simulated using a tau-leaping procedure similar to that

outlined in Section 2.4.2 where the time step, τ, is now fixed. In the previously described list of events, the size of the sample is therefore drawn from a Poisson distribution with mean proportional to the corresponding rate of the event.

For each simulation of the ABM, model output is recorded at either the end of the time step, as is the case for quantities relating to the size of a specific population, or at the instance that a particular event occurs, as is the case when the quantity corresponds to a time. In order to quantify the progression of the infection, the number of extracellular, phagosomal and cytosolic bacteria is recorded for each organ, with their sum giving the total bacterial load. In addition to this, rupture events are recorded and characterised by the time and cohort number of the event, as well as the number of bacteria released. Since the cohort number of a macrophage is determined by the cohort number of the bacterium to first infect it, macrophages infected by the initial dose of bacteria each have a cohort number equal to one. The rupture events involving these macrophages are therefore referred to as first cohort rupture events. For macrophages that are initially infected by bacteria released from first cohort rupture events, their cohort number will be equal to two and thus the rupture events that follow are called second cohort rupture events.

4.3   Cohort analysis

In this section, an approximation for the bacterial load of each organ is developed by making use of the cohort numbers defined previously. The idea of modelling individual cohorts of bacteria during *F. tularensis* infection has previously been considered by Gillard *et al.* (2014). However, their approach only considers first cohort rupture events and is constructed using the less accurate survival function depicted in Figure 4.3. This section therefore acts as an extension of this approach, drawing on results of birth processes with catastrophe to improve on the existing approximations in order to apply to higher order cohorts.

Since the lung is the first organ to be infected following inhalation of *F. tularensis* bacteria, it is suitable to begin by considering the total number of bacteria

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

within the lung of an infected mouse. Once inside the lung, the initial dose of bacteria is believed to be phagocytosed primarily by alveolar macrophages (Hall *et al.* (2008)). Given that these phagocytosis events occur quickly, it is then reasonable to assume that time zero corresponds to the time at which the initial dose of bacteria are all residing in the phagosomes of macrophages. Furthermore, if the number of macrophages exposed to the initial dose of bacteria is significantly larger than the size of the initial dose, it is also appropriate to assume that each infected macrophage contains only a single bacterium. Since the initial dose contains N bacteria, the initial state of the system is one where N macrophages are each infected with one bacterium, contained within a phagosome.

Let $P_j(t)$ and $C_j(t)$ denote the respective mean number of cohort j bacteria in macrophage phagosomes and cytosols at time t. A natural starting point is to consider $P_1(t)$, representing the initial dose of phagosome residing bacteria. Given that bacteria escape phagosomes with rate $\phi\,h^{-1}$, $P_1(t)$ decays exponentially according to

$$\frac{d}{dt}P_1(t) = -\phi P_1(t)\,, \qquad\qquad P_1(0) = N\,.$$

For the population of first cohort bacteria in cytosols, $C_1(t)$, each cell may be thought of as a birth process with catastrophe, $\mathcal{X} = \{X(t) : t \geq 0\}$, where the state of the process represents the number of intracellular bacteria. The average size of this population is then given by the expectation of $\mathcal{X}$, previously defined in (4.6), for the case $X(0) = 1$. However, this currently assumes that time zero corresponds to the time at which the bacterium enters the cytosol, but here time zero corresponds to the time immediately after a bacterium enters the cell. This delay can be accounted for by using the p.d.f. of an exponential distribution, the distribution of time a bacterium spends in the phagosome. In doing so, the number of first cohort bacteria in cytosols is given by

$$C_1(t) = N\phi \int_0^t \mathbb{E}\left[X(t-s)\right] e^{-\phi s}\,ds = N\frac{\phi}{\delta} \int f(t-s)e^{-\phi s}\,ds\,, \qquad (4.13)$$

where $f(t)$ is the p.d.f. of the time until rupture of a macrophage initially infected by a single bacterium, defined in (4.7). In order to determine the size of higher

order cohort populations, an understanding of the rupture events is required. Recall that just as $f(t)$ is the probability that a macrophage ruptures at time t, $\tilde{f}(t)$ is the average number of bacteria released by a macrophage at time t. A comparison between these functions and the average of 2.5×10^2 simulations of the ABM is provided in Figure 4.7. By averaging over a large number of realisations, the histogram matches well to the the smooth functions $f(t)$ and $\tilde{f}(t)$. However, the number of macrophages that survive to time t is not a smooth function but instead decreases in steps. In this sense, it is also interesting to see how $NS(t)$ compares with the survival curve for a single realisation. This is provided in Figure 4.8 along with a comparison between $\bar{n}(t)$ and the number of bacteria each macrophage releases on rupture.

It should be noted that since $f(t)$ and $\tilde{f}(t)$ are derived using the birth process with catastrophe, $\mathcal{X}$, they do not incorporate the time that a bacterium spends in the phagosome. As for $C_1(t)$, this delay can be accounted for by considering the convolution of each function with the density of the time spent in the phagosome. However, since this delay of approximately half an hour is significantly shorter than both the total time until macrophage rupture and the doubling time of a *F. tularensis* bacterium, there is little difference when using the convolution (Jones *et al.* (2012); Wood *et al.* (2014)). For this reason, $f(t)$ and $\tilde{f}(t)$ are continued to be expressed in their current form.

The functions $f(t)$ and $\tilde{f}(t)$ can now be used to help determine the number of second cohort bacteria residing in macrophage phagosomes. This population increases as a result of bacteria released from first cohort rupture events infecting neighbouring macrophages, and decreases due to bacteria escaping the phagosome into the cytosol. By neglecting the time spent extracellularly, $P_2(t)$ satisfies the ODE

$$\frac{d}{dt}P_2(t) = N\tilde{f}_1(t) - \phi P_2(t), \qquad P_2(0) = 0\,.$$

For $C_2(t)$, the decay in the population due to bacteria being released from second cohort rupture events can be determined by considering the convolution of $\tilde{f}(t)$ with itself. Suppose that a macrophage is initially infected by a single bacterium at time zero and releases on average $\tilde{f}(s)$ bacteria when it ruptures at time s.
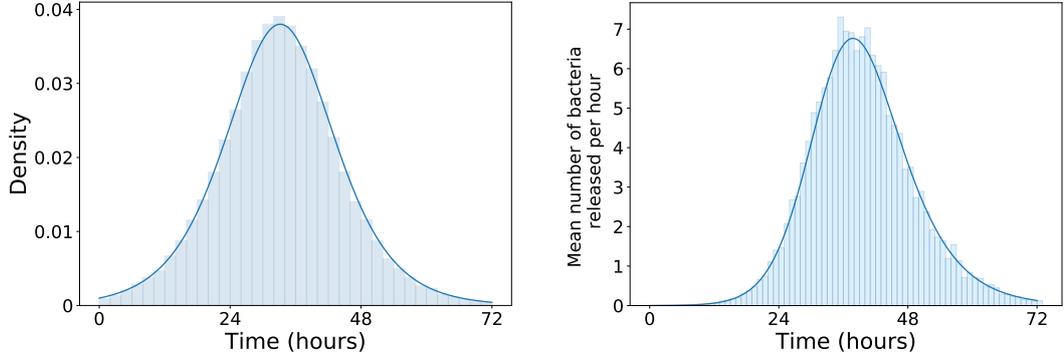
Figure 4.7: A comparison between the mean of 2.5×10^2 realisations of the ABM and the functions $f(t)$ and $\tilde{f}(t)$ ($\beta = 0.15\,h^{-1}$, $\delta = 10^{-3}\,h^{-1}$).

If these $\tilde{f}(s)$ bacteria each infect a different uninfected macrophage, then the number of bacteria released at time t is $\tilde{f}(s)\tilde{f}(t-s)$ and the average number released can be found by integrating over the possible times of the first rupture event. This argument can be extended to show that the mean number of bacteria released in n^{th} cohort rupture events is given by

$$\tilde{f}_n(t) = \int_0^t \tilde{f}_{n-1}(s)\tilde{f}(t-s)\,ds\,.$$

Similarly, the times at which n^{th} cohort rupture events occur are defined according to

$$f_n(t) = \int_0^t \tilde{f}_{n-1}(s)f(t-s)\,ds\,. \tag{4.14}$$

From this, the number of nth cohort bacteria contained within phagosomes increases due to previous cohort rupture events and decreases due to bacteria escaping the phagosome. The number of nth cohort bacteria residing in cytosols increases due to bacteria entering from phagosomes along with bacterial replication, but decreases due to current cohort rupture events. This allows the following system of ODEs to be constructed which describes the number of bacteria in each cohort within the lung
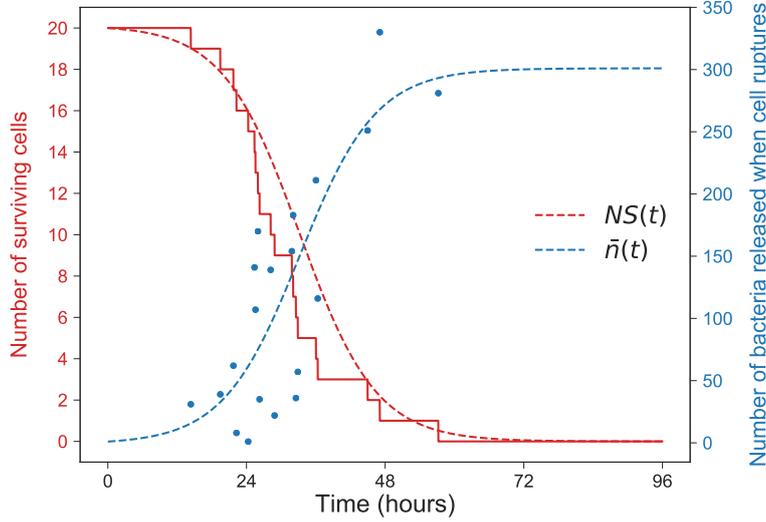
Figure 4.8: Survival curve for a single realisation of $N = 20$ macrophages, each initially infected with a single bacterium. The number of bacteria released when each macrophage ruptures is also indicated at the corresponding time of rupture ($\beta = 0.15\,h^{-1}$, $\delta = 10^{-3}\,h^{-1}$, $\phi = 2\,h^{-1}$).

$$\frac{d}{dt}P_n(t) = -\phi P_n(t) + N\tilde{f}_{n-1}(t) \qquad P_n(0) = 0\,, \quad n = 2, ..., Z\,,$$

$$\frac{d}{dt}C_n(t) = \phi P_n(t) + \beta C_n(t) - N\tilde{f}_n(t) \quad C_n(0) = 0\,, \quad n = 2, ..., Z-1\,, \quad (4.15)$$

$$\frac{d}{dt}C_Z(t) = \phi P_n(t) + \beta C_n(t) \qquad C_Z(0) = 0\,.$$

If there is sufficient migration of uninfected cells into the localised areas of infection within the lung, it is possible that bacteria released in each rupture event enter a unique uninfected macrophage. In this sense, system (4.15) could continue indefinitely. However, as the infection progresses and the total number of bacteria grows, the likelihood that this assumption holds true decreases. For this reason, the system is truncated to include only the first Z cohorts, where the interpretation of Z is that a cohort rupture event of order Z will not occur during the time period being modelled. The choice of Z should therefore reflect the time period over which the approximation is being applied, and this should in turn depend on how long this assumption is likely to hold true for. This approximation

is also independent of ρ, the rate of phagocytosis, since the time that bacteria spend extracellularly is assumed to be negligible compared to the time they spend intracellularly. The same assumption is also made by Brown *et al.* (2006) in their model of *S. enterica* infection. If the extracellular phase is believed to be important, (4.15) could be extended to account for this by including a variable for the number of extracellular cohort j bacteria. This variable would then increase due to the release of bacteria from current cohort rupture events and decrease as a result of bacteria infecting macrophages. The solution to system (4.15) is provided in Figure 4.9 along with the average number of bacteria in each cohort, as determined by the ABM.

Given (4.15), the mean total number of bacteria present in the lung is obtained by summing the individual cohorts

$$T_{lung}(t) = \sum_{n=1}^{Z} \left(P_n(t) + C_n(t) \right) ,$$

and it is now possible to consider the bacterial load in the remaining four organs. One possible mechanism for the migration of bacteria to different organs is the movement of cells containing bacteria, with Kirby *et al.* (2009) previously showing that alveolar macrophages are involved in the transport of *S. pneumoniae* to the lung draining lymph nodes. A second mechanism involves the direct entry of *F. tularensis* bacteria into the blood. Despite this, the experimental measurements used to infer the model parameters only concern the total number of bacteria present in each organ, and as a result it would not be possible to distinguish between these different mechanisms of bacterial migration. Therefore, as described by the ABM, the number of migrating bacteria is assumed to be proportional to the number of extracellular bacteria, and thus in order to find the number of bacteria in the remaining organs, the number of extracellular bacteria in the lung must first be determined. This may seem unusual given the previous assumption that the time bacteria spend extracellularly in the lung can be neglected. However, this extracellular phase, along with the time spent in phagosomes, acts only as a delay in the constant growth of the population of bacteria. Since this delay is short in comparison to the total length of the intracellular life cycle, it
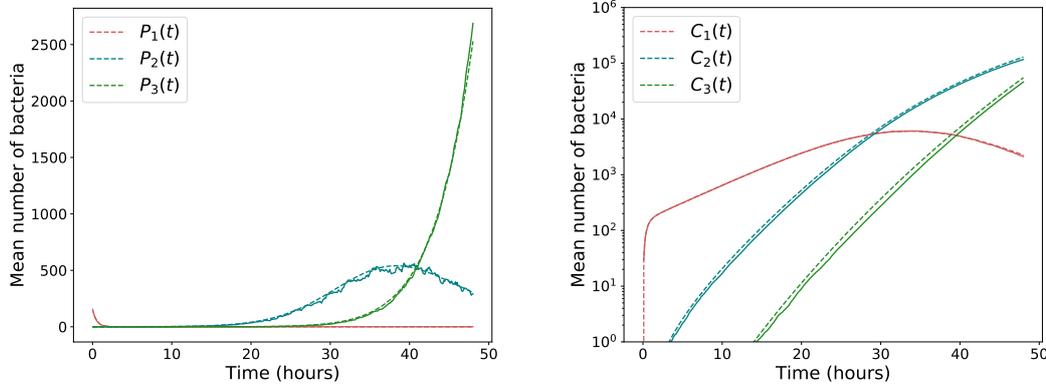
Figure 4.9: The first three cohorts of phagosome and cytosol residing bacteria as determined by system (4.15) (dashed) compared to averages obtained from 10^2 realisations of the ABM (solid). Solutions are only plotted for values exceeding one bacterium ($N = 160$ bacteria, $M = 10^4$ macrophages).

is acceptable not to account for extracellular bacteria when finding intracellular bacterial loads, yet it is essential for considering the total number of bacteria in the remaining organs.

In the ABM, bacteria are released extracellularly through successive rupture events. Once extracellular, these bacteria either infect a different macrophage with rate ρ (cell^{-1} bacterium^{-1} h^{-1}), migrate to another organ with rate γ (h^{-1}), or are killed with rate μ (h^{-1}). If $E(t)$ denotes the expected number of extracellular bacteria in the lung at time t, then

$$\frac{d}{dt}E(t) = N \sum_{n=1}^{Z-1} \tilde{f}_n(t) - E(t)\left[M\rho + \gamma + \mu\right], \qquad E(0) = 0.$$

Suppose now that $\mathcal{S}$ denotes the set of organs, aside from the lung, that bacteria can migrate towards. Let $P^j(t)$, and $C^j(t)$ denote the respective mean number of bacteria in phagosomes and cytosols of macrophages residing in organ $j \in \mathcal{S}$. Given that the average time for a macrophage in the lung to rupture is in excess of 24 hours, and only the first 48 hours of infection are currently being considered, the assumption is made that no rupture events will occur within the remaining

organs during this period. This assumption has been verified using the ABM, where fewer than 10 rupture events occurred on average in the lymph node, liver, kidney and spleen compared to greater than 10^3 in the lung. As a result of this assumption, the bacteria that migrate to another organ infect macrophages and remain there until at least the end of the 48 hour period. Since the dynamics of infection are the same in all organs, $P^j(t)$ and $C^j(t)$ obey a similar system of ODEs as their counterparts for the lung

$$
\begin{aligned}
\frac{d}{dt}P^j(t) &= E(t)\gamma w_j - \phi P^j(t) \qquad P^j(0) = 0\,, \\
\frac{d}{dt}C^j(t) &= \phi P^j(t) + \beta C^j(t) \qquad C^j(0) = 0\,,
\end{aligned}
\tag{4.16}
$$

where w_j are the weights assigned to each organ that determine how likely it is for a migrating bacterium to move to that particular organ.

Together, (4.15) and (4.16) provide a simple way of studying the mean number of bacteria in each organ of an infected mouse. One feature of the ABM is the large variability that can be observed between simulations. Following first cohort rupture events, the majority of bacteria enter macrophages in the lung, however, should migration of bacteria occur at such an early time point, the extent of the replication can often skew averages for the remaining organs. Although this can be resolved by increasing the number of simulations of the ABM, this can be time consuming and therefore means that it can be difficult to characterise the infection from simulations alone.

4.4 Parameter values

In this section, I describe how a Bayesian approach to parameter inference can be used to learn about the most important parameters by comparing the model output with *in vivo* infection data. The approximations derived in Section 4.3 provide a faster approach for evaluating the model than simulating the ABM, a necessary requirement for the ABC algorithm that is implemented here.

4.4.1 Experimental data

The data presented in this section and used in the parameter inference are provided by Dr. Roman Lukaszewski from the Defence Science and Technology Laboratory (Dstl). These quantitative data contain bacterial counts of *F. tularensis* SCHU S4 in the lungs, MLN, liver, spleen and kidney of mice infected at two different doses. A high dose of 160.33 CFUs and a medium dose of 13.7 CFUs represent an estimated average number of CFUs that were delivered to each mouse. This calculation takes into account the breathing rate of the mouse and the length of time that the mouse was exposed to the aerosolised bacteria. For each dose, six different mice were culled at multiple time points within the first 48 hours of infection and their bacterial burden recorded. To determine the bacteria load, samples of each organ were disrupted through a cell sieve to give a single cell suspension. Serial dilutions of this suspension were then plated and left to allow colonies of bacteria to form. By counting the number of colonies and accounting for the dilution factor, the number of bacterial CFUs in the original sample can be estimated.

Presented in Table 4.1 are the bacterial counts for each mouse, organ and time point, provided that at least one of the six mice had bacterial counts greater than zero. Since these data are used to infer parameter values, they must be reliable. The measurements at one hour post infection show great variability, with no bacteria present in the measured organs of some mice, whilst other mice have bacterial loads greater than the initial dose. For this reason, only the values presented in Table 4.1 for times after one hour post infection are considered reliable and used in the parameter inference.

4.4.2 Sensitivity analysis

Although the data in Table 4.1 corresponds to measurements for multiple mice, doses and time points, it would be difficult to learn about all model parameters from the data since only the total number of bacteria has been observed. For this reason, only the values of those parameters that are least known about in

High infectious dose (160.33 CFUs)

time (hours)	organ	mouse						mean	SD
		1	2	3	4	5	6		
1	lung	0	0	2.50×10^2	1.50×10^2	0	3.50×10^1	1.05×10^1	1.42×10^1
18	lung	2.60×10^3	3.25×10^3	1.50×10^3	1.55×10^3	1.20×10^3	2.50×10^3	1.97×10^1	1.48
	MLN	4.00×10^1	8.50×10^1	5.00×10^0	7.00×10^1	2.45×10^2	6.50×10^1	5.16×10^1	3.64
24	lung	4.95×10^3	3.15×10^3	3.15×10^3	4.90×10^3	1.41×10^3	1.50×10^3	2.83×10^5	1.74
48	lung	2.65×10^5	2.85×10^5	1.30×10^6	1.70×10^5	4.90×10^5	4.25×10^5	3.89×10^5	2.01
	MLN	2.05×10^3	6.15×10^3	3.40×10^3	1.50×10^2	2.55×10^3	1.20×10^3	1.64×10^3	3.64
	liver	5.00×10^2	6.00×10^2	4.00×10^2	3.00×10^2	1.00×10^3	6.00×10^2	5.28×10^2	1.51
	kidney	3.50×10^2	2.00×10^2	3.50×10^2	1.00×10^2	1.00×10^2	1.50×10^2	1.82×10^2	1.77
	spleen	3.00×10^1	5.00×10^1	8.50×10^2	3.00×10^1	5.50×10^2	5.50×10^2	1.50×10^2	4.94

Medium infectious dose (13.7 CFUs)

time (hours)	organ	mouse						mean	SD
		1	2	3	4	5	6		
1	lung	1.00×10^1	0	0	5.00×10^1	0	0	2.82	5.38
12	lung	5.00×10^1	5.00×10^0	0	0	0	2.00×10^1	4.14	5.58
18	lung	2.5×10^2	1.90×10^2	4.35×10^2	1.65×10^2	3.90×10^2	1.50×10^2	2.42×10^2	1.57
24	lung	1.25×10^3	0	0	3.00×10^2	1.50×10^2	1.85×10^2	4.67×10^1	2.15×10^1
48	lung	1.10×10^4	2.75×10^4	7.65×10^4	8.85×10^4	2.20×10^4	3.25×10^4	3.37×10^4	2.19
	MLN	1.95×10^3	4.95×10^2	2.00×10^3	5.35×10^3	2.25×10^2	1.20×10^3	1.19×10^3	3.08

Table 4.1: Bacterial counts for the lung, MLN, liver, kidney and spleen following exposure of mice to either 160.33 CFUs (top) or 13.7 CFUs (bottom) of *F. tularensis* SCHU S4 bacteria. Only data with at least one non-zero value are reported. Geometric means and standard deviations are also given, where observations showing zero bacteria have been replaced with one for the purpose of this calculation.

the literature, or those parameters that have the most influence on the model dynamics, will be inferred. For example, it is known that *F. tularensis* bacteria spend approximately 0.5–1 hours in the phagosome, so if ϕ is not identified as an important parameter, it would be suitable to fix the value of ϕ within this range.

To determine the parameters that have the greatest influence, Sobol sensitivity indices are computed for the total number of bacteria in each organ. First-order (S_1) and total-order (S_T) indices are depicted in Figure 4.10 for the initial 48 hours of infection. Since (4.15) depends on β, ϕ and δ, the Sobol indices in Figure 4.10 have only been included for these parameters. The ranges over which these parameters are varied are $\beta \in [0, 0.5]$, $\phi \in [0.5, 5]$ and $\delta \in [0, 10^{-2}]$. For the remaining organs, (4.16) has an additional dependence on the number of extracellular bacteria in the lung, and therefore the total number of bacteria in these organs also depends on both μ, γ and $M\rho$. In this case the Sobol indices have been computed for all six parameters. The ranges for these additional three parameters are $\mu \in [0, 0.1]$, $\log_{10} \gamma \in [0, 3]$ and $\log_{10} M\rho \in [-2, 5]$. The decision to include the product $M\rho$ as a single parameter relates to the interpretation of M, and the difficulty in determining the number of macrophages present in localised regions of infection within each organ. Since the dynamics in the MLN, liver, kidney and spleen are described by the same system of equations, the sensitivity analysis produces the same results and so only the Sobol indices for the MLN have been included in Figure 4.10.

For the lung, the most important parameter is β, the intracellular growth rate of bacteria, whilst ϕ has some importance during the early hours but this quickly diminishes. The rupture rate of a macrophage, δ, has a first order and total order Sobol index of zero because of the assumption that *F. tularensis* bacteria immediately infect macrophages after being released through rupture events. This again relates to the idea that rupture events act as delays in the continual growth of the bacterial population. If a rupture event causes no delay in this growth, then it can be expected that the rate at which rupture events occur will have no effect on the total number of bacteria.

For the MLN, and similarly the remaining organs, $M\rho$ and γ are the most
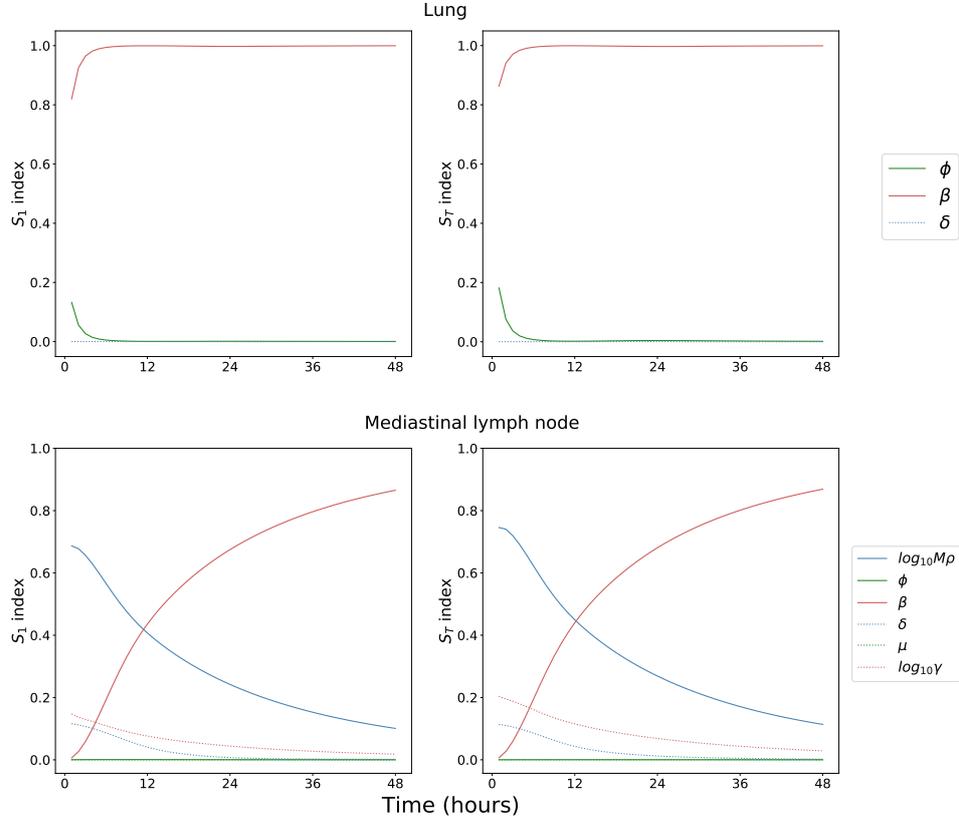
Figure 4.10: First order (S_1) and total order (S_T) Sobol sensitivity indices for the total number of bacteria in the lung and MLN during the initial 48 hours of infection with *F. tularensis*.

influential parameters during the early stages of infection. This is understandable since it is the ratio of these parameters that determines whether a bacterium released in a rupture event will remain in the lung or migrate to a different organ. As the infection progresses and it becomes increasingly likely that a bacterium has migrated to the MLN, β takes over as the most important parameter, dictating the growth of the population. Unlike for the lung, δ appears to have some importance during the early stages of infection in the other organs. Here, a small value of δ corresponds to rupture events occurring at later times. Whilst this has no effect on the way dynamics are modelled in the lung, at least one rupture event is necessary for bacterial counts to increase in the remaining organs.

Later rupture times increase the delay until this can happen, thereby reducing the number of bacteria present in these organs.

Together, the sensitivity analysis suggests that β, $M\rho$ and γ are the most important parameters for describing the total bacterial load across all organs. Although an approximate doubling time of *F. tularensis* bacteria is known, obtaining an accurate value of β is desirable given that it has the most importance (Jones *et al.* (2012)). As a result of this, the values of these three parameters will be inferred from the experimental observations.

4.4.3 Organ weights

Whilst γ is the total rate at which bacteria migrate away from the lung, the weights w_j, $j \in S$ are used to determine exactly which organ a migrating bacterium enters. As described previously, the limited data available means that it would be difficult to use ABC to learn about all four weights, in addition to β, $M\rho$ and γ. Instead, to determine the values of these weights, the proportion of bacteria that reside in the MLN, liver, spleen and kidney is computed for each of the six mice that were culled at 48 hours. These proportions are determined using total bacterial counts across all organs excluding the lung. By averaging over the different mice and then over the medium and high doses, the mean proportion of bacteria in each organ can be found. The weights are then defined using these mean proportions,

$$w_{MLN} = 0.8\,, \qquad w_{liver} = 0.11\,, \qquad w_{spleen} = 0.05\,, \qquad w_{kidney} = 0.04\,.$$

The larger weight assigned to the MLN is understandable from a biological perspective, since macrophages carrying bacteria are likely drained through the lymphatic system to the MLN earlier than it takes for bacteria to enter the blood (Bosio *et al.* (2007); Hodges *et al.* (1965)).

4.4.4 Bayesian inference

In order to learn about β, $M\rho$ and γ, an ABC rejection sampling algorithm is implemented. Prior beliefs regarding each parameter are encoded by the prior

distributions, selected here to be $\beta \sim U(0, 0.5)$, $\log_{10}(M\rho) \sim U(-2, 5)$ and
$\log_{10}\gamma \sim U(0, 3)$. The choice of prior distribution for β is due to the belief
that the doubling time of a single *F. tularensis* bacterium is approximately 6-7
hours (Jones *et al.* (2012)). Since little is known about the remaining parameters,
the ranges of these prior distributions are chosen to be large enough such that
regions of parameter space that have high posterior density are not mistakenly
omitted.

For each sampled set of parameters values, the total number of bacteria in
each organ is computed using (4.15) and (4.16), and compared to reliable exper-
imental observations as described in Section 4.4.1. When evaluating the model,
the remaining parameters are fixed according to the values provided in Table 4.3.
The model output is compared to the data using the distance

$$d^2(M, E) = \sum_{i \in \mathcal{D}} \sum_{j \in \mathcal{S}} \sum_{t \in \mathcal{T}_{i,j}} \left(\frac{\log\left(T_{i,j}^{(M)}(t)\right) - \log\left(\bar{T}_{i,j}^{(E)}(t)\right)}{\sigma_{i,j}(t)} \right)^2,$$

where $\mathcal{D} = \{160.33, 13.7\}$ is the set of initial doses and $\mathcal{T}_{i,j}$ is the set of times at
which measurements were recorded for dose $i \in \mathcal{D}$ and organ $j \in \mathcal{S}$. The variable
$\bar{T}_{i,j}^{(E)}(t)$ is the average bacterial burden in organ $j \in \mathcal{S}$ for the six mice culled at
time t, whilst $\sigma_{i,j}(t)$ is the geometric standard deviation. The values for these
variables are reported in Table 4.1.

In total, 10^6 iterations of the ABC rejection sampling algorithm were per-
formed, with an acceptance rate of 0.5% resulting in an accepted posterior sam-
ple containing 5×10^3 parameter sets. Summary statistics for these approximate
posterior samples are provided in Table 4.2. Pointwise median predictions for
the mean total number of bacteria in each organ have been constructed using
the posterior sample and are provided in Figure 4.11, showing how the model
approximations compare with the experimental data. Overall, these approxima-
tions compare well, particularly with bacterial counts in the lung, however, some
data points lie outside of the 95% credible regions. This may be due to the nar-
row posterior distribution obtained for β, that in turn is a reflection of our choice
to fix the initial doses. It is likely that there is variability in the initial doses
reported here, and accounting for this variability could yield a wider posterior
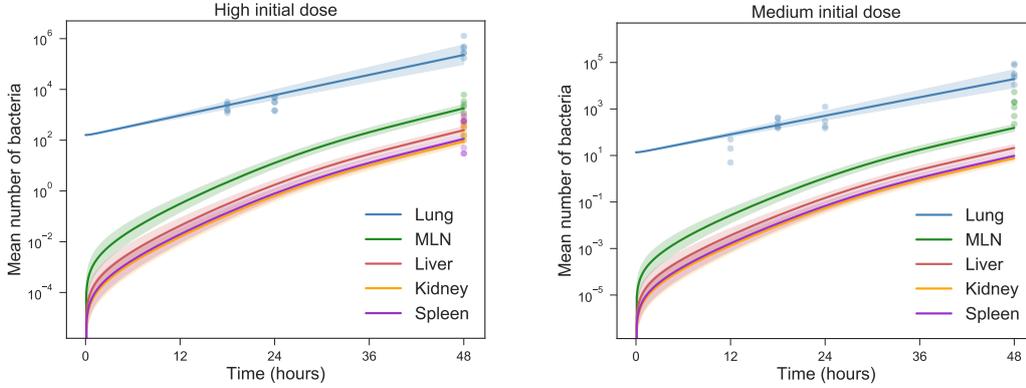
Figure 4.11: Pointwise median predictions showing the mean total number of *F. tularensis* bacteria in each organ for both medium and high initial doses. Shaded regions depict 95% credible regions.

distribution for β and thus wider credible regions that may better explain all of the measurements.

The posterior histogram for β is provided in Figure 4.12, with its narrow range indicating that it is possible to learn a lot about the intracellular growth rate. This is particularly important since the sensitivity analysis suggests that β is the most influential parameter. Furthermore, the posterior median estimate of $\beta_{median} = 0.1532\,h^{-1}$ is close to the original value of $\beta = 0.15\,h^{-1}$ predicted by Gillard *et al.* (2014). Figure 4.13 shows the bivariate posterior distribution of $M\rho$ and γ. The strong correlation between these two parameters suggests that it is difficult to learn about their individual values, however, it is possible to learn about the ratio $\gamma/M\rho$, the histogram for which is also provided in Figure 4.13. This correlation is understandable given that the inference has been performed using data that only describes total bacterial counts. Increasing $M\rho$ increases the likelihood that bacteria released from rupture events in the lung will continue to infect macrophages in the lung. In order to obtain the same level of infection in the remaining organs, the value of γ must therefore also be increased such that a greater number of bacteria also migrate away from the lung.

A closely related quantity that is of greater interest to learn about is the
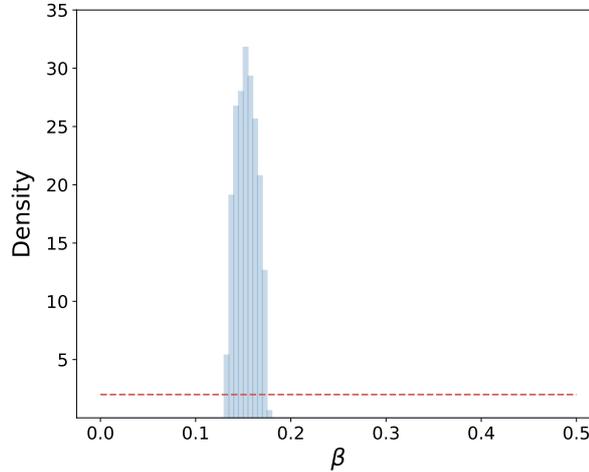
Figure 4.12: Posterior histogram for β with the corresponding prior distribution
also indicated in red.

probability that a bacterium migrates to a different organ instead of being killed
or causing further infection events in the lung. During the early stages of infec-
tion, when only a small number of rupture events have occurred and thus $M\rho$ is
approximately constant, this probability is given by $\gamma/(\gamma + M\rho + \mu)$. The poste-
rior distribution of this probability is provided in Figure 4.14 and suggests that
approximately $0 - 2\%$ of bacteria released from the first cohort rupture events
will migrate to a different organ. To obtain this distribution, a fixed value of
$\mu = 10^{-2}\,h^{-1}$ has been assumed.

The problems with identifiability that arise when trying to learn about γ and
$M\rho$ are due to only recording measurements of bacterial counts. This suggests
that, although different posterior values of these parameters result in similar
total bacterial counts, they may lead to distinct behaviour when considering a
separate quantity. For example, the ABM can also be used to identify the pro-
portion of cells in resting, activated and suppressed activation states. Figure
4.15 shows the proportion of cells in each of these activation states for multi-
ple values of $M\rho$. Since a single value of $M\rho$ corresponds to multiple values
of γ, linear regression is used to approximate the linear relationship that exists
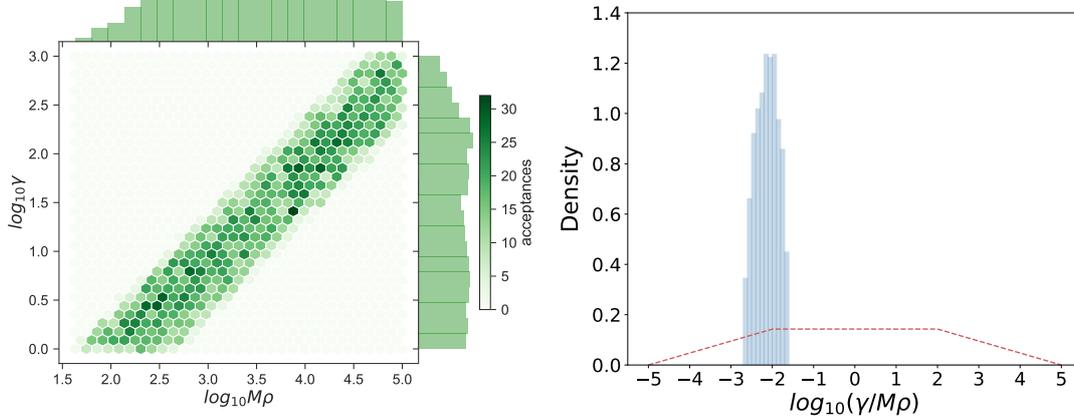
Figure 4.13: Left: bivariate posterior distribution of $\log_{10}\gamma$ and $\log_{10}(M\rho)$ showing the strong correlation that exists between these two parameters. Right: posterior distribution of the ratio $\log_{10}(\gamma/(M\rho))$ with the corresponding prior distribution indicated in red.

between $\log_{10}M\rho$ and $\log_{10}\gamma$. That is, given $M\rho$, γ is calculated according to $\log_{10}\gamma = -1.84 + 0.92\log_{10}M\rho$. Furthermore, in each of these plots a high initial dose and the posterior median estimate of β are used. It should also be noted that the ABM is parametrised by an initial number of macrophages, M, instead of the product $M\rho$, so for this purpose $\rho = 0.1$ cell^{-1} bacterium^{-1} h^{-1} is fixed and thus M is varied. Alternatively, M could have been fixed and ρ varied, however, this produces very similar behaviour for different ρ when M is large.

Since the majority of suppressed macrophages are also infected, the plots in Figure 4.15 suggest that there reaches a point where almost all macrophages in the localised areas of infection become infected, with only activated macrophages remaining uninfected. This is followed by a decline in the population of macrophages until only activated macrophages survive, since these macrophages are capable of clearing their intracellular bacterial load. This pattern is seen across each value of $M\rho$ but after a delay that increases as $M\rho$ increases. This delay is due to the assumption that the initial dose is the same for each scenario. As a result
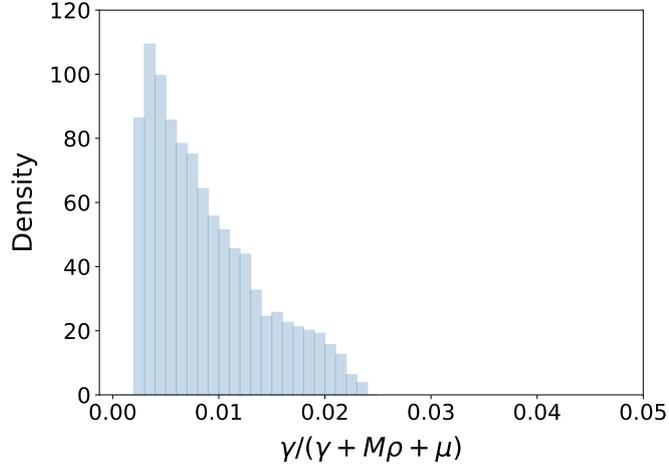
Figure 4.14: Posterior histogram for the probability that a bacterium in the lung migrates to a different organ as opposed to getting killed or infecting a different macrophage in the lung. The histogram is constructed using posterior distributions for $M\rho$ and γ, and a fixed value of $\mu = 10^{-2}\,h^{-1}$.

	Min.	1$^{\text{st}}$ Qu.	Median	Mean	3$^{\text{rd}}$ Qu.	Max.
β	1.3×10^{-1}	1.5×10^{-1}	1.5×10^{-1}	1.5×10^{-1}	1.6×10^{-1}	1.8×10^{-1}
$M\rho$	4.3	6.6×10^{2}	3.5×10^{3}	1.4×10^{4}	1.8×10^{4}	1.0×10^{5}
γ	1.0	4.9	2.5×10^{1}	1.1×10^{2}	1.3×10^{2}	1.0×10^{3}

Table 4.2: A summary of the posterior samples for each model parameter included in the Bayesian inference.

of this, the same number of macrophages initially become infected, however, this represents a larger proportion of the total population size when M is small. It then follows that the same number of bacteria will, on average, be released in first cohort rupture events. When M is small, these released bacteria then infect a larger proportion of uninfected macrophages compared to when M is large. This indicates that higher order cohort rupture events are required to infect all macrophages for larger M, therefore resulting in the observed delay.

In the plot corresponding to $\log_{10} M\rho = 2$, it can be seen that the proportion of infected cells is close to zero at 48 hours post infection. As *F. tularensis*

	Description	Parameter value	Reference
$M\rho$	per bacterium infection rate	3526.62 bacterium^{-1} h^{-1}	estimated
ϕ	rate of bacterial escape from phagosome	$2\,h^{-1}$	[1]
β	rate of intracellular bacterial growth	$0.1532\,h^{-1}$	estimated
δ	rate of macrophage rupture	$0.001\,h^{-1}$	[2]
μ	rate of extracellular bacterial death	$0.01\,h^{-1}$	[3]
γ	per bacterium migration rate	$24.86\,h^{-1}$	estimated

Table 4.3: A description and values of the parameters used in the approximation of the ABM. Posterior median values are reported for parameters inferred from infection data. [1]: Jones *et al.* (2012), [2]: Marino & Kirschner (2004), [3]: Lowrie *et al.* (1979).

replicates intracellularly, growth of the bacterial population can not be sustained in the absence of cells. There this scenario cannot agree with the data in Table 4.1, where the mean number of bacteria present in the lung exceeds 10^5. This disagreement arises from the assumption in the approximation of the ABM that there are sufficiently many macrophages for each bacteria released from rupture events to infect. Since $M = 10^3$ cells in this case, this assumption is likely not true once first cohort rupture events have occurred. Therefore, although $\log_{10} M\rho = 2$ has high posterior density, the particular parametrisation where $M = 10^3$ cells and $\rho = 0.1$ cell^{-1} bacterium^{-1} h^{-1} could not give rise to the experimental data. This shows that by inspecting model output that does not relate to total bacterial counts, it is possible to rule out certain values of parameters and hence further refine the parameter space.

During the early stages of *F. tularensis* infection, bacteria go undetected by the immune system. However, this is followed by a sharp increase in pro-inflammatory responses, such as the migration of effector cells and secretion of pro-inflammatory cytokines. In particular, the migration of neutrophils into the lung is believed to act as a disadvantage to the survival of the host, with *F. tularensis* bacteria able to prolong the lifespan of neutrophils (Schwartz *et al.* (2012)). In prolonging the lifespan, the process of neutrophil apoptosis is perturbed and neutrophils can
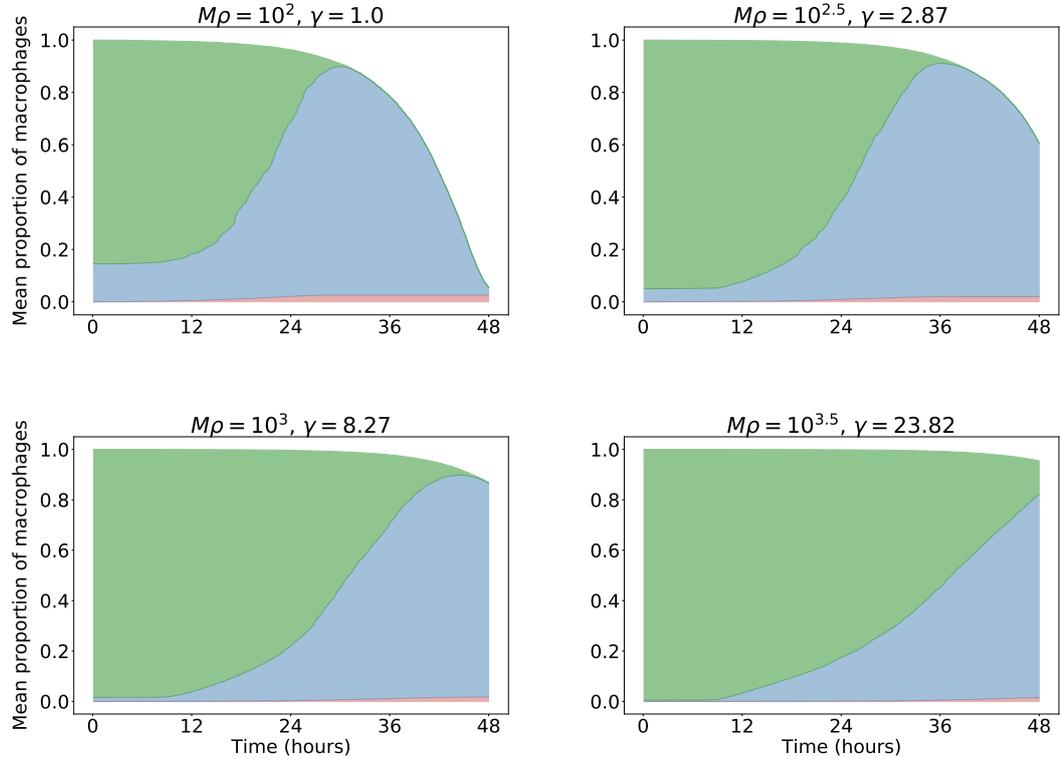
Figure 4.15: The proportion of macrophages that are in resting (green), suppressed (blue) and activated (red) activation states for different values of $M\rho$ and corresponding values of γ. In each plot, $\rho = 0.1$ cell^{-1} bacterium^{-1} h^{-1} whilst M is varied.

develop a pro-inflammatory phenotype that ultimately contributes to sustaining infection (Allen (2013)). It is therefore interesting to consider the time at which this migration of cells into the lung occurs, as this could represent a marker of disease progression. The ABM simulations indicate that there is a time at which all cells in the localised areas of infection are either activated or have ruptured, and so the migration of cells is likely to occur before this time. Therefore, the time until only activated macrophages survive could act as this marker, corresponding to the mouse succumbing to infection in the absence of treatment. Figure 4.16 shows the mean time until only activated macrophages survive for both high and medium initial doses and for different values of $M\rho$. As in Figure 4.15, $\rho = 0.1$
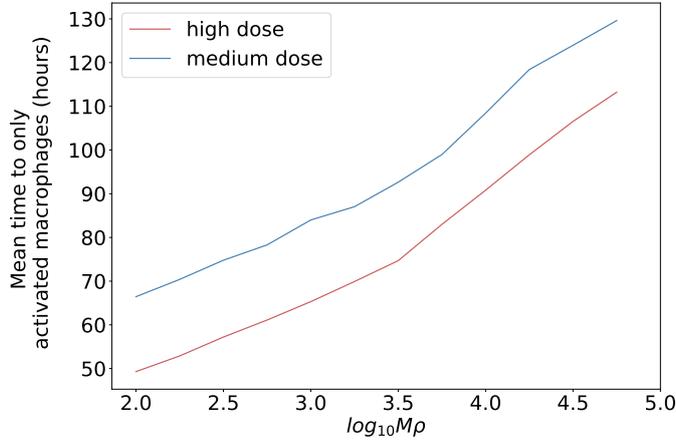
Figure 4.16: The mean time taken for ABM simulations to show that only activated macrophages are present within infected areas of the lung for varying values of $M\rho$.

cell^{-1} bacterium^{-1} h^{-1} remains fixed whilst M is varied. This results in a longer mean time when both $M\rho$ is larger, since there are more macrophages to infect, and when the initial dose is lower, since higher order rupture events are required to infect all cells.

4.5 Discussion

In this chapter, I have firstly provided a summary of results regarding birth processes and birth-and-death processes with catastrophe, with emphasis towards the time at which the catastrophe event occurs. As in Chapter 3, this process then acts as a model for the infection of a single macrophage with *F. tularensis* bacteria, however, now under the assumption that the rate at which an infected macrophage ruptures is proportional to its intracellular bacterial load. Previously, the constant catastrophe rate was chosen to reflect the experimentally observed log-normally distributed rupture times, whereas here, the choice reflects the belief that a macrophage containing a greater number of bacteria is more likely to rupture. Attempting to determine the distribution of rupture times is a preferable approach, with the choice of a log-normal distribution not motivated by biolog-

4. AN AGENT BASED MODEL FOR *F. TULARENSIS* INFECTION

ical mechanisms, only by goodness-of-fit. The density of the time until rupture, depicted in Figure 4.7, may not exactly match that of the $logN(3.72, 0.385)$ distribution used in Chapter 3, however, this is likely due to the assumption that both the birth and rupture rates are linear. Future consideration could therefore be given to investigating how the structure of these rates influences the shape of the density of the time until rupture. Despite this, through linking the replication and rupture rates, the approach used in this chapter has the advantage of being able to predict the effect that changing the rate of bacterial replication has on the time until rupture. In the within-phagocyte level of the multi-scale model, the distribution of rupture times is approximated from *in vitro* results, and is independent of the stochastic logistic growth process. For this model, it would therefore not be possible to make these same predictions.

To consider the interactions between populations of cells and *F. tularensis* bacteria, an ABM has also been developed, the behaviour of which can be approximated through applying results from the model of a single infected cell. These ODE approximations consider the change in successive cohorts of bacteria and are consistent with the ABM during the initial 48 hours of infection. An alternative approximation has since been formulated that offers greater accuracy for later time points, and is reported here to avoid confusion regarding the exact method used in the previous sensitivity analysis and parameter inference. Consider, for example, the second cohort of cytosol residing bacteria, $C_2(t)$. This mean number of bacteria can instead be expressed as

$$
\begin{aligned}
C_2(t) &= N \int_0^t \mathbb{E}\left[X(t-s) \mid X(0) = 1\right] \tilde{f}_1(s)\, ds, \\
&= N \int_0^t \frac{f_1(t-s)}{\delta} \tilde{f}_1(s)\, ds = \frac{N f_2(t)}{\delta}.
\end{aligned} \tag{4.17}
$$

where $f_2(t)$ is equivalent to the definition provided in (4.14). Although this new approach is more accurate at later time points, it is almost equivalent to the previous approximations during the initial 48 hours, as seen in Figure 4.17. Little difference could therefore be expected in the sensitivity analysis and Bayesian inference presented in Sections 4.4.2 and 4.4.4, and so the same conclusions hold
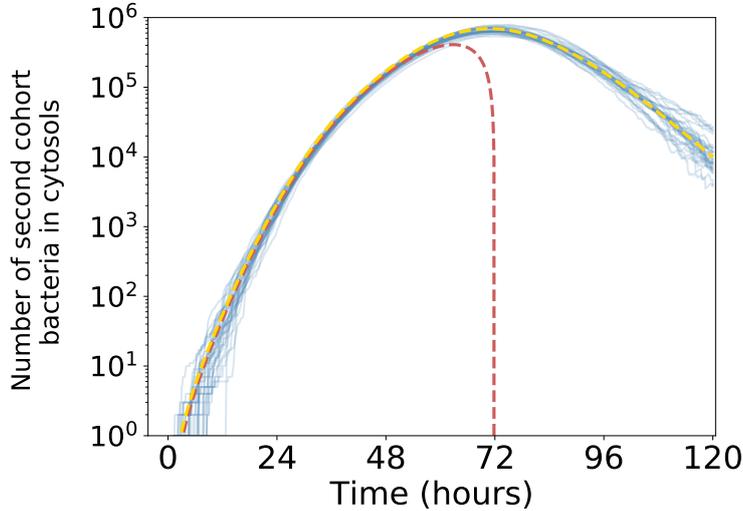
Figure 4.17: A comparison between the mean number of second cohort bacteria in macrophage cytosols computed using (4.15) (red), the new approach described in (4.17) (yellow), and 3×10^1 simulations of the ABM (blue).

true. These conclusions include obtaining an accurate estimate of the intracellular replication rate, β, of *F. tularensis* SCHU S4 bacteria. Given that, relative to the time spent in the cytosol, little time is spent extracellularly or in the phagosome, this parameter has also been shown to be the most important for determining how quickly the disease progresses. Despite providing an improvement, there will ultimately be a time where this new approximation fails due to multiple bacteria being phagocytosed by the same macrophage. At such a time, obtaining mathematical expressions for the size of each cohort becomes difficult, as the effect that subsequent phagocytosis events have on the time until rupture must be accounted for. This has previously been considered in Section 3.4.1 for the within-phagocyte model, but only numerically through the use of the Gillespie algorithm. Furthermore it may also be necessary to consider a spatial aspect, with the proximity of neighbouring macrophages affecting which cells the released bacteria enter. Together, these difficulties in obtaining analytical expressions highlight the benefits of agent based models, where simple sets of rules are able to produce complex behaviour.

Chapter 5

A within-host model for Ebola virus infection

Mathematical models have previously been used to great effect in understanding the infection dynamics of viruses and in aiding the design of novel experiments, with particular successes including models of HIV, hepatitis and influenza infection (Baccam *et al.* (2006); Perelson (2002); Ribeiro *et al.* (2002)). In many cases, these models are variations of an original target cell-limited model that considers the dynamics of uninfected target cells, infected cells and extracellular virus (Nowak & May (2000)). Extensions of this original model often then involve the introduction of innate and adaptive immune responses, or the effects that antiviral drug treatments have on the infectivity and replication of the virus (Best & Perelson (2018)). Although antiviral treatments can target specific intracellular viral components, their efficacy is accounted for in these models by adjusting the rate of virus production, a rate that is assumed to be proportional to the number of infected cells. As a result, these models focus very much on the extracellular dynamics of infection and less so on the intracellular dynamics.

Detailed models of the intracellular dynamics that incorporate transcription of the viral genome and translation of viral proteins have been considered by Srivastava (2002) and Sidorenko & Reichl (2004). However, the greater level of detail that is included in these models means that they are only used to study the in-

fection cycle of a single cell. There are therefore relatively few models that bridge both extracellular and intracellular dynamics. One approach, used by Haseltine *et al.* (2005) and Duvigneau *et al.* (2018) is to derive cell population balances from the equation of continuity. This approach is able to differentiate between infected cells within a population whilst also providing a multi-component description of each infected cell at an intracellular level. The benefit of this modelling framework has been observed when considering infection at a low MOI, such that not all target cells initially become infected. In this case, Haseltine *et al.* (2005) show how the original target cell-limited model is not able to correctly describe the multiple peaks in the number of infected cells that is associated with more than one round of infection events. Although this is an attractive approach, the incorporation of multiple interacting populations requires the estimation of many parameters, which can be difficult when limited experimental data are available.

An alternative approach that is considered by Koelle *et al.* (2018) looks at the number of virions that infect target cells rather than the exact intracellular viral load. In this target cell-limited 'macroparasite' model, heterogeneity between cells is accounted for by allowing cells to become infected at different cellular MOI. Simplifications then arise from assuming that rates of virus production and infected cell death are proportional to the cellular MOI. Although this results in a model whose structure is similar to that of traditional models of viral kinetics, it only describes the increase in the intracellular viral load due to internalisation. Since this increase is also due to replication of the viral genome, this type of model is not practical when comparing to time course data of intracellular viral loads.

In this chapter, the aim is to therefore develop a model for studying viral infections that can account for both extracellular and intracellular dynamics whilst minimising the number of additional parameters required to describe the intracellular stages. Despite being applicable to different viruses, the model is studied and parametrised with regards to Ebola virus (EBOV) infection. Following and during the most recent outbreak of EBOV in West Africa, efforts have focused on developing epidemiological models that can predict the spread of infection,

with few considering the within-host dynamics (Althaus (2014); Rachah & Torres (2015)). Exceptions to this include Madelain *et al.* (2015), who were the first to model the effect that the polymerase inhibitor favipiravir has on reducing viral loads in mice, and later in non-human primates (Madelain *et al.* (2018)). A similar model without treatment is described by Nguyen *et al.* (2015) and is parametrised using *in vitro* infection data. In particular, it was noted how reliable estimates of each model parameter could not be obtained from measurements of viral load alone due to issues that arise with identifiability. It is suggested that measurements of infected and uninfected cells could contribute to alleviating these problems, along with experiments that can determine the clearance rate of EBOV in the absence of target cells. For this reason, the model developed here is not only parametrised using measurements of extracellular viral load, but also intracellular viral load, cell death and virus decay.

This chapter begins by providing a detailed description of the mathematical model in Section 5.1, with particular emphasis towards the novel approach used to account for intracellular dynamics. Importantly, this model requires just a single additional parameter compared to an equivalent traditional model of viral dynamics and also includes a per virus budding rate that is more realistic from a biological perspective. In Section 5.2, Bayesian analysis is performed for each data set separately and all data sets simultaneously in order to gain a better understanding of the parameters that can be inferred from each. Posterior samples for each parameter obtained from data concerning the initial 72 hours of infection are then used to predict late time behaviour and can be compared to corresponding experimental measurements. Finally, the basic reproduction number, a measure of whether the infection will spread or be cleared, is derived for the model described in Section 5.1.

5.1 Wild-type model

In this section, I describe how a deterministic mathematical model may be developed to represent a series of *in vitro* experiments involving the infection of

target cells with EBOV. The model also adopts a stochastic approach to take into account the intracellular dynamics of an infected cell. However, since it is important that the model variables are comparable to the measured quantities, an understanding of the data provided is first required.

5.1.1 Experimental data

The data presented in this section and applied in the parameter inference that follows are provided by Dr. Thomas Laws and Dr. Sophie Smither from the Defence Science and Technology Laboratory (Dstl). These quantitative results consist of measurements of infectious and total viral load, viral decay in the absence of cells, and cell death. Where applicable, infection dynamics are studied in Vero cells, a cell line derived from kidney epithelial cells of African green monkeys. In each case, 10^5 Vero cells are infected at a MOI of 5, where a MOI of 1 would represents an average of one $TCID_{50}$ per cell (Iwami *et al.* (2012)).

In order to study the infectious viral load, the 50% tissue culture infectious dose ($TCID_{50}$) is determined at regular time points during the first 24 hours and also at 48 and 72 hours post infection. The $TCID_{50}$ assay is an endpoint assay that involves infecting cells at specific dilutions and determining, after a suitable period of time, the number of wells in which cytopathic effects can be observed. From this, it is possible to determine the dilution such that each cell has a 50% chance of becoming infected (Smither *et al.* (2013)). Since $TCID_{50}$ is the dose required to infect 50% of cells, it is likely that this corresponds to multiple copies of viral RNA. The recorded extracellular $TCID_{50}$ are provided in Table 5.1.

Real-time polymerase chain reaction (PCR) is a fast and popular approach for determining viral load, however, since it measures all genetic material present in a sample, it quantifies total viral load. PCR uses fluorescent dyes to detect viral genomes that are amplified in a cyclic manner such that the number of genomes approximately doubles with each cycle. The Ct value is then defined as the number of cycles necessary to achieve a predetermined threshold of fluorescence (Knipe *et al.* (2007)). Therefore, a larger Ct value corresponds to more cycles required to reach the threshold and thus a smaller number of viral genomes in

time (hours)	TCID$_{50}$	extracellular Ct value	intracellular Ct value
t_0	4.71×10^5	16.45	22.61
2	5.62×10^5	16.37	19.06
4	1.70×10^6	16.36	18.68
6	3.83×10^5	16.32	19.04
8	5.01×10^5	16.18	19.16
10	4.22×10^5	16.19	19.05
12	3.16×10^5	16.32	19.13
14	3.83×10^5	16.98	19.55
16	4.22×10^5	16.15	19.67
18	5.01×10^5	16.45	19.47
20	3.83×10^5	16.48	19.43
22	4.22×10^5	16.57	18.57
24	-	17.13	18.33
48	3.16×10^6	15.57	13.47
72	1.91×10^7	14.64	12.88

Table 5.1: Extracellular TCID$_{50}$ and Ct values along with intracellular Ct values during the initial 72 hours of infection with EBOV at an MOI of 5. The time $t_0 > 0$ denotes the time at which the first measurements were recorded.

the initial sample. Both extracellular and intracellular Ct values for the EBOV experiment are reported in Table 5.1.

In order to study the degradation of EBOV, TCID$_{50}$ and Ct values are also determined daily in the absence of target cells, such that no infection occurs. This experiment is performed in triplicate with the results presented in Table 5.2.

In order to quantify cell death, the dye trypan blue is used to identify the number of cells that have clear cytoplasms and are viable, and the number of non-viable cells whose cyotplasms are stained blue (Strober (2015)). The fraction of total cells that are non-viable is determined in duplicate daily between three and seven days post infection, with the results presented in Table 5.3.

time (hours)	TCID$_{50}$			Ct value		
	set 1	set 2	set 3	set 1	set 2	set 3
0	4.64×10^5	1.00×10^6	1.58×10^6	15.52	15.80	15.87
24	3.16×10^5	2.15×10^5	1.31×10^5	16.02	15.93	16.02
48	4.11×10^4	4.64×10^4	3.16×10^4	16.07	16.09	16.14
72	1.00×10^4	1.00×10^4	1.31×10^4	16.36	16.17	16.33

Table 5.2: TCID$_{50}$ and Ct values for the decay of virus in the absence of infection.

time (hours)	set 1	set 2	mean
72	0.229	0.156	0.192
96	0.345	0.238	0.291
120	0.415	0.303	0.359
144	0.325	0.303	0.314
168	0.330	0.286	0.308

Table 5.3: The proportion of dead cells between three and seven days, following infection with EBOV at an MOI of 5, as determined using trypan blue staining.

5.1.2 Extracellular dynamics

The description of the experimental data suggests that in order to fully utilise these results, the mathematical model should be able to specify extracellular and intracellular total viral loads, extracellular infectious viral loads and the proportion of dead cells at each observed time point. In this sense, a target cell-limited model can act as a starting point for studying the infection dynamics. Initially, target cells are infected by free infectious virus according to mass-action kinetics. From the extracellular TCID$_{50}$ and Ct values in Table 5.1, it can be seen that the extracellular viral load does not begin to increase until approximately 24 hours. This period of time, where viral replication is occurring within the cell but release of virus is yet to occur, is referred to as the eclipse phase. This relates to the period where synthesis of viral proteins and replication of the viral genome are

ongoing within the cell, but are yet to yield sufficient numbers of each such that viral assembly and budding can occur (see Section 1.1.2). To incorporate this into the model, it is assumed that the population of infected cells is partitioned into eclipse phase cells and infectious phase cells. When target cells become infected, they enter an eclipse phase state before transitioning into infectious cells after an average time of τ_E hours. Infectious phase cells then remain in this state for a period of τ_I hours, during which the release of virus occurs. This released virus is assumed to be infectious with probability ε and non-infectious with probability $1 - \varepsilon$, representing the presence of lethal mutations during the replication of the viral genome. A depiction of this model is presented in Figure 5.1, for which the model variables may be defined as

$T(t) = $ "*mean number of target cells at time t*",

$E_i(t) = $ "*mean number of stage i eclipse phase cells at time t*",

$I_j(t) = $ "*mean number of stage j infectious phase cells at time t*",

$V_I(t) = $ "*mean number of extracellular infectious virus at time t*",

$V_{NI}(t) = $ "*mean number of extracellular non-infectious virus at time t*",

and obey the system of ODEs

$$\frac{d}{dt}T(t) = -\beta_c V_I(t)T(t),$$

$$\frac{d}{dt}E_1(t) = \beta_c V_I(t)T(t) - \frac{n_E}{\tau_E}E_1(t),$$

$$\frac{d}{dt}E_i(t) = \frac{n_E}{\tau_E}\left(E_{i-1}(t) - E_i(t)\right), \qquad i = 2,...,n_E,$$

$$\frac{d}{dt}I_1(t) = \frac{n_E}{\tau_E}E_{n_E}(t) - \frac{n_I}{\tau_I}I_1(t),$$

$$\frac{d}{dt}I_j(t) = \frac{n_I}{\tau_I}\left(I_{j-1}(t) - I_j(t)\right), \qquad j = 2,...,n_I,$$

$$\frac{d}{dt}V_I(t) = -\beta_v V_I(t)T(t) - \sum_{i=1}^{n_C}\beta_v V_I(t)E_i(t) + \varepsilon b v_I(t) - (\mu_V + \mu_S)V_I(t),$$

$$\frac{d}{dt}V_{NI}(t) = (1 - \varepsilon)b v_I(t) - \mu_V V_{NI}(t) + \mu_S V_I(t).$$
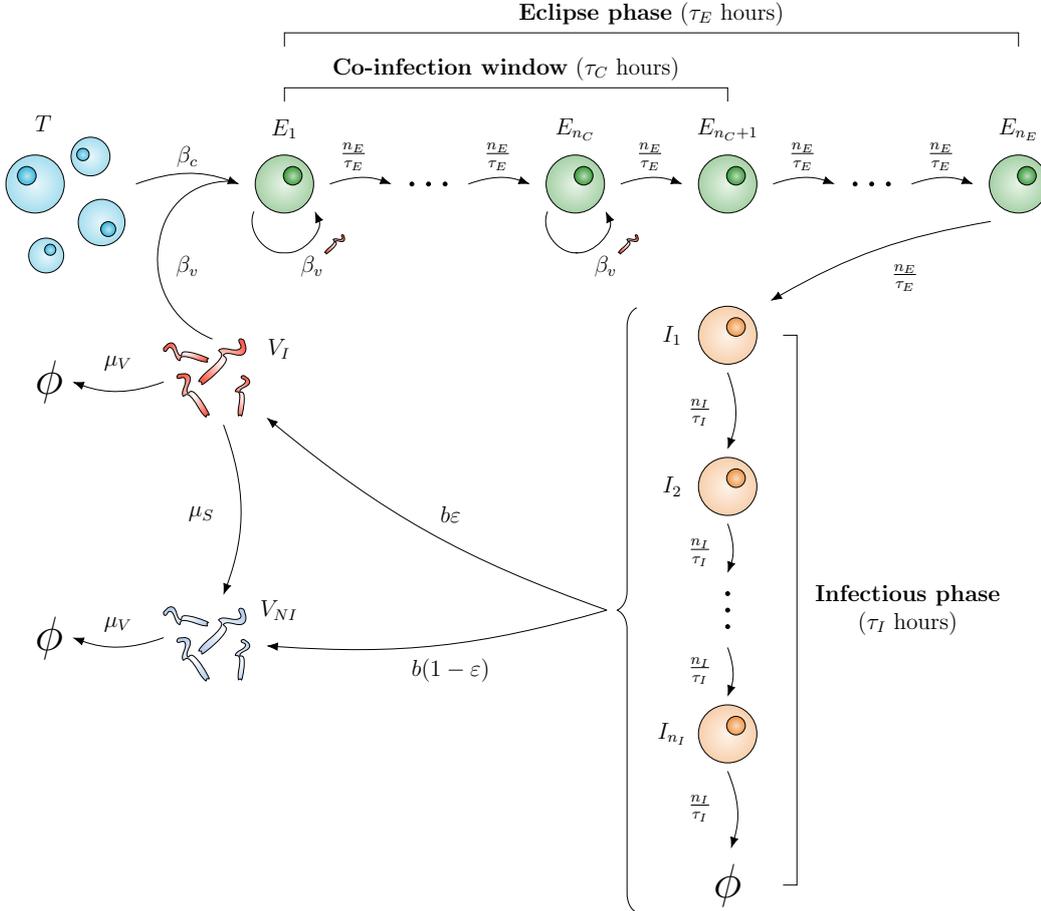
Figure 5.1: A depiction of the wild type model for EBOV infection. Target cells (T) are infected by infectious EBOV and enter an eclipse phase (E_i). During the co-infection window, represented by the first n_C states of the eclipse phase, infectious EBOV can further infect eclipse cells but this does not alter the state of the cell. Following the eclipse phase, infected cells enter an infectious phase (I_j) where production of virus occurs at rate $b\,h^{-1}$, proportional to the amount of intracellular virus (v_I). The released virus is either infectious (V_I) with probability ε, or non-infectious (V_{NI}) with probability $1-\varepsilon$. Degradation of virus occurs with rate $\mu_V\,h^{-1}$ whilst infectious virus loses infectivity with rate $\mu_S\,h^{-1}$.

Here, the eclipse phase and infectious phases are further partitioned into a series of states. This allows for the possibility that the time an infected cell spends in these phases is not exponentially distributed, but instead follows an Erlang distribution. As the number of states in the partition increases, the distribution of the time spent in each phase becomes narrower until the duration is assumed to be constant for all cells (Best & Perelson (2018)). This approach has also been applied in models of HIV and influenza virus (Mittler *et al.* (1998); Pinilla *et al.* (2012)).

When considering terms representing the infection of target cells, the loss of virus has also been included in the ODE for infectious virus. This term is often omitted under the assumption that the number of virions is large enough that any loss due to infection is negligible, but is included here for completeness. Furthermore, different infection rates β_c $(h^{-1}(\text{RNA})^{-1})$ and β_v $(h^{-1}(\text{cells})^{-1})$ have been used to reflect that $\beta_c V_I T$ is in units of cells when describing the change in T and E_1, but in units of virus when describing the change in V_I (Yuan & Allen (2011)).

This model also accounts for the possibility that a single cell becomes co-infected with multiple virus in a manner that further reduces the infectious viral load but does not alter the state of the cell. The time period during which co-infection is possible is referred to as the co-infection window and has length $\tau_C \, h$, although in the parameter inference that follows this will instead be defined by the ratio τ_C/τ_E. In this way, the co-infection window comprises the initial $n_C = n_E \tau_C / \tau_E$ states of the eclipse phase.

In addition to infection events, loss of virus also occurs due to viral degradation with rate $\mu_V \, h^{-1}$ and affects both infectious and non-infectious virus, whilst infectious virus can become non-infectious with rate $\mu_S \, h^{-1}$. Production of virus occurs at rate $b \, h^{-1}$ and is proportional to $v_I(t)$, the average amount of intracellular virus in all infectious phase cells. An equation for $v_I(t)$ is derived in Section 5.1.3 when the intracellular dynamics are taken into account. However, it can be noted now that unlike traditional models of viral dynamics, the production of virus is not proportional to the number of infected cells. This is an unrealistic

assumption as it suggests that production of virus occurs at an equal rate from cells just entering the infectious phase and cells that are close to dying. Instead, by considering that virus production is proportional to the amount of intracellular virus, this can reflect the increased production, as viral proteins accumulate in the cytoplasm. Non-constant production rates have previously been considered in age-structured models, where the production rate is a function of how long the cell has been infected for (Nelson *et al.* (2004)).

Currently, both $V_I(t)$ and $V_{NI}(t)$ are defined in units of RNA copies/ml and therefore cannot be compared with the experimentally measured TCID$_{50}$ and Ct values. However, if α (TCID$_{50}$/RNA) is a constant conversion factor from infectious viral RNA copies to TCID$_{50}$, then the infectious viral load in units of TCID$_{50}$/ml is $v_{50}(t) = \alpha V_I(t)$. Similarly, since Ct values are related to a measure of total viral load, the transformation $v_T(t) = V_I(t) + V_{NI}(t)$ can be used, where $v_T(t)$ now denotes the total viral load in units of RNA copies/ml (Iwami *et al.* (2012)). Ct values are usually converted into RNA copies by means of a standard curve that is determined by performing PCR on serial dilutions of the sample. Given that the number of RNA copies approximately doubles with each cycle, and the Ct value represents a number of cycles to reach a fluorescence threshold, the standard curve has the general form

$$\text{number of RNA copies at time } t = \kappa 2^{-C(t)} \,,$$

where $C(t)$ is the Ct value at time t. From here, it can be seen that when normalising with respect to the number of RNA copies at a fixed time point, the result is independent of the constant κ and depends only on the difference between Ct values. For this reason, the model variable $v_T(t)$ is normalised with respect to $v_T(0)$, and similarly $v_{50}(t)$ is normalised with respect to $v_{50}(0)$. If $w_{50}(t) = v_{50}(t)/v_{50}(0)$ and $w_T(t) = v_T(t)/v_T(0)$ are the normalised variables, the model can now be described by the system of ODEs

$$\frac{d}{dt}T(t) = -\frac{\beta_c v_{50}(0)}{\alpha} w_{50}(t) T(t) \,,$$
$$\frac{d}{dt}E_1(t) = \frac{\beta_c v_{50}(0)}{\alpha} w_{50}(t) T(t) - \frac{n_E}{\tau_E} E_1(t) \,,$$

$$\frac{d}{dt}E_i(t) = \frac{n_E}{\tau_E}\left(E_{i-1}(t) - E_i(t)\right), \qquad i = 2, ..., n_E,$$

$$\frac{d}{dt}I_1(t) = \frac{n_E}{\tau_E}E_{n_E}(t) - \frac{n_I}{\tau_I}I_1(t),$$

$$\frac{d}{dt}I_j(t) = \frac{n_I}{\tau_I}\left(I_{j-1}(t) - I_j(t)\right), \qquad j = 2, ..., n_I,$$

$$\frac{d}{dt}w_{50}(t) = -\beta_v w_{50}(t)T(t) - \sum_{i=1}^{n_C}\beta_v w_{50}(t)E_i(t) + \frac{\alpha\varepsilon b}{v_{50}(0)}v_I(t) - (\mu_V + \mu_S)w_{50}(t),$$

$$\frac{d}{dt}w_T(t) = -\frac{\beta_v v_{50}(0)}{\alpha v_T(0)}w_{50}(t)T(t) - \sum_{i=1}^{n_C}\frac{\beta_v v_{50}(0)}{\alpha v_T(0)}w_{50}(t)E_i(t) + \frac{b}{v_T(0)}v_I(t) - \mu_V w_T(t).$$

$$(5.1)$$

Initially, $w_{50}(0) = w_T(0) = 1$ and the number of target cells is $T(0) = 10^5$ cells. Since there are no infected cells to begin with, all remaining variables are zero. This system of ODEs currently describes the dynamics of target cells, infected cells and extracellular virus; in order to determine the production rate of virus it is now necessary to consider the intracellular viral load of an infected cell.

5.1.3 Intracellular dynamics

To study the intracellular dynamics, the following assumptions are made regarding the different phases that an infected cell transitions through.

- **Co-infection window**: For cells in the co-infection window, infectious virus can enter already infected cells at the same rate, but significant viral replication and protein synthesis is yet to occur within these infected cells.

- **Eclipse phase**: Following the end of the co-infection window until the end of the eclipse phase, viral replication occurs within infected cells at rate $\lambda\,h^{-1}$ per virus.

- **Infectious phase**: For cells in the infectious phase, viral replication is still ongoing, however, a decrease in intracellular viral load occurs due to the assembly and budding of virus at rate $b\,h^{-1}$ per virus from the cell membrane.

For the purpose of studying the intracellular dynamics, let the eclipse phase denote the period between the end of the co-infection window and the start of the infectious phase. In this way, the intracellular dynamics of the three separate phases can be thought of as three distinct processes. For cells in the co-infection phase, the constant increase in viral load associated with infection events may be thought of as a Poisson process. Following the co-infection window, the replication of the viral genome is naturally described by a birth process where the rate of replication is proportional to the number of genomes. The production of virus that occurs from infectious phase cells may be interpreted as the loss of intracellular virus and is therefore analogous to the 'death' of virus. Since viral replication continues within cells in the infectious phase, the intracellular dynamics during this phase can be modelled as a birth-and-death process.

In order to compare to intracellular Ct values, the total intracellular viral load is required as a function of time, however, under these assumptions this can be considered as the sum of the viral loads across all cells in each phase. The intracellular model variables may therefore be defined as

$v_C(t) = $ "*mean number of intracellular virus in co-infection phase cells at time t*",

$v_E(t) = $ "*mean number of intracellular virus in eclipse phase cells at time t*",

$v_I(t) = $ "*mean number of intracellular virus in infectious phase cells at time t*",

and the total intracellular viral load is $v_{int}(t) = v_C(t) + v_E(t) + v_I(t)$. The difficulty with this approach is then in determining how each variable increases and decreases as cells transition between phases. For example, given a cell in state E_{n_E}, the final state of the eclipse phase, the intracellular viral load of the cell contributes to the total $v_E(t)$. However, when the cell transitions into state I_1, the intracellular viral load must now contribute to the total $v_I(t)$. To overcome this, the distribution of time until a single cell exits each phase can be considered.

Let $t = 0$ represent the start of the experiment with all cells initially uninfected. Provided that there is initially $v_{50}(0)$ TCID$_{50}$ of infectious virus, the time until a single cell becomes infected is an exponential random variable with rate $\beta_c v_{50}(0)/\alpha$. Once infected, this cell transitions through the first n_C states of the

eclipse phase until the end of the co-infection window. Since the time spent in each state is an exponential random variable with rate n_E/τ_E, the time spent between infection and the end of the co-infection window follows an Erlang(n_E/τ_E, n_C) distribution. The total time from $t = 0$ until the end of the co-infection window is then the sum of these exponential and Erlang random variables, the density of which is obtained using the convolution of the individual densities. Suppose that $f_C(t)$ denotes the time at which a single cell, initially uninfected, exits the co-infection window,

$$f_C(t) = \int_0^t \beta_{50} v_{50}(0) e^{-(\beta_{50} v_{50}(0))(t-s)} \frac{\delta_E^{n_C} s^{n_C-1} e^{-\delta_E s}}{(n_C - 1)!} \, ds$$

$$= \left(\frac{\delta_E}{\delta_E - \beta_{50} v_{50}(0)} \right)^{n_C} \beta_{50} v_{50}(0) e^{-\beta_{50} v_{50}(0)t} F(t; \delta_E - \beta_{50} v_{50}(0), n_C) \,,$$

where $\beta_{50} = \beta_c/\alpha$, $\delta_E = n_E/\tau_E$ and $F(t; a, k)$ is the distribution function of an Erlang(a, k) random variable,

$$F(t; a, k) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-at} (at)^n \,, \qquad a, t \geq 0 \,.$$

As the density must be non-negative, the above expression is only true when $\delta_E > \beta_{50} v_{50}(0)$. When $\delta_E \leq \beta_{50} v_{50}(0)$, this suggests that the initial infection of a cell occurs at a faster rate than the rate at which an infected cell transitions through states in the co-infection window. As a result of this, it is assumed that the contribution from the time until the first infection event can be neglected from the total time to exit the co-infection window. In doing so, this total time is now equal to the time taken for a cell to transition from state E_1 to state E_{n_C+1}. Since this time is an Erlang(δ_E, n_C) random variable, $f_C(t)$ is equal to the corresponding density when $\delta_E \leq \beta_{50} v_{50}(0)$. The full definition of $f_C(t)$ may then be given by

$$f_C(t) = \begin{cases} \left(\dfrac{\delta_E}{\delta_E - \beta_{50} v_{50}(0)} \right)^{n_C} \beta_{50} v_{50}(0) e^{-\beta_{50} v_{50}(0)t} F(t; \delta_E - \beta_{50} v_{50}(0), n_C) & \\ & \delta_E > \beta_{50} v_{50}(0) \,, \\[2ex] \dfrac{\delta_E^{n_C} t^{n_C-1} e^{-\delta_E t}}{(n_C - 1)!} & \\ & \delta_E \leq \beta_{50} v_{50}(0) \,. \end{cases}$$
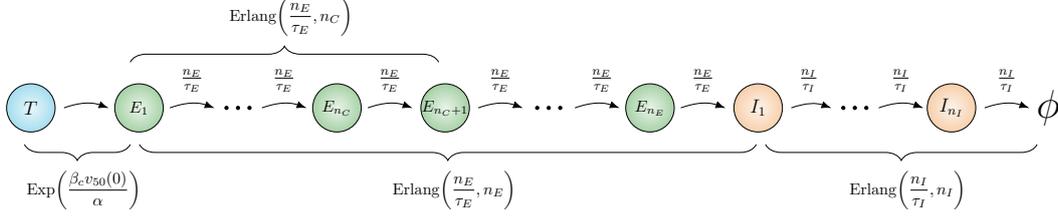
Figure 5.2: A depiction of the life cycle of a single cell from infection to the end of the infectious phase, including the distribution of the time spent in each of the phases in between.

Since the co-infection window represents a subset of states of the whole eclipse phase, the density for the time until a cell exits the eclipse phase can be found in a similar way. This time is again a sum of exponential random variables, each with rate δ_E. The difference now is that the number of variables that are summed is n_E, the total number of states in the co-infection window and eclipse phase. If $f_E(t)$ denotes the density of the time for a cell to exit the eclipse phase,

$$
f_E(t) = \begin{cases}
\left(\dfrac{\delta_E}{\delta_E - \beta_{50}v_{50}(0)}\right)^{n_E} \beta_{50}v_{50}(0)e^{-\beta_{50}v_{50}(0)t}F(t; \delta_E - \beta_{50}v_{50}(0), n_E) \\
\hspace{8cm} \delta_E > \beta_{50}v_{50}(0)\,, \\[2mm]
\dfrac{\delta_E^{n_E}t^{n_E-1}e^{-\delta_E t}}{(n_E - 1)!} \hspace{5cm} \delta_E \leq \beta_{50}v_{50}(0)\,.
\end{cases}
$$

Finally, let $f_I(t)$ be the density of the time for an initially uninfected cell to exit the infectious phase and stop producing virus. This can be described as the sum of the time to exit the eclipse phase and the time spent in the infectious phase, the latter of which is an Erlang(δ_I, n_I) distributed random variable, where $\delta_I = n_I/\tau_I$. As a result of this

$$
f_I(t) = \int_0^t f_E(s)\frac{\delta_I^{n_I}(t-s)^{n_I-1}e^{-\delta_I(t-s)}}{(n_I - 1)!}\,ds\,. \tag{5.2}
$$

With $f_C(t)$, $f_E(t)$ and $f_I(t)$ it is now possible to determine the exchange in the contribution to the intracellular viral load that occurs as cells transition between

phases. Consider, for example, the case where cells transition from the eclipse phase to the infectious phase. The mean number of cells that make this transition at time t is $T(0)f_E(t)$, and the average viral load of each cell may be thought of as the ratio of the total viral load across all cells in the eclipse phase, $v_E(t)$, and the number of cells in the eclipse phase, $\sum_{i=n_C+1}^{n_E} E_i(t)$. By applying this reasoning to each phase, the intracellular viral loads satisfy the ODEs

$$\frac{d}{dt}v_C(t) = \frac{\beta_v v_{50}(0)}{\alpha}w_{50}(t)T(t) + \frac{\beta_v v_{50}(0)}{\alpha}w_{50}(t)\sum_{i=1}^{n_C}E_i(t) - \frac{v_C(t)T(0)f_C(t)}{\sum_{i=1}^{n_C}E_i(t) + \gamma},$$

$$\frac{d}{dt}v_E(t) = \frac{v_C(t)T(0)f_C(t)}{\sum_{i=1}^{n_C}E_i(t) + \gamma} + \lambda v_E(t) - \frac{v_E(t)T(0)f_E(t)}{\sum_{i=n_C+1}^{n_E}E_i(t) + \gamma},$$

$$\frac{d}{dt}v_I(t) = \frac{v_E(t)T(0)f_E(t)}{\sum_{i=n_C+1}^{n_E}E_i(t) + \gamma} + (\lambda - b)v_I(t) - \frac{v_I(t)T(0)f_I(t)}{\sum_{j=1}^{n_I}I_j(t) + \gamma},$$

$$(5.3)$$

where $v_C(0) = v_E(0) = v_I(0) = 0$ and $\gamma \ll 1$ ensures that the denominator is never zero when there are initially no infected cells, but does not interfere with the dynamics. The terms $\lambda v_E(t)$ and $(\lambda - b)v_I(t)$ reflect the growth of the intracellular population of virus as described by the birth process and birth-and-death process. Here it is assumed that the rate of viral replication is linearly proportional to the number of viral genomes, however, this could be changed to represent the limitation of intracellular resources or to account for antiviral treatment. In this way, the effects of antiviral therapies that target viral replication can be explicitly represented, instead of the commonly used approach where a reduction in the per cell production rate corresponds to the efficacy of the treatment (Best & Perelson (2018)).

In the parameter inference that follows it is possible for there to be no co-infection window, that is, each cell is only ever infected by a single virus. In this case, $v_C(t)$ is no longer required in the model and $v_E(t)$ instead satisfies

$$\frac{d}{dt}v_E(t) = \frac{\beta_v v_{50}(0)}{\alpha}w_{50}(t)T(t) + \lambda v_E(t) - \frac{v_E(t)T(0)f_E(t)}{\sum_{i=n_C+1}^{n_E}E_i(t) + \gamma}.$$

The most attractive feature of this model is that it only requires a single additional parameter, λ, compared to equivalent traditional models for extracellular

dynamics, yet allows for additional data, in the form of intracellular measurements, to be used to help infer the value of each parameter.

5.1.4 Cell death

With the current description of the model, it is possible to compare with experimental measurements of intracellular and extracellular viral load. In order to compare with the measurements of cell death, let $D(t)$ denote the mean fraction of cells that have died before time t given an initial number of cells, $T(0)$. When a cell exits state I_{n_I}, this represents that production of virus no longer occurs, but this may not always coincide with the death of the cell. Instead, it can be assumed that a fraction, f, of cells that stop producing virus do so because of cell death. In this way, $D(t)$ satisfies the ODE

$$\frac{d}{dt}D(t) = \frac{f}{T(0)}\frac{n_I}{\tau_I}I_{n_I}(t)\,, \qquad\qquad D(0) = 0\,. \qquad\qquad (5.4)$$

Together with systems (5.1) and (5.3), equation (5.4) fully defines the mathematical model of EBOV infection. However, when interested in only the cell death dynamics, the full solution of the model is still required. An alternative approach for obtaining $D(t)$ can be described by utilising the densities for the time taken for a cell to exit each phase. Since $f_I(t)$ is the probability that a cell exits the infectious phase at time t, it is also true that

$$D(t) = f\int_0^t f_I(s)\,ds\,. \qquad\qquad (5.5)$$

5.2 Parameter inference

In this section, Bayesian inference is used to determine how much can be learned about each model parameter given individual data sets for extracellular and intracellular viral loads and cell death. Before this, global sensitivity analysis is used to identify the most important parameters with respect to those model variables for which experimental measurements have been recorded. A nonlinear least squares

approach is also used to fix the rates of viral degradation and denaturation given measurements of virus decay (5.2).

5.2.1 Sensitivity analysis

To determine the most influential parameters, the Sobol approach to global sensitivity analysis, as described in Section 2.6, has been performed. First-order (S_1) and total-order (S_T) sensitivity indices have been evaluated for each of the model variables $\log_{10} v_{50}(t)$, $\log_{10} v_T(t)$, $\log_{10} v_{int}(t)$ and $D(t)$, recalling that $v_{int}(t)$ is the total viral load of all infected cells. For $\log_{10} v_{50}(t)$, $\log_{10} v_T(t)$ and $\log_{10} v_{int}(t)$, both indices are evaluated across the initial 72 hours of infection, whilst indices for $D(t)$ are evaluated for the first 168 hours. This choice reflects the period of time over which corresponding experimental measurements are recorded. Figures 5.3 and 5.4 show these sensitivity indices for each model parameter given fixed initial conditions of $v_{50}(0) = 5 \times 10^5$ and $v_T(0) = 10^9$, suitable values given an MOI of 5 and $T(0) = 10^5$ cells. The ranges over which each parameter is varied correspond to the ranges of the prior distributions provided in Table 5.5. For a more clear visualisation of the most important parameters over the course of infection, the mean and standard deviation of S_T for each parameter and model variable are also reported in Table 5.4.

For $v_{50}(t)$, the amount of extracellular infectious virus, β_v is the most important parameter, particularly during the early stages of infection, as this dictates the rate at which infectious virus infects target cells. The length of the eclipse phase is also important since a longer eclipse phase prolongs the time until virus production can occur. Similarly, the length of the infectious phase, τ_I, is increasingly important as the infection progresses. If the length of both the eclipse and infectious phases are short, viral production cannot be sustained, resulting in the decay of the population of extracellular virus. The rate at which viral genomes are replicated, λ, also increases in importance with time. This can be expected since the production of virus is now proportional to the intracellular viral load, such that a slower rate of replication may result in reduced viral production.

For $v_T(t)$, the total amount of extracellular virus, the most influential pa-
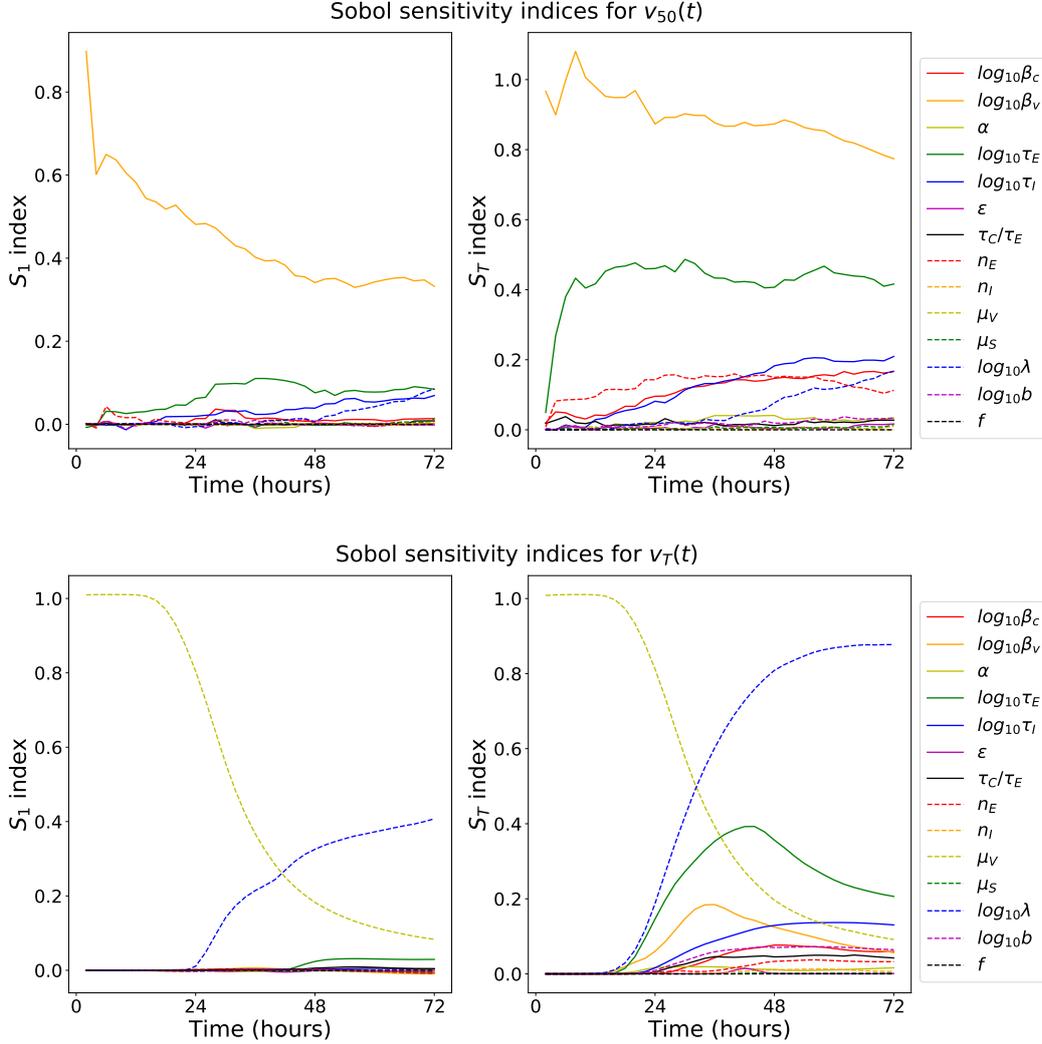
Figure 5.3: First-order (S_1) and total-order (S_T) Sobol sensitivity indices for each model parameter with respect to the variables $\log_{10} v_{50}(t)$ (top) and $\log_{10} v_T(t)$ (bottom).

rameter during the early stages of infection is μ_V, the rate of viral degradation. This is understandable since $v_T(t)$ also accounts for non-infectious virus which is unable to infect cells and is therefore not directly involved in the infection dynamics. Given that the initial number of infectious virus particles is $v_{50}(0)/\alpha$, for larger values of α the pool of infectious virus represents a small fraction of the $v_T(0) = 10^9$ total virus particles. In this case, the majority of total virus is then
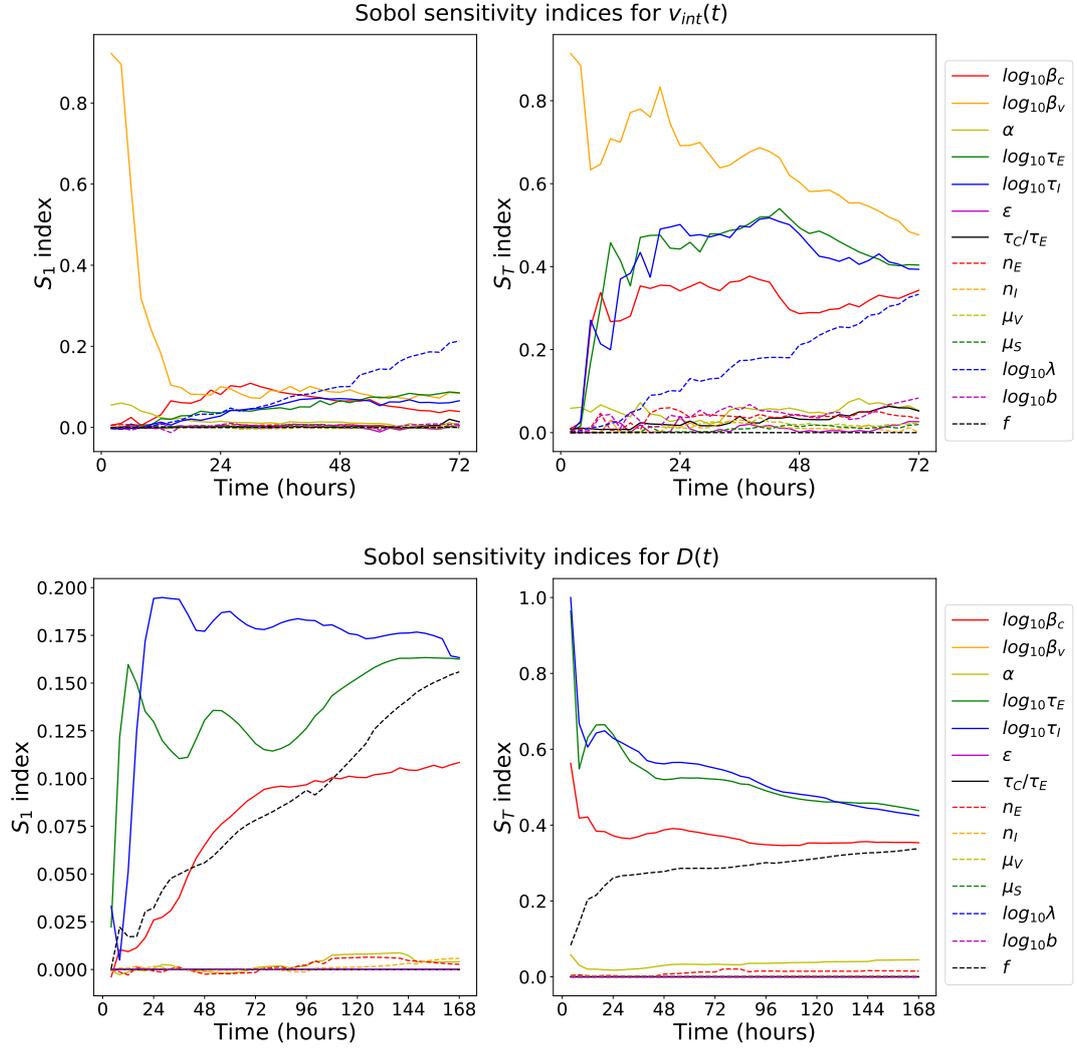
Figure 5.4: First-order (S_1) and total-order (S_T) Sobol sensitivity indices for each model parameter with respect to the variables $\log_{10} v_{int}(t)$ (top) and $D(t)$ (bottom).

non-infectious and can only degrade, so that changes in μ_V have a large effect on the total. Later on in the infection, λ, τ_E, β_v and τ_I become more important as virus production begins to affect the population size.

For $v_{int}(t)$, β_v is important for the same reasons that it is for $v_{50}(t)$, in that it determines the rate at which infectious virus infects target cells. The lengths of the eclipse and infectious phase also have high sensitivity indices as they define

the times at which the underlying birth and birth-and-death process start and end. It is these processes that determine how quickly the intracellular population grows. Furthermore, since the changes that occur in $v_C(t)$, $v_E(t)$ and $v_I(t)$ as cells transition between phases do not effect the total intracellular viral load, $v_{int}(t)$ grows exponentially following the end of the co-infection window. As a result, it is understandable that λ becomes increasingly important.

Finally, for $D(t)$ it should be noted that the sensitivity indices for a number of parameters are exactly zero. The reason for this is more clearly seen in equation (5.5) and the parameters that $f_I(t)$ depends on. Since this density depends only on the amount of virus through the initial condition, $v_{50}(0)$, parameters such as μ_V, μ_S, λ and b that relate solely to viral quantities have no effect on $D(t)$. The parameters that have the most influence on $D(t)$ are τ_E and τ_I as their sum specifies the average number of hours until an infected cell stops producing virus. Since the fraction of cells that die after virus production stops is determined by f, this also has a high sensitivity index. The rate at which target cells initially become infected depends on the per cell infection rate β_c. When this is small, the infection of cells does not occur as rapidly and the delay that this causes persists through the eclipse and infectious phases, resulting in a smaller fraction of cells dying at earlier times. For this reason, β_c is also important when describing cell death.

Collectively, by performing the sensitivity analysis for the four model variables of interest, β_v, β_c, τ_E, τ_I, μ_V, λ and f have been identified as the most important parameters. Although it is preferable to learn about all parameters, the hope is that in the inference that follows, it is possible to at least learn more about these specific parameters.

5.2.2 Viral decay

Nguyen *et al.* (2015) have previously suggested that measurements of viral decay in the absence of infection could greatly enhance the ability to successfully parametrise models of viral dynamics. With no target cells to infect, there is no subsequent production of virus, and thus the size of the viral populations only

	$\log_{10} v_{50}$		$\log_{10} v_T$		$\log_{10} v_{int}$		D	
	S_T mean	S_T s.d.	S_T mean	S_T s.d.	S_T mean	S_T s.d.	S_T mean	S_T s.d.
β_c	0.11	0.05	0.04	0.03	0.31	0.08	0.37	0.04
β_v	0.89	0.07	0.08	0.06	0.65	0.11	$< 10^{-2}$	$< 10^{-2}$
α	0.02	0.01	0.01	$< 10^{-2}$	0.05	0.02	0.03	0.01
τ_E	0.42	0.07	0.20	0.14	0.42	0.12	0.52	0.09
τ_I	0.12	0.07	0.07	0.06	0.41	0.12	0.53	0.10
ε	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
τ_C/τ_E	0.02	$< 10^{-2}$	0.03	0.02	0.03	0.02	$< 10^{-2}$	$< 10^{-2}$
n_E	0.13	0.03	0.02	0.02	0.04	0.01	0.01	$< 10^{-2}$
n_I	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	0.01	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
μ_V	$< 10^{-2}$	$< 10^{-2}$	0.50	0.37	0.02	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
μ_S	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$
λ	0.06	0.05	0.50	0.37	0.16	0.10	$< 10^{-2}$	$< 10^{-2}$
b	0.02	0.01	0.04	0.03	0.04	0.02	$< 10^{-2}$	$< 10^{-2}$
f	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	0.29	0.05

Table 5.4: Mean and standard deviation of the total-order Sobol sensitivity indices (S_T) for each model parameter and variable. These summaries are evaluated over the time periods shown in Figures 5.3 and 5.4.

decrease. Therefore, if $w_{50}(t)$ denotes the infectious viral load at time t, normalised with respect to the initial condition $v_{50}(0)$, and $w_T(t)$ is the normalised total viral load, then both variables decay exponentially according to

$$w_{50}(t) = e^{-(\mu_V + \mu_S)t}, \qquad w_T(t) = e^{-\mu_V t}.$$

If only measurements of infectious viral load were available, only the sum $\mu_V + \mu_S$ would be identifiable. Here, however, measurements of viral decay have been provided in both units of $TCID_{50}$ and Ct values (Table 5.2) and it is therefore possible to estimate μ_V and μ_S. By considering $\tilde{w}_{50}(t) = \log w_{50}(t)$ and $\tilde{w}_T(t) = \log w_T(t)$, the estimation problem reduces to finding the gradient of a straight line. For this reason, a Bayesian approach is not necessary and a least-squares approach is favoured instead. By regressing $\tilde{w}_{50}(t)$ and $\tilde{w}_T(t)$ on t, the least-squares estimates can be obtained using the statistical software R to be $\mu_V = 5.73 \times 10^{-3} \, h^{-1}$ and
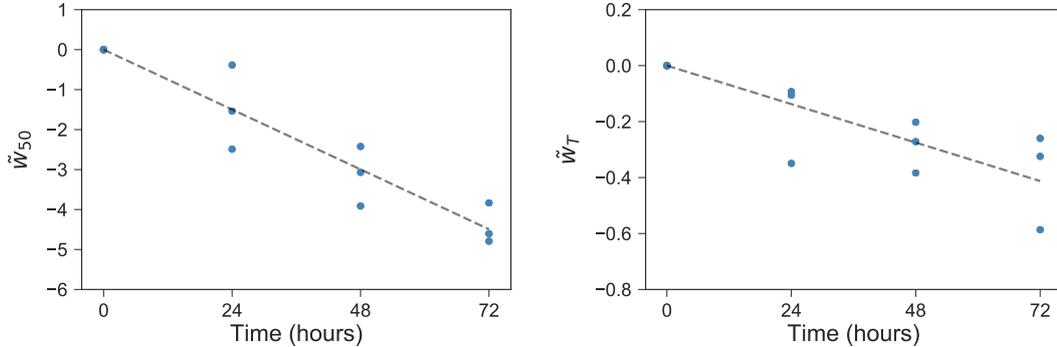
Figure 5.5: Fitted curves showing the decay of infectious virus $(\tilde{w}_{50})$ and total virus $(\tilde{w}_T)$ using the least-squares estimates $\mu_V = 5.73 \times 10^{-3}\, h^{-1}$ and $\mu_S = 5.67 \times 10^{-2}\, h^{-1}$.

$\mu_S = 5.67 \times 10^{-2}\, h^{-1}$. The respective p-values for these estimates are 9.6×10^{-2} and 1.32×10^{-9}, whilst the fitted lines are provided in Figure 5.5 alongside the observed values. To estimate μ_V and μ_S, the ratio of the total viral load has been expressed in terms of the experimentally measured Ct values using the expression $w_T(t) = 2^{C(0)-C(t)}$.

In the Bayesian inference that follows, the values of μ_V and μ_S are fixed as these least-squares estimates, thereby reducing the dimension of the parameter space. Furthermore, the sensitivity analysis identified μ_V as an influential parameter, particularly when considering $v_T(t)$, and so it is beneficial that an accurate estimate has been obtained.

5.2.3 Bayesian inference

With least-squares estimates for μ_V and μ_S, the aim is to now use the experimental data described in Section 5.1.1 in order to infer values for the remaining parameters. An ABC rejection sampling algorithm can be applied for each data set individually and all data sets simultaneously to determine how different measurements contribute to the learning of each parameter. The prior distributions for each parameter are reported in Table 5.5. Uniform distributions are selected

to reflect the prior beliefs that little is known about each parameter, with the range of each distribution chosen to be large enough such that values with high posterior density are not omitted. For the parameters ε, τ_C/τ_E and f, the range of the prior distribution reflects that these parameters represent fractions, whilst the range of the prior distribution for α is based on the assumption that the number of infectious virus particles that infect a cell follows a Poisson distribution (Iwami *et al.* (2012)). For t_0, the prior distribution reflects the uncertainty in the time of the first measurement, but the belief that it was conducted within 30 minutes of the start. To ensure even sampling for parameters whose ranges span multiple orders of magnitude, the corresponding prior distributions have also instead been specified for the logarithm of the parameter.

The model variables $w_{50}(t)$, $w_T(t)$ and $D(t)$ have previously been defined so that they are comparable to the experimentally observed quantities. For $v_{int}(t)$, it is not possible to normalise with respect to the initial number of intracellular copies of viral RNA, since this is zero. Instead, $v_{int}(t)$ can be normalised with respect to the number at t_0, the time of the first measurement, and thus $w_{int}(t) = v_{int}(t)/v_{int}(t_0)$ is now comparable to the intracellular Ct values in a similar way to which $w_T(t)$ is comparable to extracellular Ct values.

When deciding on a measure of distance between the model output and experimental observations, it is important to consider the weight that is assigned to each point and how this affects the total distance. For example, from Table 5.1 it can be noted that the extracellular Ct and TCID$_{50}$ values remain approximately constant between 2 and 24 hours post infection. This period, where infection of cells has occurred but significant release of virus has not, is represented in the model by the eclipse phase. Therefore, these measurements during the first day of infection may help to infer the length of the eclipse phase, τ_E. However, similar conclusions could be reached by considering only the measurements at 2 and 24 hours. Since there are multiple measurements at these early time points but only a single measurement at 48 and 72 hours, choosing a distance function that gives equal weight to all points may result in posterior predictions that fit the data well at early time points and not at late times. A previous analysis with such

	Units	Prior distribution
β_c	$(\text{RNA copies}/ml\ h)^{-1}$	$\log_{10}\beta_c \sim U(-10,-4)$
β_v	$(\text{cells}/ml\ h)^{-1}$	$\log_{10}\beta_v \sim U(-10,-4)$
α	$\text{TCID}_{50}/\text{RNA copies}$	$U(0, 1.44)$
τ_E	h	$\log_{10}\tau_E \sim U(0,3)$
τ_I	h	$\log_{10}\tau_I \sim U(0,3)$
ε	$-$	$U(0,1)$
$v_{50}(0)$	TCID_{50}/ml	$\log_{10}v_{50}(0) \sim U(4,7)$
$v_T(0)$	$\text{RNA copies}/ml$	$\log_{10}v_T(0) \sim U(5,12)$
τ_C/τ_E	$-$	$U(0,1)$
n_E	$-$	$U(1,100)$
n_I	$-$	$U(1,100)$
μ_V	h^{-1}	(least squares estimate)
μ_S	h^{-1}	(least squares estimate)
t_0	h	$U(0, 0.5)$
λ	h^{-1}	$\log_{10}\lambda \sim U(-3,0)$
b	h^{-1}	$\log_{10}b \sim U(-3,0)$
f	$-$	$U(0,1)$

Table 5.5: List of parameters for the wild-type Ebola model, their units, and the corresponding prior distributions used in the Bayesian inference.

a distance function produced posterior predictions with little variability at early times but this variability increased sharply through to 72 hours. At 72 hours, 95% credible regions were observed to span multiple orders of magnitude. As the aim is to ultimately use the posterior samples to make predictions about the behaviour after 72 hours, it would be favourable to have posterior predictions that are more confined.

One approach to avoid such problems is to give less weight to the individual points during the initial 24 hours by asking that the model output is close to these observations through some summary statistic. Here, this summary statistic is chosen to be the mean, such that the averages of the variables $w_{50}(t)$, $w_T(t)$

and $w_{int}(t)$ over the first 24 hours are given respectively by

$$\bar{w}_{50} = \frac{1}{|T_{50}|} \sum_{t \in T_{50}} w_{50}(t)\,, \quad \bar{w}_T = \frac{1}{|T_T|} \sum_{t \in T_T} w_T(t)\,, \quad \bar{w}_{int} = \frac{1}{|T_{int}|} \sum_{t \in T_{int}} w_{int}(t)\,.$$

In these expressions, T_{50}, T_T and T_{int} are the respective sets of times at which measurements of extracellular TCID_{50} and extracellular and intracellular Ct values are recorded during the initial 24 hours. If $d(E, M)$ then denotes the distance between experimental observations and model output, this may be defined as

$$d(E, M)^2 = \left(\log \bar{w}_{50}^{(E)} - \log \bar{w}_{50}^{(M)}\right)^2 + \sum_{t \in \{48,72\}} \left(\log w_{50}^{(E)}(t) - \log w_{50}^{(M)}(t)\right)^2$$

$$+\omega_1 \left(\log \bar{w}_T^{(E)} - \log \bar{w}_T^{(M)}\right)^2 + \omega_1 \sum_{t \in \{48,72\}} \left(\log w_T^{(E)}(t) - \log w_T^{(M)}(t)\right)^2$$

$$+\omega_2 \left(\log \bar{w}_{int}^{(E)} - \log \bar{w}_{int}^{(M)}\right)^2 + \omega_2 \sum_{t \in \{2,48,72\}} \left(\log w_{int}^{(E)}(t) - \log w_{int}^{(M)}(t)\right)^2$$

$$+\omega_3 \sum_{t \in T_D} \left(D^{(E)}(t) - D^{(M)}(t)\right)^2\,,$$

where T_D is the set of all times at which measurements of cell death are recorded. Since the total distance can be thought of as the sum of individual Euclidean distances for each data set, it is important that equal weight is given to each data set. For example, suppose that the contribution to $d(E, M)$ from the extracellular Ct values is significantly larger than the contribution from the remaining data sets. To minimise the total distance, it may then be sufficient to minimise the distance between $w_T(t)$ and the extracellular Ct values, thus ignoring the remaining data sets. Therefore, in selecting a distance for ABC, it is important to consider the weighting of points within a single data set and between data sets. For this reason, the constants ω_1, ω_2 and ω_3 scale the individual contributions from data concerning total viral load and cell death so they are of a similar magnitude to the contribution from the infectious viral load data. To obtain these constants, the ABC algorithm has initially been performed for a reduced number of 10^5 iterations. Summary statistics, such as the median and interquartile range, are then used to study the average contribution to the distance from each data set, with the three constants selected such that the magnitudes of each contribution

are similar. Here, the values of these constants are given by $\omega_1 = 16$, $\omega_2 = 1$ and $\omega_3 = 20$.

With an appropriate distance selected, the ABC rejection sampling algorithm is performed for four cases, with each case using a different combination of the available data sets. These may be described as follows:

$i)$ measurements of total and infectious extracellular viral load for the initial 72 hours are used to infer parameter values,

$ii)$ measurements of total intracellular viral load for the initial 72 hours are used to infer parameter values,

$iii)$ measurements of cell death for the initial 168 hours are used to infer parameter values,

$iv)$ measurements of total and infectious extracellular viral load, total intracellular viral load and cell death are used to infer parameter values.

For each case, 4×10^6 iterations of the rejection sampling algorithm are performed, with an acceptance rate of 0.05% resulting in an approximate posterior sample containing 2×10^3 parameter sets. Pointwise median predictions are provided for each case in Figures 5.6, 5.7, 5.8 and 5.9, and match well to the observed data. The predictions appear worse overall when utilising all data sets simultaneously compared to using each data set separately, particularly for predictions of cell death. This could suggest that the model is not capable of correctly describing the cell death dynamics, but it is more likely due to only one time course being available for the majority of data sets. When performing ABC, it is believed that the model is representative of the average true behaviour, and for this reason the model variables are compared to the mean of the observed values. However, here it is not possible to determine whether the single time course provided reflects what would be observed on average, and if it does not, how greatly it differs from the mean behaviour.

The posterior histograms are provided for each parameter in Figures 5.10, 5.11, 5.12 and 5.13, along with the corresponding prior distributions. These show
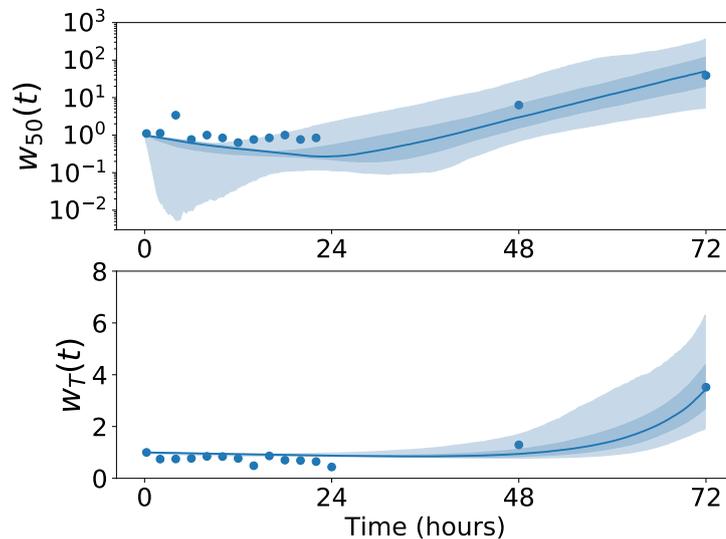
Figure 5.6: Pointwise median predictions for the variables $w_{50}(t)$ and $w_T(t)$ given the posterior sample obtained by performing ABC with only the extracellular data. Shaded regions represent 50% and 95% credible regions.
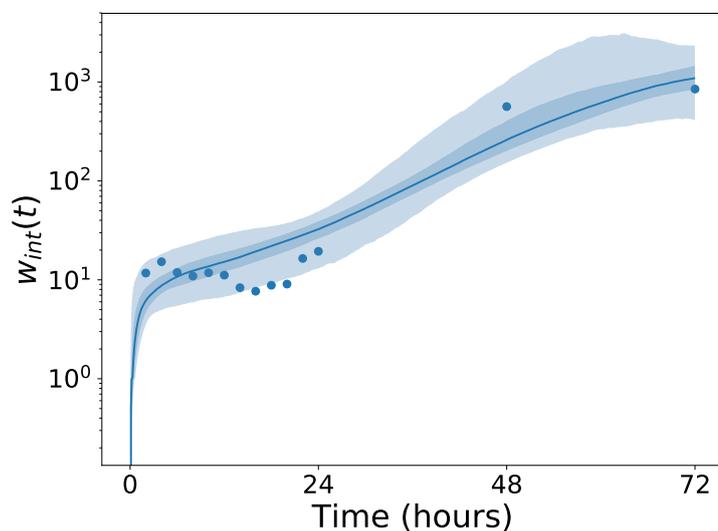
Figure 5.7: Pointwise median predictions for the variable $w_{int}(t)$ given the posterior sample obtained by performing ABC with only the intracellular Ct values. Shaded regions represent 50% and 95% credible regions.
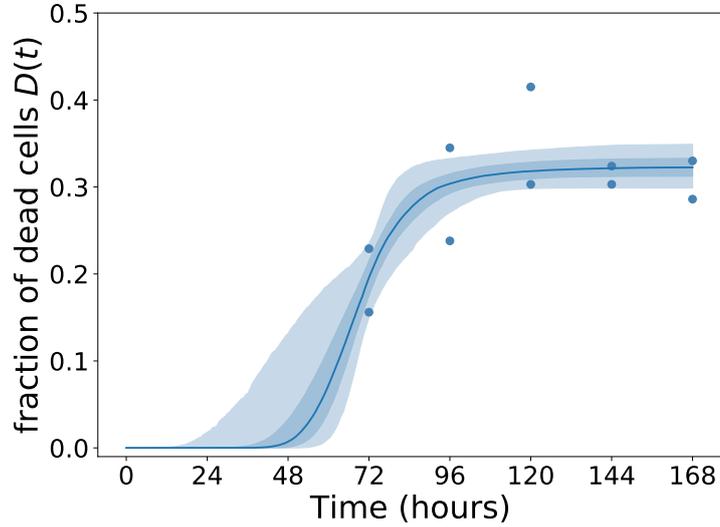
Figure 5.8: Pointwise median predictions for the variable $D(t)$ given the posterior sample obtained by performing ABC with only measurements of cell death. Shaded regions represent 50% and 95% credible regions.
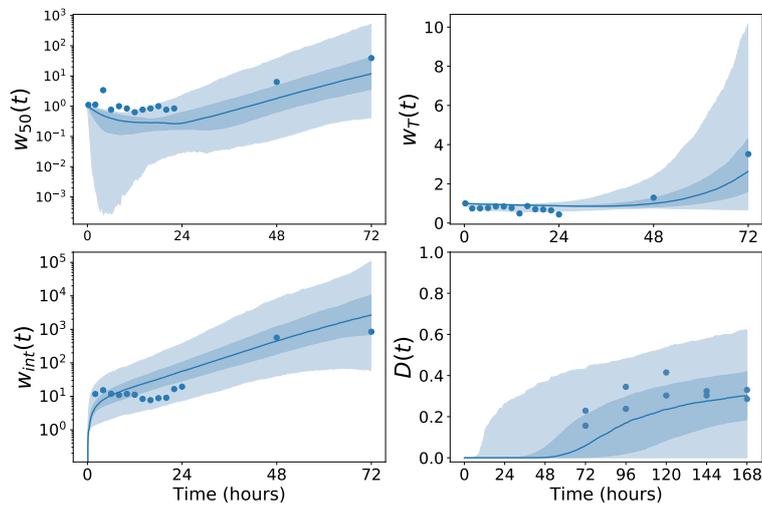
Figure 5.9: Pointwise median predictions for $w_{50}(t)$, $w_T(t)$, $w_{int}(t)$ and $D(t)$ given the posterior sample obtained by performing ABC with all data sets. Shaded regions represent 50% and 95% credible regions.

that from the extracellular measurements, it is possible to learn a lot about important parameters such as β_v and λ. The length of the eclipse phase can be confined to smaller values since larger values would result in too long a delay in the release of virus, and therefore a delayed increase in the extracellular populations of virus. Similarly, smaller values of τ_I can be dismissed, since a short infectious phase results in limited viral production, after which the extracellular viral population would begin to decay. Despite being able to confine the values of τ_E and τ_I, the histograms suggest that within these regions, all values are equally likely.

The intracellular measurements provide further insight into the values of τ_E and β_v. For β_v, this additional learning compared to the extracellular values can likely be explained by the sharp increase in the intracellular viral load between t_0 hours and two hours post infection, with only specific values of β_v able to recreate this behaviour. Aside from these two parameters there is a little additional learning compared to extracellular measurements, with the posterior distribution for λ consistent with that found previously in Figure 5.10.

As expected, measurements of cell death say little about parameters that explicitly dictate viral growth, such as λ and b, since the progression of cells through the different phases is independent of the amount of virus a cell contains. The time until an infected cell stops producing virus is given by the sum $\tau_E + \tau_I$ and so it is not surprising that these parameters are strongly negatively correlated (Figure 5.15), with smaller values of τ_E corresponding to larger values of τ_I. This strong correlation suggests that it would be possible to learn about the ratio of τ_E and τ_I, so it is interesting to see from Figure 5.11 that it is possible to learn about each parameter individually. Some learning can also be seen for β_c as this dictates the rate at which cells become infected, however, the parameter for which the most can be be learnt is f. It is worth noting that, along with the summary statistics provided in Table 5.8, the posterior histogram for f indicates that its value lies close to the observed proportion of dead cells after seven days. Since this proportion takes into account the total number of cells, the model therefore suggests that all cells ultimately become infected.
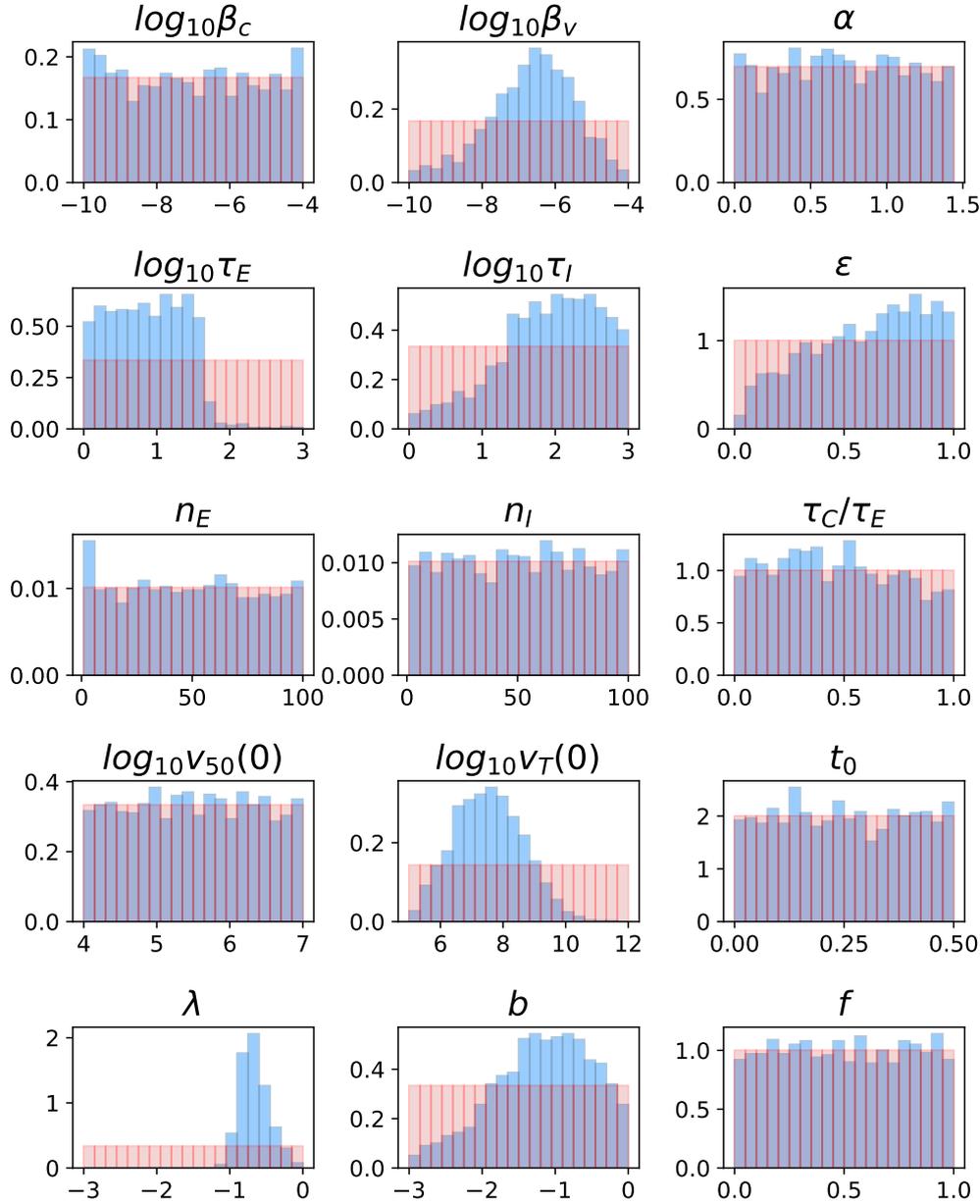
Figure 5.10: Posterior histograms (blue) obtained by performing 4×10^6 iterations of an ABC rejection sampling algorithm using extracellular Ct values and extracellular TCID$_{50}$. The prior distributions are provided in red to indicate how much can be learnt about each parameter.

156
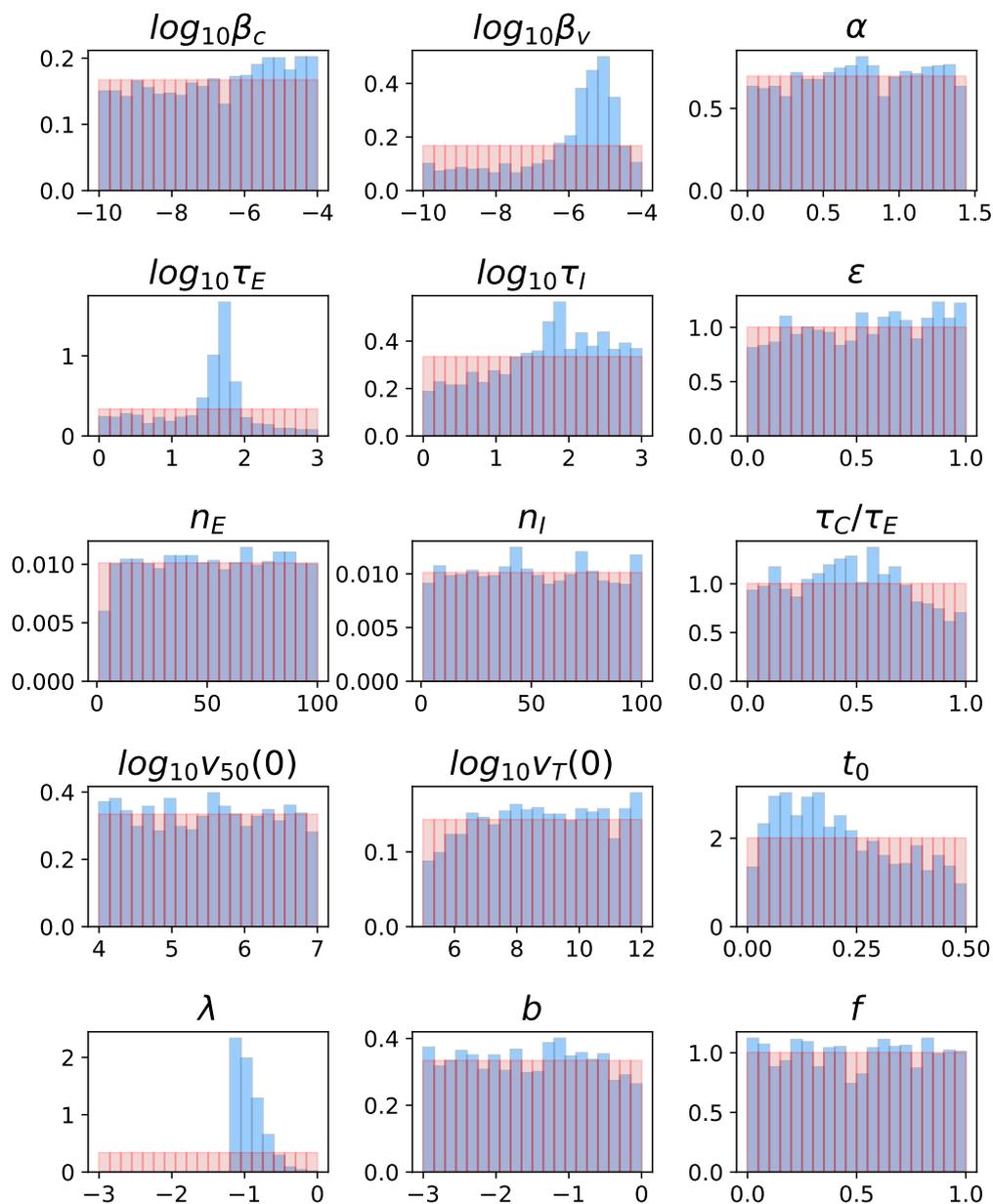
Figure 5.11: Posterior histograms obtained by performing 4×10^6 iterations of an ABC rejection sampling algorithm using only intracellular Ct values. The prior distributions are provided in red to indicate how much can be learnt about each parameter.

	Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
β_c	1.0×10^{-10}	2.3×10^{-9}	8.8×10^{-8}	9.1×10^{-8}	3.2×10^{-6}	1.0×10^{-4}
β_v	1.0×10^{-10}	4.2×10^{-9}	2.9×10^{-7}	2.3×10^{-7}	1.6×10^{-6}	9.9×10^{-5}
α	3.0×10^{-4}	3.7×10^{-1}	7.1×10^{-1}	5.3×10^{-1}	1.1×10^{0}	1.4×10^{0}
τ_E	1.0×10^{0}	2.8×10^{0}	7.5×10^{0}	7.5×10^{0}	1.9×10^{1}	9.2×10^{2}
τ_I	1.0×10^{0}	2.9×10^{1}	9.7×10^{1}	8.2×10^{1}	2.9×10^{2}	1.0×10^{3}
ε	8.5×10^{-3}	3.9×10^{-1}	6.3×10^{-1}	5.1×10^{-1}	8.2×10^{-1}	1.0×10^{0}
n_E	1.0×10^{0}	2.4×10^{1}	4.9×10^{1}	3.5×10^{1}	7.3×10^{1}	1.0×10^{2}
n_I	1.0×10^{0}	2.5×10^{1}	5.1×10^{1}	3.8×10^{1}	7.5×10^{1}	1.0×10^{2}
τ_C/τ_E	1.0×10^{-4}	2.4×10^{-1}	4.7×10^{-1}	3.5×10^{-1}	7.1×10^{-1}	1.0×10^{0}
$v_{50}(0)$	1.0×10^{4}	6.0×10^{4}	3.1×10^{5}	3.2×10^{5}	1.7×10^{6}	1.0×10^{7}
$v_T(0)$	1.0×10^{5}	5.4×10^{6}	3.2×10^{7}	3.9×10^{7}	2.0×10^{8}	3.5×10^{11}
t_0	1.0×10^{-4}	1.3×10^{-1}	2.5×10^{-1}	1.8×10^{-1}	3.8×10^{-1}	5.0×10^{-1}
λ	6.7×10^{-2}	1.6×10^{-1}	2.1×10^{-1}	2.2×10^{-1}	2.9×10^{-1}	9.9×10^{-1}
b	1.0×10^{-3}	2.3×10^{-2}	7.4×10^{-2}	6.6×10^{-2}	2.2×10^{-1}	9.9×10^{-1}
f	3.0×10^{-4}	2.6×10^{-1}	5.0×10^{-1}	3.7×10^{-1}	7.6×10^{-1}	1.0×10^{0}

Table 5.6: Summary statistics for the posterior sample of each parameter following ABC performed using only extracellular measurements of viral load.

	Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
β_c	1.0×10^{-10}	4.4×10^{-9}	1.9×10^{-7}	1.5×10^{-7}	5.4×10^{-6}	9.9×10^{-5}
β_v	1.0×10^{-10}	1.2×10^{-7}	2.8×10^{-6}	7.4×10^{-7}	9.3×10^{-6}	9.9×10^{-5}
α	5.0×10^{-4}	3.9×10^{-1}	7.4×10^{-1}	5.6×10^{-1}	1.1×10^{0}	1.4×10^{0}
τ_E	1.0×10^{0}	1.4×10^{1}	4.3×10^{1}	2.8×10^{1}	6.0×10^{1}	9.9×10^{2}
τ_I	1.0×10^{0}	1.3×10^{1}	6.1×10^{1}	4.9×10^{1}	2.3×10^{2}	1.0×10^{3}
ε	9.0×10^{-4}	2.8×10^{-1}	5.4×10^{-1}	4.0×10^{-1}	7.7×10^{-1}	1.0×10^{0}
n_E	1.0×10^{0}	2.7×10^{1}	5.1×10^{1}	4.0×10^{1}	7.6×10^{1}	1.0×10^{2}
n_I	1.0×10^{0}	2.6×10^{1}	4.9×10^{1}	3.8×10^{1}	7.5×10^{1}	1.0×10^{2}
τ_C/τ_E	9.0×10^{-4}	2.6×10^{-1}	4.7×10^{-1}	3.6×10^{-1}	6.8×10^{-1}	1.0×10^{0}
$v_{50}(0)$	1.0×10^{4}	5.3×10^{4}	3.2×10^{5}	3.0×10^{5}	1.7×10^{6}	9.9×10^{6}
$v_T(0)$	1.1×10^{5}	1.2×10^{7}	4.6×10^{8}	5.0×10^{8}	2.3×10^{10}	1.0×10^{12}
t_0	4.8×10^{-3}	1.0×10^{-1}	2.0×10^{-1}	1.7×10^{-1}	3.2×10^{-1}	5.0×10^{-1}
λ	6.4×10^{-2}	8.3×10^{-2}	1.0×10^{-1}	1.2×10^{-1}	1.5×10^{-1}	8.9×10^{-1}
b	1.0×10^{-3}	5.1×10^{-3}	3.1×10^{-2}	2.9×10^{-2}	1.6×10^{-1}	1.0×10^{0}
f	4.0×10^{-4}	2.4×10^{-1}	5.0×10^{-1}	3.6×10^{-1}	7.5×10^{-1}	1.0×10^{0}

Table 5.7: Summary statistics for the posterior sample of each parameter following ABC performed using only intracellular Ct values.
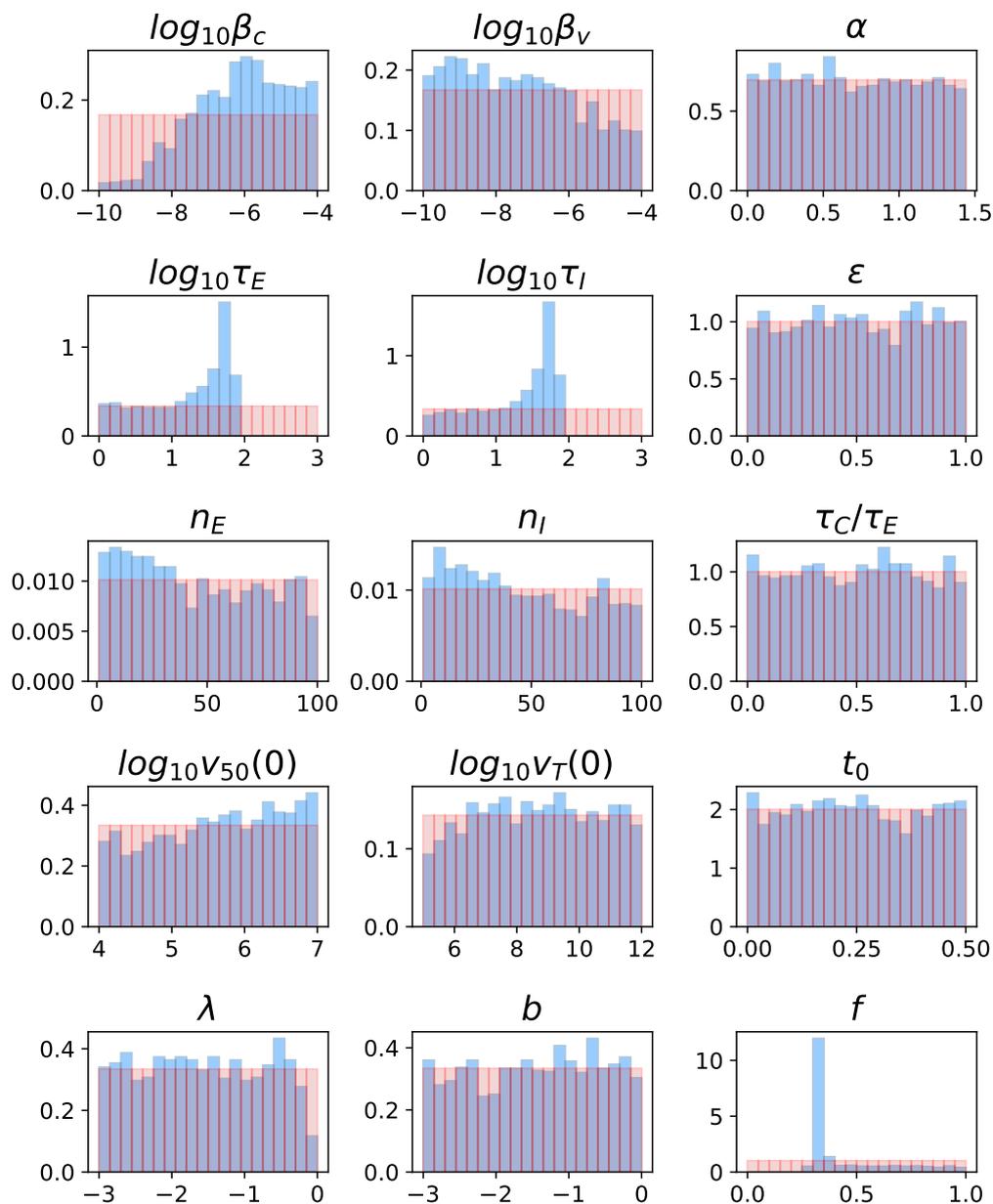
Figure 5.12: Posterior histograms obtained by performing 4×10^6 iterations of an ABC rejection sampling algorithm using only measurements of cell death. The prior distributions are provided in red to indicate how much can be learnt about each parameter.
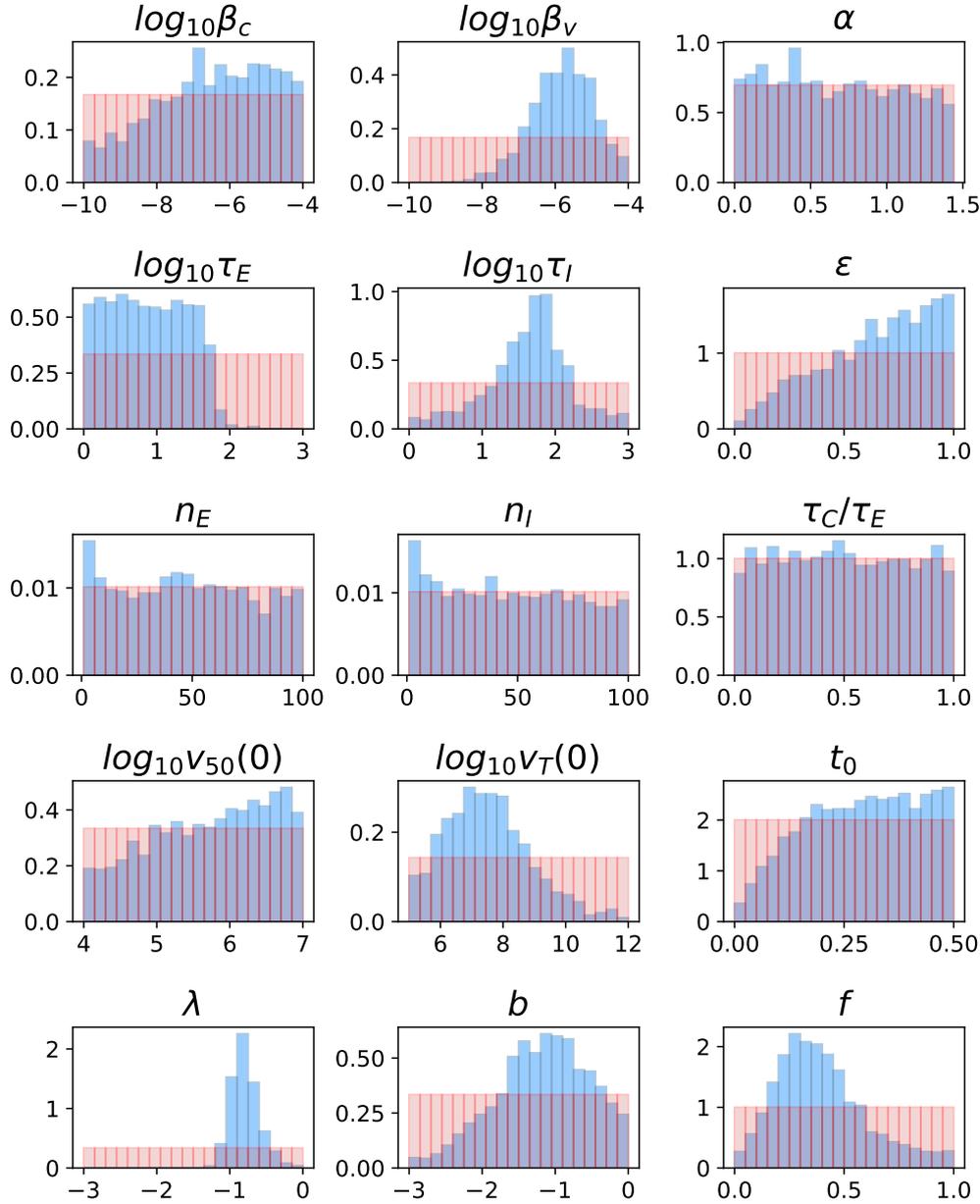
Figure 5.13: Posterior histograms obtained by performing 4×10^6 iterations of an ABC rejection sampling algorithm using all experimental measurements. The prior distributions are provided in red to indicate how much can be learnt about each parameter.

|	Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
β_c	1.1×10^{-10}	8.6×10^{-8}	1.0×10^{-6}	7.5×10^{-7}	8.3×10^{-6}	1.0×10^{-4}
β_v	1.0×10^{-10}	1.6×10^{-9}	3.3×10^{-8}	4.3×10^{-8}	8.0×10^{-7}	9.9×10^{-5}
α	1.0×10^{-4}	3.5×10^{-1}	7.0×10^{-1}	5.2×10^{-1}	1.1×10^{0}	1.4×10^{0}
τ_E	1.0×10^{0}	5.5×10^{0}	2.4×10^{1}	1.6×10^{1}	5.2×10^{1}	7.5×10^{1}
τ_I	1.0×10^{0}	7.2×10^{0}	3.0×10^{1}	1.8×10^{1}	5.4×10^{1}	7.4×10^{1}
ε	3.0×10^{-4}	2.6×10^{-1}	5.0×10^{-1}	3.7×10^{-1}	7.6×10^{-1}	1.0×10^{0}
n_E	1.0×10^{0}	2.0×10^{1}	4.4×10^{1}	3.4×10^{1}	7.2×10^{1}	1.0×10^{2}
n_I	1.0×10^{0}	2.0×10^{1}	4.3×10^{1}	3.4×10^{1}	7.2×10^{1}	1.0×10^{2}
τ_C/τ_E	1.7×10^{-3}	2.5×10^{-1}	5.1×10^{-1}	3.6×10^{-1}	7.4×10^{-1}	1.0×10^{0}
$v_{50}(0)$	1.0×10^{4}	8.1×10^{4}	4.8×10^{5}	4.2×10^{5}	2.4×10^{6}	1.0×10^{7}
$v_T(0)$	1.0×10^{5}	9.9×10^{6}	4.5×10^{8}	4.1×10^{8}	1.9×10^{10}	1.0×10^{12}
t_0	1.0×10^{-4}	1.3×10^{-1}	2.5×10^{-1}	1.8×10^{-1}	3.8×10^{-1}	5.0×10^{-1}
λ	1.0×10^{-3}	5.5×10^{-3}	2.7×10^{-2}	2.8×10^{-2}	1.5×10^{-1}	9.2×10^{-1}
b	1.0×10^{-3}	6.1×10^{-3}	3.8×10^{-2}	3.5×10^{-2}	1.9×10^{-1}	1.0×10^{0}
f	3.0×10^{-1}	3.2×10^{-1}	3.3×10^{-1}	4.1×10^{-1}	5.0×10^{-1}	1.0×10^{0}

Table 5.8: Summary statistics for the posterior sample of each parameter following ABC performed using only measurements of cell death.

	Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
β_c	1.0×10^{-10}	2.3×10^{-8}	4.2×10^{-7}	3.2×10^{-7}	6.8×10^{-6}	9.9×10^{-5}
β_v	1.1×10^{-9}	4.5×10^{-7}	1.7×10^{-6}	1.6×10^{-6}	6.4×10^{-6}	9.9×10^{-5}
α	6.0×10^{-4}	3.4×10^{-1}	6.7×10^{-1}	5.0×10^{-1}	1.0×10^{0}	1.4×10^{0}
τ_E	1.0×10^{0}	2.7×10^{0}	7.4×10^{0}	7.7×10^{0}	2.1×10^{1}	2.3×10^{2}
τ_I	1.0×10^{1}	2.2×10^{1}	5.0×10^{1}	4.3×10^{1}	9.1×10^{1}	1.0×10^{3}
ε	2.1×10^{-2}	4.6×10^{-1}	6.8×10^{-1}	5.7×10^{-1}	8.6×10^{-1}	1.0×10^{0}
n_E	1.0×10^{0}	2.3×10^{1}	4.7×10^{1}	3.4×10^{1}	7.2×10^{1}	1.0×10^{2}
n_I	1.0×10^{0}	2.1×10^{1}	4.6×10^{1}	3.3×10^{1}	7.2×10^{1}	1.0×10^{2}
τ_C/τ_E	5.0×10^{-4}	2.5×10^{-1}	4.9×10^{-1}	3.7×10^{-1}	7.5×10^{-1}	1.0×10^{0}
$v_{50}(0)$	1.0×10^{4}	1.1×10^{5}	6.4×10^{5}	5.2×10^{5}	2.7×10^{6}	1.0×10^{7}
$v_T(0)$	1.0×10^{5}	3.3×10^{6}	2.5×10^{7}	3.3×10^{7}	2.2×10^{8}	9.0×10^{11}
t_0	6.3×10^{-3}	1.9×10^{-1}	3.0×10^{-1}	2.5×10^{-1}	4.0×10^{-1}	5.0×10^{-1}
λ	4.9×10^{-2}	1.2×10^{-1}	1.5×10^{-1}	1.6×10^{-1}	2.1×10^{-1}	9.6×10^{-1}
b	1.0×10^{-3}	2.5×10^{-2}	7.3×10^{-2}	6.6×10^{-2}	2.0×10^{-1}	1.0×10^{0}
f	1.0×10^{-3}	2.5×10^{-1}	3.7×10^{-1}	3.4×10^{-1}	5.1×10^{-1}	9.9×10^{-1}

Table 5.9: Summary statistics for the posterior sample of each parameter following ABC performed using all four data sets.

The learning from the individual data sets is collectively brought together when considering all four types of data at once. The histogram for β_c shows similar learning to when only cell death measurements are used, with extracellular and intracellular viral loads previously unable to provide any insight into this parameter. This is also similar for τ_I, with some additional weight given to larger values, as suggested by the measurements of extracellular viral load. For ε, λ, $v_T(0)$ and b, the posterior histograms closely resemble those in Figure 5.10 since the intracellular viral load and cell death measurements offer little additional learning. Furthermore, due to the consistency observed between the histograms for λ in Figures 5.10 and 5.11, a similar shape is seen when all data sets are considered. For β_v, the peak of the distribution is close to that suggested from the intracellular Ct values, but now the smallest values of β_v have zero density.

Since it is possible to learn about the ratio of strongly correlated parameters, the Pearson correlation coefficient has been computed for the posterior samples of pairs of parameters, with the results presented in Figures 5.14 and 5.15 for the different combinations of data used in the inference. Noticeably, when only comparing to extracellular data, there are strong correlations between β_v and λ, as well as $v_T(0)$ and $v_{50}(0)$. The negative correlation between β_v and λ can be explained by considering that a larger value of β_v results in a greater number of virus particles entering host cells, and therefore a lower replication rate is required to counteract this. For $v_T(0)$ and $v_{50}(0)$, the positive correlation reflects that a sample of EBOV containing greater levels of total virus will naturally contain greater levels of infectious virus. As a result of this, it is possible to learn about the initial ratio of infectious virus particles to total virus particles, given here by $v_{50}(0)/(\alpha v_T(0))$. The posterior samples of $v_{50}(0)$, $v_T(0)$ and α may be used to construct a posterior sample for this ratio, giving a median estimate of 1.7×10^{-2}.

When learning from the intracellular measurements, the strong correlation between $v_{50}(0)$ and $v_T(0)$ is not present. This is likely due to only the total intracellular viral load being measured, and not the infectious viral load. Instead, many of the stronger correlations observed here involve t_0, since the intracellular viral load at this point is used in the normalisation of $v_{int}(t)$. As larger values
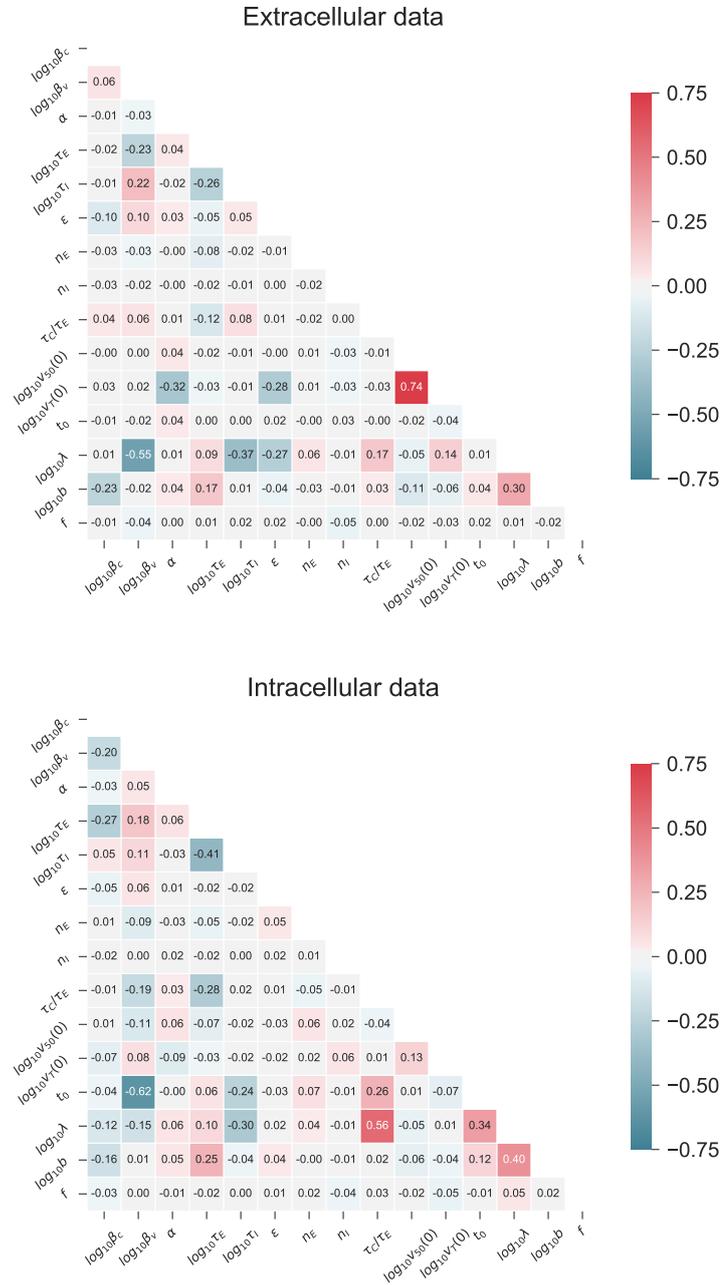
Figure 5.14: Correlations between the posterior samples of pairs of parameters following ABC performed using only extracellular measurements (top) and only intracellular measurements (bottom).
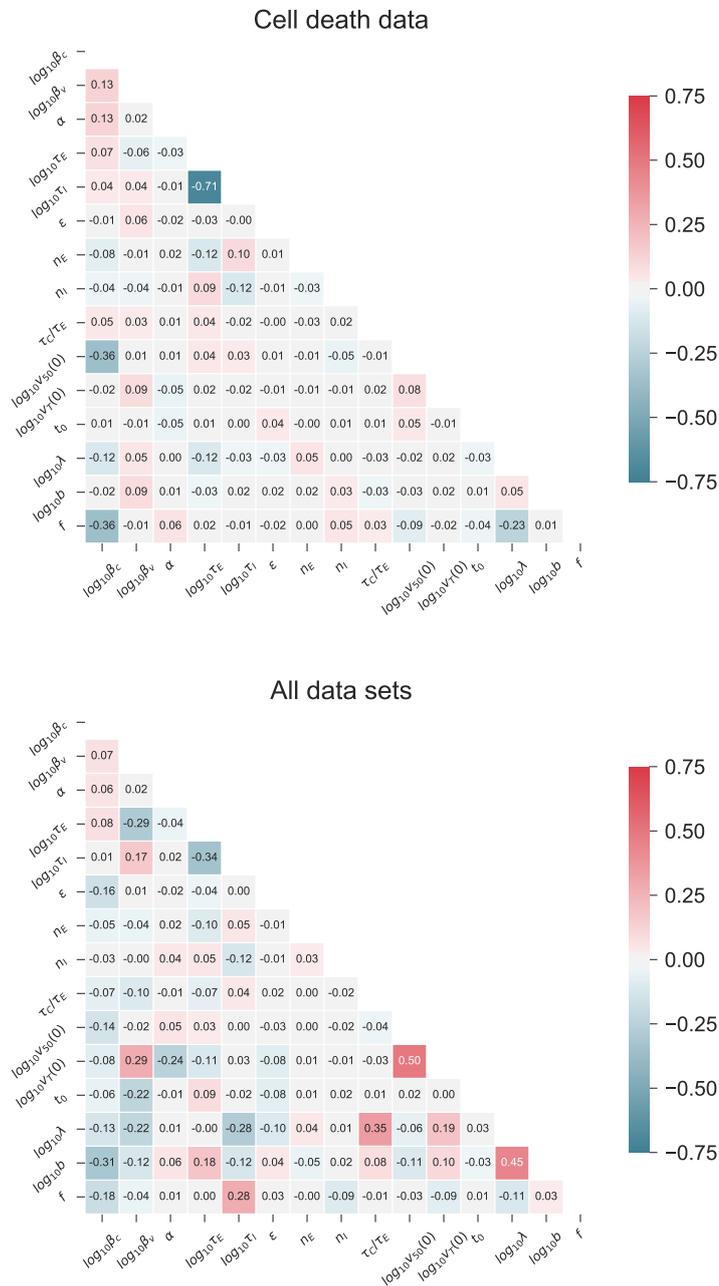
Figure 5.15: Correlations between the posterior samples of pairs of parameters following ABC performed using only cell death measurements (top) and using all available data sets (bottom).

of τ_C/τ_E delay the point at which viral replication can begin, it is not surprising that this parameter is strongly correlated with λ.

From the cell death data, there are generally weak correlations between the majority of parameters, with the strong negative correlation between τ_E and τ_I mentioned previously. The correlation between $v_{50}(0)$ and β_c allows more to be learnt about the ratio $\beta_c v_{50}(0)/\alpha$. It is likely that more can be learnt about this ratio rather than the individual parameters since this is the scaled rate at which target cells become infected, as described in the model equations.

By combining all data sets, the strong correlation between $v_{50}(0)$ and $v_T(0)$ is again present, although the median of $v_{50}(0)/(\alpha v_T(0))$ is now 4.4×10^{-2}, suggesting that infectious virus represents a slightly larger fraction of total virus particles compared to the value inferred from only the extracellular data. Other strongly correlated parameters include λ and b, which is understandable since together they define the net growth of intracellular virus in infectious phase cells, τ_I and τ_E, and λ and τ_C/τ_E.

Overall, the majority of the learning arises from comparing the model to measurements of extracellular viral load. Although the intracellular viral load allows a better estimate of τ_E to be obtained, this can also be achieved by comparing to measurements of cell death. The cell death data also has the added benefit of providing greater learning for β_c, τ_I and f, and would therefore be preferred over the intracellular viral load. Notable parameters that are difficult to learn about are α, the scale factor between TCID_{50} and RNA copies. The difficulty in estimating this parameter is also observed by Iwami *et al.* (2012), who, given data regarding infectious and total viral loads of influenza virus, are able to estimate that one TCID_{50} corresponds to between 3.3×10^2 and 1.2×10^5 infectious virus particles. Little can also be learnt about the number of states in the eclipse and infectious phases, with the magnitude of the correlation coefficient between these parameters and each remaining parameter less than 0.12. This suggests that in future it may be plausible to fix the values of both n_E and n_I.

5.3 Results

In this section, I describe how the posterior distributions obtained from the Bayesian inference can be used to make predictions regarding different model quantities, such as the number of cells in each phase and the basic reproduction number.

5.3.1 Cellular and late time predictions

The wild-type model of EBOV infection has primarily been parametrised using measurements of viral load, with the only exception being the use of cell death data. It is therefore also interesting to consider posterior predictions of other model variables, in particular, the number of cells in each state. Here, the number of cells in the whole eclipse phase is given by $E(t) = \sum_{i=1}^{n_E} E_i(t)$ whilst the number of cells in the infectious phase is $I(t) = \sum_{j=1}^{n_I} I_j(t)$. Figure 5.16 shows posterior predictions for these variables as well as the number of target cells, $T(t)$, using posterior samples obtained by performing ABC with only extracellular measurements and also from ABC performed with all data sets. When comparing to only extracellular measurements, the credible regions of $E(t)$ show two peaks, possibly suggesting that two rounds of infection occur. However, the pointwise median and credible regions only provide a summary of the overall dynamics and cannot provide insight into individual solutions. When instead considering individual solutions, it is possible to see that solutions either contribute to the first or second peak and not both. To determine what causes this difference in dynamics, let $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively denote the sets of posterior samples for which the maximum of $E(t)$ occurs within 30 hours or occurs after 30 hours post infection. As shown in Figure 5.17, the posterior distributions of β_c and $v_{50}(0)$ are skewed towards larger values for parameter samples in $\mathcal{P}_1$, whilst for parameter samples in $\mathcal{P}_2$, the same distributions are skewed towards smaller values. Figure 5.18 shows a histogram of the cellular infection rate, $\beta_c v_{50}(0)/\alpha$, for parameter samples in $\mathcal{P}_1$ and $\mathcal{P}_2$, from which it is evident that the two peaks observed in Figure 5.16 are attributed to contrasting rates of infection. For parameter
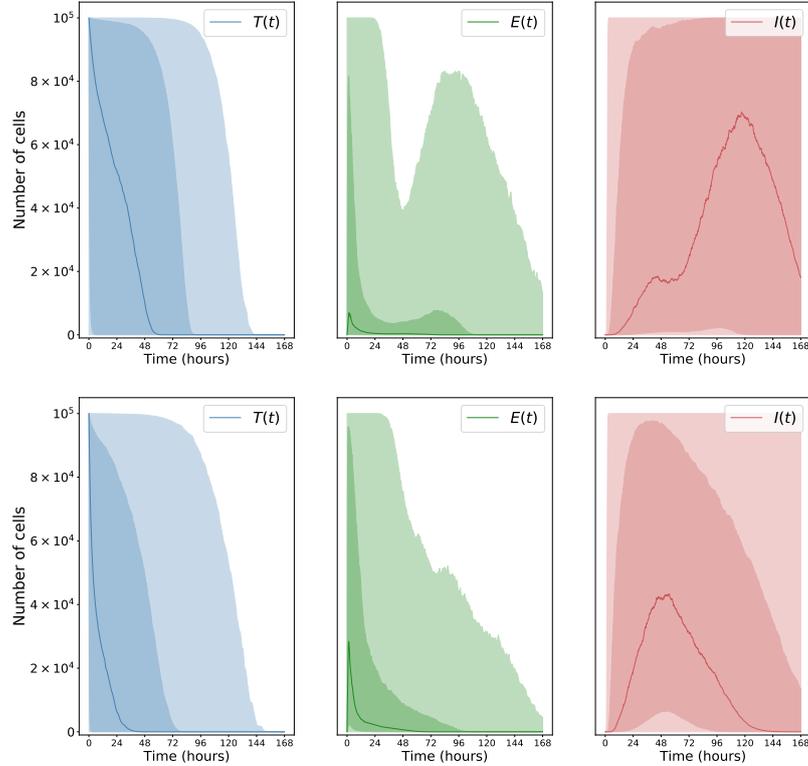
Figure 5.16: Posterior predictions for the total number of target cells, $T(t)$, eclipse phase cells, $E(t)$, and infectious phase cells, $I(t)$. Posterior samples are obtained by performing ABC using only extracellular measurements (top) and all data sets (bottom). Shaded regions represent 50% and 95% credible regions.

samples in $\mathcal{P}_1$, infection events occur quickly with almost all target cells infected in the first 24 hours, whereas for parameter samples in $\mathcal{P}_2$, the infection of target cells is much more gradual. This idea also explains why the median number of infectious phase cells appears to increase in two stages when the inference is conducted using only extracellular measurements, and also why it takes longer for the median number of target cells to reach zero. Previously, the posterior histograms in Figure 5.10 suggested that little could be learnt about β_c and $v_{50}(0)$, so it is interesting that the distribution of the infection rate is distinctly bimodal. Furthermore, Figure 5.14 shows no indication of this relationship, with the correlation between β_c and $v_{50}(0)$ approximately zero. As in Chapter 4, where
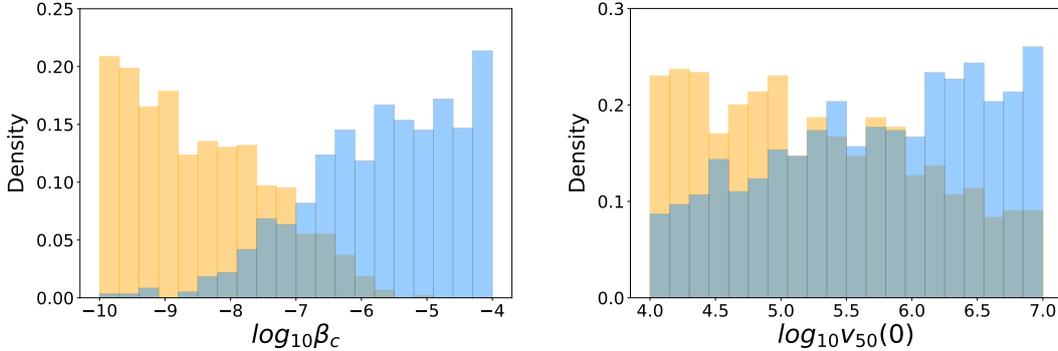
Figure 5.17: Posterior histograms of β_c and $v_{50}(0)$ for the sets $\mathcal{P}_1$ (blue) and $\mathcal{P}_2$ (orange).

it was described how additional measurements regarding the activation states of cells could help improve the inference, these qualitative results relating to the speed of infection could also act as a method for eliminating parameter sets, thus refining the posterior samples.

When performing ABC that utilises all data sets, the measurements of cell death indicate that β_c is more likely to attain larger values. For this reason, the corresponding predictions of $T(t)$ and $E(t)$ suggest a scenario where target cells become infected quickly and thus only a single peak is seen in the number of eclipse phase cells. Since the learning of β_c is due to the cell death measurements, similar predictions for $T(t)$, $E(t)$ and $I(t)$ are observed when using the posterior sample from the ABC that compares only to these measurements. When comparing to only intracellular Ct values, similar dynamics are observed to those obtained when comparing to extracellular measurements.

Although model parameters have in part been inferred from measurements of extracellular TCID_{50} and extracellular Ct values during the first 72 hours of infection, additional observations for these variables are available for the subsequent four days of infection. It is therefore also possible to assess the predictive ability of the model at these later time points. Figure 5.19 shows pointwise median predictions along with 50% and 95% credible regions for each variable during the initial seven days of infection, with the additional data for extracellular TCID_{50}
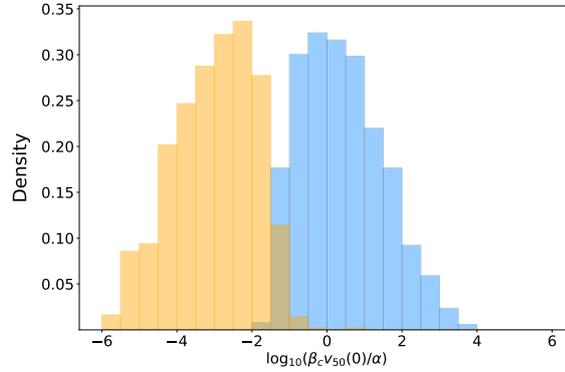
Figure 5.18: Posterior histograms of the infection rate $\beta_c v_{50}(0)/\alpha$ for parameter samples in $\mathcal{P}_1$ (blue) and $\mathcal{P}_2$ (orange).

and extracellular Ct values indicated. These predictions suggest that the median model output is not able to sufficiently account for the decay in the extracellular viral counts that occur at later time points, instead over-predicting the amount of virus. However, all of the measurements at late times lie within or close to the 50% credible regions, suggesting that there are parameter regimes that can describe these dynamics.

Given the current description of the model, there are two reasons why the extracellular viral load decreases at late time points. The first is that the budding rate of virus, b, is greater than the replication rate, λ. In this case, when cells enter the infectious phase, their intracellular viral load will decrease such that there is not sufficient release of virus to counter the degradation of extracellular virus. However, analysis of the posterior sample shows that $\lambda > b$ for over 75% of parameter sets, indicating that the intracellular viral load would continue to grow whilst cells remain in the infectious phase. The other, more probable reason, is due to cells no longer producing virus after they exit the infectious phase. Once all cells exit the infectious phase, degradation and denaturation of virus are then the only events that take place. This suggests that the large variability and over-predictions observed in $w_{50}(t)$ and $w_T(t)$ at late time points could be due to less accurate estimates of the total time a cell remains infected, $\tau_E + \tau_I$.
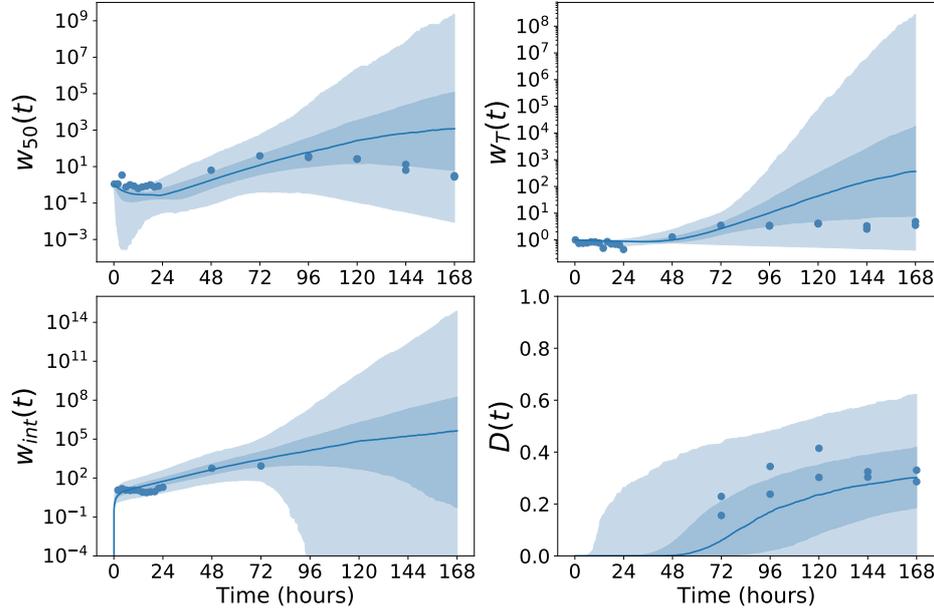
Figure 5.19: Pointwise median predictions for the first seven days of infection where shaded regions represent 50% and 95% credible regions.

5.3.2 Basic reproduction number

When describing the potential spread of an infectious disease in an epidemic, a commonly associated quantity is the basic reproduction number, $\mathcal{R}_0$. This also has an interpretation in models of viral kinetics, where it is defined as the expected number of secondary infections produced by a single infected cell in a completely susceptible population (Nowak & May (2000)). When $\mathcal{R}_0 < 1$, infected cells will on average fail to replace themselves and so the infection is contained. On the other hand, if $\mathcal{R}_0 > 1$, the number of infected cells increases and the disease will spread (Van den Driessche & Watmough (2008)). The basic reproduction number can be expressed as the product of the expected duration of the infectious period and the rate at which secondary infections occur. In calculating the rate at which secondary infections occur, this takes into account the rate at which virus is produced by infectious cells. For previous models of viral dynamics, this is simple to determine since the production of virus occurs at a constant rate that is proportional to the number of infected cells. Here, however, difficulties arise

due to the assumption that viral production is proportional to the intracellular viral load. As a result, the production of virus from a single infected cell is no longer constant, but instead a function of time.

In order to find $\mathcal{R}_0$ for the mathematical model described in Section 5.1, recall that the intracellular viral load of an infectious phase cell is modelled as a birth-and-death process, such that a death event represents the release of a single virus particle from the cell. The average amount of virus released by a cell throughout the duration of the infectious phase may then be thought of as the average number of death events that occur in the birth-and-death process. To determine this average, it is therefore necessary to take into account both the dynamics of the intracellular viral population as well as the time that an infected cell spends in the infectious phase. For this reason, let $\mathcal{X} = \{\boldsymbol{X}(t) = (X(t), Y(t)) : t \geq 0\}$ be a bivariate Markov process where $X(t)$ denotes the intracellular viral load of an infectious phase cell at time t, and $Y(t)$ represents the state of the cell within the infectious phase. The state space of $\mathcal{X}$ is given by $\mathcal{S}_{\mathcal{X}} = \{(i, j) : i \geq 0, 1 \leq j \leq n_I\} \cup \{\phi\}$. If the transition probabilities are defined as

$$p_{(i,j),(m,n)}(\Delta t) = \Pr(\boldsymbol{X}(t + \Delta t) = (m, n) \mid \boldsymbol{X}(t) = (i, j)),$$

then $\mathcal{X}$ may be defined by its one-step transition probabilities

$$p_{(i,j),(m,n)}(\Delta t) = \begin{cases} \lambda i \Delta t + o(\Delta t) & (m, n) = (i + 1, j), \\ b i \Delta t + o(\Delta t) & (m, n) = (i - 1, j), \\ \delta_I \Delta t + o(\Delta t) & (m, n) = (i, j + 1), \\ 1 - ((\lambda + b)i + \delta_I)\Delta t + o(\Delta t) & (i, j) = (m, n), \\ o(\Delta t) & \text{otherwise}, \end{cases}$$

for states $(i, j) \in \mathcal{S}_{\mathcal{X}}$. A depiction of $\mathcal{X}$ is provided in Figure 5.20. Since the rate at which cells transition through infectious phase states is independent of the birth-and-death process used to represent viral replication and budding, $\mathcal{X}$ has a similar interpretation to the within-phagocyte level of the multi-scale model described in Chapter 3. The transitioning of $\mathcal{X}$ down columns of states in Figure 5.20
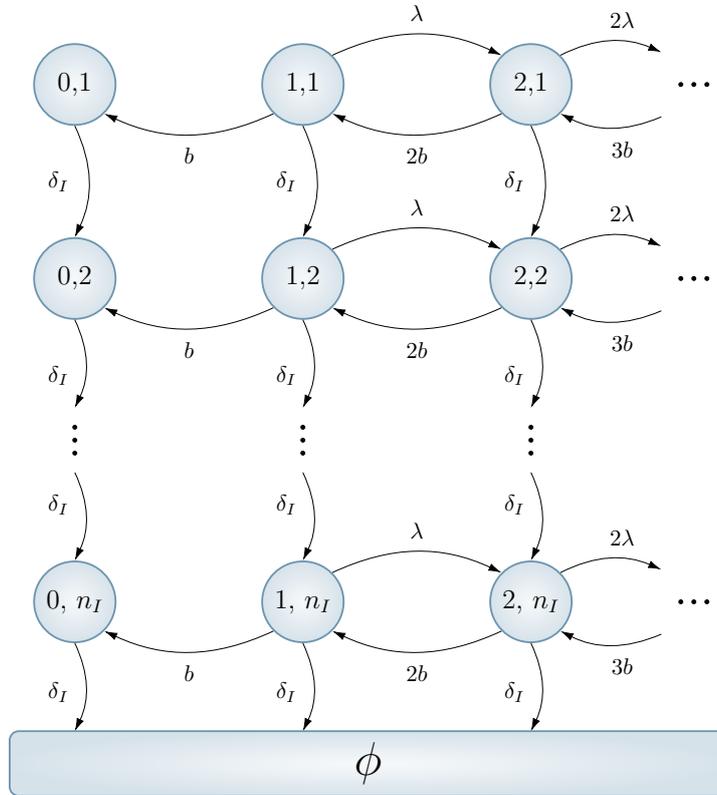
Figure 5.20: A depiction of the Markov process $\mathcal{X}$ used to model a single cell during the infectious phase. State (i, j) indicates that the cell contains i virus and is currently in the jth state of the infectious phase.

represents that cell death becomes increasingly nearer and is therefore analogous to the rupture clock from Section 3.1.1. The process $\mathcal{X}$ has a single absorbing state, ϕ, representing that viral production no longer occurs from the infected cell. State ϕ has an equivalent interpretation to that depicted in Figure 5.1.

Let R be a random variable representing the amount of virus released by a cell whilst in the infectious phase. With reference to Figure 5.20, this may be understood as the number of jumps the process $\mathcal{X}$ makes to the left before it reaches the absorbing state ϕ. A stochastic descriptor for the average number of virus released may then be defined as

$$m_{i,j} = \mathbb{E}\left[R \mid \boldsymbol{X}(0) = (i, j)\right],$$

where $m_{0,j} = 0$ for all j. By conditioning on the subsequent state of $\mathcal{X}$, first step analysis can be used to construct a system of equations relating $m_{i,j}$ for each state in $\mathcal{S}_{\mathcal{X}}$

$$m_{i,j} = \frac{\lambda i}{(\lambda + b)i + \delta_I}m_{i+1,j} + \frac{bi}{(\lambda + b)i + \delta_I}(m_{i-1,j}+1) + \frac{\delta_I}{(\lambda + b)i + \delta_I}m_{i,j+1}. \quad (5.6)$$

where $i = 0, 1, 2, ...$ and $j = 1, 2, ..., n_I$. The system described by equation (5.6) can be solved by first considering the final row of states in Figure 5.20, relating to the final infectious phase state. In this case, m_{1,n_I} satisfies

$$m_{1,n_I} = \frac{\lambda}{\lambda + b + \delta_I}m_{2,n_I} + \frac{b}{\lambda + b + \delta_I}.$$

In Section 4.1.2 it is noted that prior to a catastrophe event occurring, a linear birth-and-death process with catastrophe starting in state k may be thought of as k independent processes starting in state 1. The same reasoning can also be applied here to obtain m_{i,n_I}. If a transition from state (i, n_I) into the absorbing state is thought of as a catastrophe event, then a cell containing two virus particles at the start of the infectious phase may be thought of as two cells each containing one. Since these cells are identical and independent, the average number of virus particles released during the infectious phase is the same for each cell, and thus $m_{2,n_I} = 2m_{1,n_I}$. In general, it is then true that $m_{i,n_I} = im_{1,n_I}$ and the system formed by equation (5.6) can be solved recursively to yield

$$m_{i,j} = im_{1,j} = \frac{ib}{b + \delta_I - \lambda}\sum_{k=0}^{n_I-j}\left(\frac{\delta_I}{b + \delta_I - \lambda}\right)^k, \qquad i \geq 1, \, 1 \leq j \leq n_I. \quad (5.7)$$

Since a cell always enters the infectious phase through state I_1, the case $j = 1$ is of particular interest, with equation (5.7) simplifying to

$$m_{i,1} = im_{1,1} = \frac{ib}{\lambda - b}\left(\left(\frac{\delta_I}{b + \delta_I - \lambda}\right)^{n_I} - 1\right), \qquad i \geq 1. \quad (5.8)$$

It should be noted that this mean is only defined when $b + \delta_I > \lambda$. This condition arises due to the difficulty in taking the mean of a birth-and-death process with constant catastrophe rate that has previously been discussed in Section 4.1.4. With this, it is now possible to determine the average number of virus released

during the infectious phase, provided that the viral load of the cell at the start of the infectious phase is known. In order to determine the distribution of cellular viral loads at the start of the infectious phase, recall that the population of intracellular virus is modelled as a Poisson process during the co-infection window and a birth process during the eclipse phase.

Let $\mathcal{W} = \{W(t) : t \geq 0\}$ be a Poisson process, where $W(t)$ is the intracellular viral load at time t of an infected cell during the co-infection window. Since a target cell immediately enters the co-infection window following infection with a single virus, it is suitable to let $W(0) = 1$. The viral load then increases due to further infection events that occur at constant rate $\beta = \beta_c v_{50}(0)/\alpha$ and the transition probabilities of $\mathcal{W}$ are shown by Allen (2003) to have the form

$$\Pr(W(t) = n \,|\, W(0) = 1) = \frac{\beta^{n-1}e^{-\beta t}}{(n-1)!}, \qquad n \geq 1.$$

The distribution of the viral load at the end of the co-infection window can then be obtained by taking into account the $\text{Erlang}(\delta_E, n_C)$ distributed duration of the co-infection window. If $p_n^{(C)}$ is the probability that a cell contains n virus at the end of the co-infection window, then

$$p_n^{(C)} = \int_0^{+\infty} \frac{(\beta t)^{n-1}e^{-\beta t}}{(n-1)!} \frac{\delta_E^{n_C} t^{n_C-1}e^{-\delta_E t}}{(n_C-1)!} \, dt,$$

$$= \binom{(n_C - 1) + (n - 1)}{n - 1} \left(\frac{\beta}{\beta + \delta_E}\right)^{n-1} \left(\frac{\delta_E}{\beta + \delta_E}\right)^{n_C},$$

where the integral has been evaluated by recognising that, up to a constant factor, the integrand is equivalent to the p.d.f of an $\text{Erlang}(\lambda+\delta_E, n+n_C-1)$ distribution.

The probabilities $p_n^{(C)}$ can now be used to form the distribution of initial viral loads for a cell as it enters the eclipse phase. To determine the distribution of viral loads for a cell at the end of the eclipse phase, let $\mathcal{Z} = \{Z(t) : t \geq 0\}$ be a linear birth process where $Z(t)$ represents the viral load of an eclipse phase cell, and consider first the general case where $Z(0) = k$. It is known that, after the co-infection window, the duration of the eclipse phase follows an $\text{Erlang}(\delta_E, n_E - n_C)$ distribution, and the transition probabilities for a linear birth process have previously been described in Section 2.2.7. Therefore, if $p_n^{(E)}(k)$

denotes the probability that a cell contains n virus at the end of the eclipse phase, provided that it contained k virus after the co-infection window, then

$$p_n^{(E)}(k) = \int_0^{+\infty} \binom{n-1}{n-k} e^{-\lambda k t} \left(1 - e^{-\lambda t}\right)^{n-k} \frac{\delta_E^{n_E - n_C} t^{n_E - n_C - 1} e^{-\delta_E t}}{(n_E - n_C - 1)!} \, dt \,,$$

$$= \int_0^{+\infty} \binom{n-1}{n-k} e^{-\lambda k t} \sum_{j=0}^{n-k} \binom{n-k}{j} (-1)^j e^{-\lambda j t} \frac{\delta_E^{n_E - n_C} t^{n_E - n_C - 1} e^{-\delta_E t}}{(n_E - n_C - 1)!} \, dt \,,$$

$$= \sum_{j=0}^{n-k} (-1)^j \binom{n-1}{n-k} \binom{n-k}{j} \left(\frac{\delta_E}{\lambda(k+j) + \delta_E}\right)^{n_E - n_C} \,, \qquad n \geq k \,.$$

The expressions for $p_n^{(C)}$ and $p_n^{(E)}(k)$ can now be combined to find $p_n^{(E)}$, the probability that a cell contains n virus at the end of the eclipse phase

$$p_n^{(E)} = \sum_{k=1}^{n} p_n^{(E)}(k) p_k^{(C)}$$

$$= \sum_{k=1}^{n} \left[\sum_{j=0}^{n-k} (-1)^j \binom{n-1}{n-k} \binom{n-k}{j} \left(\frac{\delta_E}{\lambda(k+j) + \delta_E}\right)^{n_E - n_C} \right]$$

$$\times \binom{(n_C - 1) + (k - 1)}{k - 1} \left(\frac{\beta}{\beta + \delta_E}\right)^{k-1} \left(\frac{\delta_E}{\beta + \delta_E}\right)^{n_C} \,, \qquad n \geq 1 \,.$$

Since the end of the eclipse phase marks the beginning of the infectious phase, these probabilities form the distribution of initial states for finding the average amount of virus released from a cell during the infectious phase, which may now be written as

$$\mathbb{E}[R] = \sum_{i=1}^{+\infty} \mathbb{E}[R \mid \boldsymbol{X}(0) = (i, 1)] \Pr(\boldsymbol{X}(0) = (i, 1)) = \sum_{i=1}^{+\infty} m_{i,1} p_i^{(E)} \,. \qquad (5.9)$$

With an expression for the mean number of virus released from a cell during the period it remains infectious, it is now possible to define $\mathcal{R}_0$. For this, it is also necessary to take into account that only a fraction ε of produced virus is infectious and thus able to cause secondary infections, and that extracellular infectious virus degrades and loses infectivity with rate $\mu_V + \mu_S$. Altogether, an

appropriate expression for $\mathcal{R}_0$ is given by

$$\mathcal{R}_0 = \frac{\varepsilon \mathbb{E}\left[R\right] \beta_c T(0)}{\mu_V + \mu_S} \,.$$ (5.10)

Figure 5.21 shows the distribution of $\mathcal{R}_0$, with a summary also provided in Table 5.10. This distribution has been constructed using the posterior sample obtained by performing ABC with all four data sets, along with equations (5.9) and (5.10). Only parameters satisfying the condition $b + \delta_I > \lambda$ have been used to find $\mathcal{R}_0$.

Previously, Nguyen *et al.* (2015) used bootstrapping to obtain a distribution of $\mathcal{R}_0$ that has a median of approximately $10^4 - 10^5$. Such large values of $\mathcal{R}_0$ have also been observed by Pinilla *et al.* (2012) for influenza virus and can be explained by the use of *in vitro* data to parametrise the model. It is likely that *in vivo*, both the rate of viral clearance and infected cell clearance are greater due to the presence of immune responses, which would then result in smaller estimates for $\mathcal{R}_0$. Although the distribution in Figure 5.21 suggests that large values of $\mathcal{R}_0$ are also possible here, the median value of approximately 16 appears more sensible.

Given that experimental measurements show an increase in the infectious viral load through time, this would suggest that the infection is spreading and thus $\mathcal{R}_0 > 1$. Therefore, as a further approach to refine the posterior distributions, it would also be appropriate to eliminate those parameter sets from the posterior sample that yield $\mathcal{R}_0 < 1$.

5.4 Discussion

In this chapter, I have described how a mathematical model can be constructed to replicate a series of *in vitro* experiments involving Ebola virus. The model is based on traditional target cell-limited models but is unique in its ability to also describe intracellular dynamics following the inclusion of just a single additional parameter. Parameter inference has subsequently been conducted in a Bayesian manner to compare the learning from different experimental measurements. For the measurements of extracellular and intracellular viral load, the limited number of observations means that it is difficult to determine how accurate the data are,
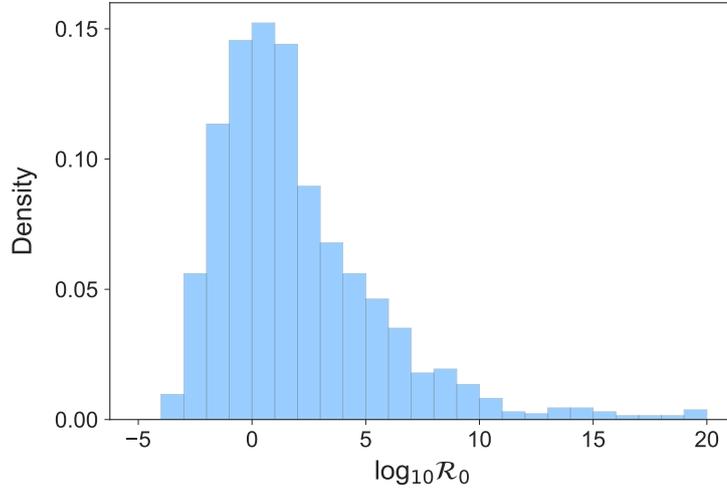
Figure 5.21: Distribution of the basic reproductive number, $\mathcal{R}_0$, constructed using the posterior sample following ABC performed using all available data sets.

	Min.	1$^{\text{st}}$ Qu.	Median	Geo. Mean	3$^{\text{rd}}$ Qu.	Max.
$\mathcal{R}_0$	4.7×10^{-4}	3.8×10^{-1}	1.6×10^{1}	3.8×10^{2}	6.8×10^{3}	5.0×10^{89}

Table 5.10: Summary statistics for the basic reproductive number, $\mathcal{R}_0$.

and thus how trustworthy the posterior distributions can be. However, whilst lacking in the number of observations, this is balanced by the diverse nature of the measurements, with results for viral counts in both the presence and absence of cells, as well as cell death also made available. Ultimately, multiple replicates of a single experiment would provide more accurate estimates of the mean behaviour, to which the model is being compared when performing ABC. This may subsequently result in more certain estimates for some individual model parameters, however, a wide variety of observations can provide greater learning about the whole parameter space, albeit with less certainty. Furthermore, even when only a little can be learnt about a certain parameter, it may still be possible to confine the range of the parameter, as is the case for τ_E in Figure 5.13. With the cost and time associated with performing these experiments, it is important that experimentalists collaborate effectively with modellers in order to determine the

types of experiments that provide the most useful data.

From the analysis performed here, measurements of infectious and total viral counts in the absence of infection appear important, since a simple exponential decay model can then be used to obtain accurate estimates for the rates of viral decay, thus reducing the dimension of the parameter space when performing ABC. Extracellular viral counts are a commonly measured quantity and these are shown here to provide significant learning for a number of parameters, including the rates of infection and intracellular viral replication. However, in order to determine the length of the eclipse and infectious phases, measurements of cell death are required. The intracellular viral load provides information about the length of the eclipse phase, although the little information that it provides for the remaining parameters suggests that it is not beneficial compared to measurements of cell death. Preference should therefore be given to measuring cell death instead of the intracellular viral load. Other cellular measurements such as the proportion of infected cells may be equally successful for providing estimates of the time cells spend in each phase, and could also yield a better estimate for the cellular infection rate, β_c.

One advantage of using ABC to conduct the inference is that it can identify parameters that are not important in the model. For example, when using any of the data sets, the prior and posterior distributions are approximately equal for both n_E and n_I, the number of states that comprise the eclipse and infectious phases. Since an uninformative prior has been assumed, it is not possible to determine whether this is because no learning occurs, or whether all values of n_E and n_I are equally likely. In future, the ABC could be repeated for different choices of prior distribution. If the posterior distribution remains the same as found here, this would suggest that all values are equally likely. In such a case, it may then be more appropriate to fix the values of n_E and n_I in any subsequent analysis.

Chapter 6

Concluding remarks

In this thesis, mathematical models for two highly infectious and lethal pathogens have been developed. Alongside experimental results, these models are important for helping to understand the early stages of pathogenesis and can ultimately be applied to predict the effects of potential treatment strategies. Stochastic approaches to modelling have been used to study *F. tularensis* in order to account for the inherent variability associated with low bacterial counts, which is of particular importance for this specific bacterium given its low infectious dose. The large number of Ebola virus particles present in the *in vitro* experiments suggests that a deterministic approach is more suitable, however, given that a single cell is initially infected by only a few virus particles, a stochastic approach has again been used to model the intracellular dynamics.

In Chapter 3, a novel multi-scale model has been developed to describe the effects that the release of *F. tularensis* has on a population. Through incorporating the intracellular and within-host dynamics, the probability of symptom onset and mean time to symptom onset can be obtained given the initial exposure dose. This dose-and-time response model is an improvement on existing models by explicitly accounting for biological mechanisms, rather than relying on goodness of fit, as is the case for existing statistical models. The model has been parametrised using both *in vivo* and *in vitro* data, and is capable of explaining data obtained following human trials well.

One limitation of this model is the use of matrix analytic methods to analyse the probability of response and mean time to response. Although these methods are efficient in the sense that they utilise partitions of considerably larger matrices, the computational cost of storing and performing operations with matrices restricts their use to stochastic processes with smaller state spaces. When difficulties arise in studying the behaviour of the model using these methods, numerical simulations offer a solution, however, this often requires many realisations and can therefore also be computationally expensive. One solution that has been identified for the within-host model is to consider a 'point of no return', representing a threshold of bacteria that, if attained, indicates that a response is almost certain to occur. For the probability of response, this allows the size of the state space to be reduced, such that matrix analytic methods can be applied and are no longer a limiting factor. For the mean time to response, the 'point of no return' could be used to develop a hybrid approach, where stochastic dynamics are considered when the level of extracellular bacteria is below this threshold and deterministic ODEs are used to describe subsequent dynamics when the population size is larger.

Currently, the multi-scale model does not account for treatment of infected individuals, however, this could be a future development. Still focusing on infection within the lung, one method to include treatment would be to construct a pharmacokinetic model for the concentration of antibiotic. For bactericidal antibiotics, the rate of extracellular bacterial killing in the within-host model may then be specified as an increasing function of this concentration. Therefore, when treatment is administered early enough, the antibiotics are able to resolve the infection through increased bacterial killing before symptom onset can occur. However, even if antibiotics are administered at a later time, the reduction in net bacterial growth would result in a lengthened response period, thus extending the window of opportunity for further treatment.

In Chapter 4, an ABM has been developed for the early stages of *F. tularensis* pathogenesis. Through extending existing results regarding birth processes with catastrophe, it has been possible to characterise the infection cycle of a single

6. CONCLUDING REMARKS

cell. Under the assumption that cells act independently of each other during the first 48 hours of infection, this has then enabled approximations of the ABM to be derived for populations of cells. As in each chapter, a Bayesian approach has been used to infer parameter values using infection data. Whilst this yields an accurate estimate for the intracellular growth rate of *F. tularensis* bacteria, it highlights the difficulties with identifiability that are also observed in Chapters 3 and 5. These problems can sometimes be avoided by combining reactions, such as in the extracellular killing rate in the multi-scale model. However, even in this within-host model with only three reactions, the high correlation between the extracellular killing rate and rate of phagocytosis mean they are difficult to estimate individually. More important are the types of data with which the inference is performed. For the ABM approximations, only bacterial counts are available and this explains why only an estimate for the replication rate can be obtained. Despite this, by comparing ABM simulations for other quantities, such as the number of surviving cells or the proportion of cells in each activation state, it is evident that measurements of this type would enable better learning.

In Chapter 5, a deterministic model for Ebola virus infection is developed that also accounts for stochastic intracellular dynamics using a novel approach that not only requires just a single additional parameter compared to analogous models of extracellular dynamics, but also models the release of virus in what is believed to be a more biologically realistic manner. Through the Bayesian analysis, it is now possible to determine which types of data offer the most learning, and thus we can advise on the specific experiments that should be conducted. This knowledge, and the model description, is not only useful for Ebola virus but is also applicable for different viruses.

Understanding infection with wild-type Ebola virus is the first stage to designing treatment. One treatment strategy currently being considered is the use of defective interfering particles (DIPs), attenuated virus particles that interfere during the replication of Ebola virus by competing for viral proteins (Calain *et al.* (1999)). A mathematical model that incorporates DIPs would include all reactions that currently exist in the wild-type model, with additional reactions

relating to the entry and competition of the DIPs. Future work could therefore make use of the posterior distributions obtained here. By already having suitable parameter values inferred from *in vitro* experiments for the wild-type model, a more effective analysis of the parameters influencing DIPs could be conducted. For example, it would be interesting to consider how the rate of entry of DIPs and the level of interference affects the concentration of infectious virus particles. Naturally, however, since these experiments have been performed *in vitro*, care must be taken when using these results to reach conclusions about DIP interference at an *in vivo* level, and thus the effect of DIPs as therapeutic agents.

Appendix A

Code samples

A.1   Python code for a tau-leaping algorithm

For the within-host model described in Section 3.1.2, there are two species and $N + 2$ reactions. The multiple rupture events are represented in N of these reactions, whilst phagocytosis followed by bacterial survival and extracellular bacterial death encompass the remaining reactions. Therefore, with reference to the tau-leaping algorithm outlined in Algorithm 2.2, $I = \{B(t), P(t)\}$ and $M = N + 2$. By ignoring the absorbing state at 0, the critical reactions may be determined by considering that $B(t) \geq P(t)$ and $B(t) > 0$. From this, either every reaction is critical, every rupture event is critical, only phagocytosis and extracellular death events are critical, or no reactions are critical. Using this information, the following code may be implemented to simulate one realisation of the within-host model.

```python
import random, numpy, math

# Evaluates the propensity function for each reaction
def Make_Rates(nT,nP):
    return [mu*(nT-nP), alpha*(nT-nP), delta*nP]

# Functions to return the state change vector when the reaction occurs kj times.
def ExDeath(kj):
    return -kj, 0
def Phagocytosis(kj):
```

```python
        return 0, kj

def Rupture(kj):
    return sum([(random.choices(num_bacteria_released,
                        cum_weights = Rcum)[0]-1) for i in range(kj)]), -kj

# A function to determine which reactions are critical given the current state
# of the process
def Critical_Reactions(nT,nP):
    critical_reaction_list = []
    if 0 < (nT-nP) < nC:
        if nP < nC:
            critical_reaction_list.extend([0,1,2])
        else:
            critical_reaction_list.extend([0,1])
    elif 0 < nP < nC:
        critical_reaction_list.append(2)
    return critical_reaction_list

def Finding_Tau_Dash(nT,nP):
    non_critical_reactions = [x for x in events
                                if x not in Critical_Reactions(nT,nP)]
    if non_critical_reactions == []:
        # all reactions are critical
        tau_dash = tmax
    elif all(rates[i] == 0 for i in non_critical_reactions):
        # all reactions are critical
        tau_dash = tmax
    else:
        if non_critical_reactions == [2]:
            # both extracellular and phagocytosis are critical
            mu_list = [delta*nP*sum_k_minus_1_Rk, -delta*nP]
            sigma_squared_list = [delta*nP*sum_k_minus_1_Rk_squared, delta*nP]
        elif non_critical_reactions == [0,1]:
            # rupture is the only critical reaction
            mu_list = [-mu*(nT-nP), alpha*(nT-nP)]
            sigma_squared_list = [mu*(nT-nP), alpha*(nT-nP)]
        elif non_critical_reactions == [0,1,2]:
            # no reactions are critical
            mu_list = [delta*nP*sum_k_minus_1_Rk - mu*(nT-nP),
                                    -delta*nP + alpha*(nT-nP)]
            sigma_squared_list = [delta*nP*sum_k_minus_1_Rk_squared+mu*(nT-nP),
                                    delta*nP + alpha*(nT-nP)]
        Ttau = min(max(epsilon*abs(nT),1)/abs(mu_list[0]),
                        max(epsilon*abs(nT),1)**2/sigma_squared_list[0])
        Ptau = min(max(epsilon*abs(nP),1)/abs(mu_list[1]),
                        max(epsilon*abs(nP),1)**2/sigma_squared_list[1])
        return min(Ttau,Ptau)
```

184

```python
def Finding_Tau_Double_Dash(nT,nP):
    if Critical_Reactions(nT,nP) == []:
        tau_double_dash = tmax
    else:
        rates = Make_Rates(nT,nP)
        tau_double_dash = random.expovariate(
                         sum([rates[i] for i in Critical_Reactions(nT,nP)]))
    return tau_double_dash


# A function that determines how many bacteria are released when a rupture
# event occurs
def Choose_Rupture_Event(num_bac,num_phag):
    reaction = random.choices(num_bacteria_released, cum_weights=Rcum)[0]
    num_bac += reaction - 1
    num_phag -= 1
    return num_bac,num_phag


# A function that peforms one step of the Gillespie algorithm for when it is
# not advantageous to perform tau-leaping.
def Gillespie_Step(nT,nP,tt):
    if nT == 0:
        return nT, nP, tt
    else:
        rates = Make_Rates(nT,nP)
        sum_rates = sum(rates)
        tt += -math.log(random.random())/sum_rates
        u1 = random.random()
        if u1 < rates[0]/sum_rates:
            nT -= 1
        elif u1 < (rates[0]+rates[1])/sum_rates:
            nP += 1
        else:
            nT,nP = Choose_Rupture_Event(nT,nP)
        return nT, nP, tt

# A function that performs one step of the tau-leaping algorithm
def One_Step(nT,nP,tt):
    state_change = [0,0]
    rates = Make_Rates(nT,nP)
    tau_dash = Finding_Tau_Dash(nT,nP)
    if tau_dash < 10/sum(rates):
        for j in range(100):
            nT, nP, tt = Gillespie_Step(nT,nP,tt)
        return nT, nP, tt
    else:
        tau_double_dash = Finding_Tau_Double_Dash(nT,nP)
        tau = min(tau_dash, tau_double_dash)
```

```python
        for i in [x for x in events if x not in Critical_Reactions(nT,nP)]:
            state_change = numpy.add(state_change,
                        events[i](numpy.random.poisson(tau*rates[i])))
        if tau == tau_double_dash:
            critical = Critical_Reactions(nT,nP)
            critical_reaction_rates = [rates[i] for i in critical]
            state_change = numpy.add(state_change,
                events[random.choices(critical,
                cum_weights=numpy.cumsum(critical_reaction_rates))[0]](1))
        return nT + state_change[0], nP + state_change[1], tt + tau


# Defining parameters
N = 10  # maximum number of bacteria released from a rupturing phagocyte
num_bacteria_released = [i for i in range(1,N+1)]
R = [1/N for _ in range(N)]# distribution of number of bacteria released
Rcum = numpy.cumsum(R)

M = 5*1e3        # the number of bacteria required for a response to occur
alpha = 0.3325        # rate of phagocytosis resulting in bacterial survival
mu = 8.63   # rate of extracellular bacterial death
delta = 0.0241  # total rate at which an infected phagocyte ruptures

nC = 10  # the threshold used to define the critical region
epsilon = 0.03 # the error in the tau-leaping algorithm
tmax = 1e10 # the maximum time for which the tau-leaping algorithm can run for
numruns = 1 # the number of realisations of the Markov process to be simulated
T0 = 1e3 # the initial number of Francisella tularensis bacteria
P0 = 0 # the initial number of infected phagocytes

events = {0:ExDeath, 1:Phagocytosis, 2:Rupture}
sum_k_minus_1_Rk = sum((num_bacteria_released[k]-1)*R[k] for k in range(N))
sum_k_minus_1_Rk_squared = sum(((num_bacteria_released[k]-1)**2)*R[k]
                                            for k in range(N))

# Constructing one realisation of the within-host model
for _ in range(numruns):
    T = [T0]
    P = [P0]
    tt = [0]
    while tt[-1] < tmax:
        if T[-1] == 0 or T[-1] >= M:
            # the process has reached an absorbing
            tt.append(tmax)
        else:
            rates = Make_Rates(T[-1],P[-1])
            single_step = One_Step(T[-1],P[-1],tt[-1])
            T.append(single_step[0])
            P.append(single_step[1])
            tt.append(single_step[2])
```

A.2 Python code for simulating the ABM

This Python code can be used to simulate the agent based model (ABM) detailed in Section 4.2 and reproduce results such as the size of each cohort.

```python
import copy, numpy as np
from random import choice, sample

# Defining the classes for bacteria and macrophages
class Bacterium:
    def __init__(self,location):
        self.l=location   # initial location
        self.cohort=0     # initial bacteria have cohort number = 0
        self.int=0
    def tophag(self):
        ''' move bacterium to phagosome and increase cohort number '''
        self.int = 'p'
        self.cohort += 1
    def tocyt(self):
        ''' move bacterium to cytosol '''
        self.int = 'c'
    def toorgan(self,place):
        self.l = place
        self.cohort = 0

class Cell(object):
    number = 0
    def __init__(self):
        type(self).number += 1
    def __del__(self):
        type(self).number -= 1

class Mphi(Cell):
    def __init__(self,location):
        Cell.__init__(self)
        self.a = 0, self.b = 0, self.c = 0
        self.cohort_num = [] # cohort number for classifying rupture events
        self.bac = []        # bacteria contained inside
        self.l=location      # initial location
    def phagb(self,nb):
        ''' Macrophage ingests nb bacteria and becomes suppressed '''
        if self.a != 1:
            self.b += nb
            self.a = -1
    def activate(self):
        ''' macrophage is activated '''
        self.a = 1
```

```python
# Defining the reactions
def MeatBn(n,place,FreeBlist,Mlist,time):
    ''' macrophages eat n bacteria '''
    sublist=[i for i,b in enumerate(FreeBlist) if b.l==place]
    movelist = list(set(mysample(sublist,n)))
    killlist = []
    n = len(movelist)
    Msublist=[i for i,M in enumerate(Mlist) if M.l==place]
    if n > len(Msublist)/10:
        Mphaglist = np.bincount([choice(Msublist) for i in range(n)])
        for ind,nb in enumerate(Mphaglist):
            if nb > 0:
                Mlist[ind].phagb(nb)
                phagindlist = np.random.choice(movelist,nb,replace=False)
                baclist = [FreeBlist[k] for k in phagindlist]
                killlist.extend(baclist)
                for i in baclist: i.tophag()
                if Mlist[ind].a != 1:
                    Mlist[ind].bac.extend(baclist)
                    Mlist[ind].cohort_num.append(min([i.cohort for i in baclist]))
                movelist = [i for i in movelist if i not in phagindlist]
    else:
        for ind in [choice(Msublist) for i in range(n)]:
            Mlist[ind].phagb(1)
            phagindlist = np.random.choice(movelist,1,replace=False)
            baclist = [FreeBlist[k] for k in phagindlist]
            killlist.extend(baclist)
            for i in baclist: i.tophag()
            if Mlist[ind].a != 1:
                Mlist[ind].bac.extend(baclist)
                Mlist[ind].cohort_num.append(min([i.cohort for i in baclist]))
            movelist = [i for i in movelist if i not in phagindlist]
    return [i for i in FreeBlist if i not in killlist], Mlist


def Bdien(n,place,FreeBlist,Mlist,time):
    ''' n extracellular bacteria die '''
    sublist=[i for i,b in enumerate(FreeBlist) if b.l==place]
    killlist = set(mysample(sublist,n))
    return [FreeBlist[i] for i in range(len(FreeBlist))
                                if i not in killlist], Mlist


def Bescapen(n,place,FreeBlist,Mlist,time):
    ''' n bacteria escape from the phagosome into the cytosol '''
    thislist = [M for M in Mlist if M.l==place]
    Mbcum = np.cumsum([len([M.bac[i] for i in range(len(M.bac))
                                if M.bac[i].int=='p']) for M in thislist])
    weights = np.insert(Mbcum/(1.0*Mbcum[-1]),0,0.0)
    counts = np.histogram(np.random.random_sample(size=n), bins=weights)[0]
    for i,thism in enumerate(thislist):
```

```
        thisnb = counts[i]
        # if the number moving to the cytosol is less than that in the phagosome
        if thisnb < len([thism.bac[k] for k in range(len(thism.bac))
                                if thism.bac[k].int=='p']):
            for j in sample([thism.bac[q] for q in range(len(thism.bac))
                                    if thism.bac[q].int=='p'],thisnb): j.tocyt()
            thism.b -= thisnb
            thism.c += thisnb
        else:
            for j in thism.bac:
                if j.int=='p':
                    j.tocyt()
            thism.c += thism.b
            thism.b = 0
    return FreeBlist,Mlist


def Bbirthn(n,place,FreeBlist,Mlist,time):
    ''' n bacteria reproduce '''
    thislist = [M for M in Mlist if M.l==place]
    Mccum = np.cumsum([len([M.bac[i] for i in range(len(M.bac))
                            if M.bac[i].int=='c']) for M in thislist])
    weights = np.insert(Mccum/(1.0*Mccum[-1]),0,0.0)
    counts = np.histogram(np.random.random_sample(size=n), bins=weights)[0]
    for i,thism in enumerate(thislist):
        thism.c += counts[i]
        if counts[i] > 0:
            thism.bac += [copy.deepcopy(k) for k in
                np.random.choice([thism.bac[q] for q in range(len(thism.bac))
                    if thism.bac[q].int=='c'],size=counts[i],replace=True)]
    return FreeBlist,Mlist


def Mruptn(n,place,FreeBlist,Mlist,time):
    ''' n macrophages rupture - neighbouring macrophages activated '''
    thisindlist = [i for i in range(len(Mlist)) if Mlist[i].l==place]
    biglist = []
    for i in thisindlist:
        biglist.extend([i]*len(Mlist[i].bac))
    ruptlist = set(sample(biglist,n))
    nbreleased = []
    nbreleased.extend(Mlist[i].bac[j]
            for i in ruptlist for j in range(len(Mlist[i].bac)))
    for i in mysample(thisindlist,n):  # indices of activating M
        if Mlist[i].a==0: Mlist[i].activate()
    Mlist = [Mlist[i] for i in range(len(Mlist)) if i not in ruptlist]
    return FreeBlist + nbreleased, Mlist


def Bmigraten(n,thisplace,FreeBlist,Mlist,time):
    ''' n bacteria migrate to a different compartment '''
    sublist=[i for i,b in enumerate(FreeBlist) if b.l==thisplace]
```

```python
    movelist = mysample(sublist,n)
    n = len(movelist)
    if n>0:
        sumsizes = sum([sizes[place] for place in locations])
        bs = np.random.binomial(n,sizes['S']/sumsizes)
        bp = np.random.binomial(n,sizes['P']/sumsizes)
        bk = np.random.binomial(n,sizes['K']/sumsizes)
        bmln = np.random.binomial(n,sizes['MLN']/sumsizes)
        for ib in movelist[:bs]:
            FreeBlist[ib].toorgan('S')
        for ib in movelist[bs:bs+bp]:
            FreeBlist[ib].toorgan('P')
        for ib in movelist[bs+bp:bs+bp+bk]:
            FreeBlist[ib].toorgan('K')
        for ib in movelist[bs+bp+bk:bs+bp+bk+bmln]:
            FreeBlist[ib].toorgan('MLN')
        for ib in movelist[bs+bp+bk+bmln:]:
            FreeBlist[ib].toorgan('L')
    return FreeBlist, Mlist


def Mdownn(n,place,FreeBlist,Mlist,time):
    ''' n macrophages are suppressed '''
    sublist = [i for i in range(len(Mlist)) if Mlist[i].l==place
                                            and Mlist[i].a==0]
    alist = mysample(sublist,n)
    for i in alist:
        Mlist[i].a = -1
    return FreeBlist, Mlist


def Mupn(n,place,FreeBlist,Mlist,time):
    ''' n macrophages become activated '''
    sublist = [i for i in range(len(Mlist)) if Mlist[i].l==place
                                            and Mlist[i].a==0]
    if len(sublist) !=0:
        alist = mysample(sublist,n)
        for i in alist:
            Mlist[i].activate()
    return FreeBlist, Mlist


def ifn(x):
    ''' rate of macrophage activation as function of IFN-gamma level '''
    return 1 if x > 100 else 0


def tgf(x):
    ''' rate of macrophage suppression as a function of TFG-beta level '''
    return 1 if x > 100 else 0


def mysample(list,n):
    return sample(list,min(n,len(list)))
```

190

```
def makerates(FreeBlist,Mlist,G,T):
    myrates = []
    for place in locations:
        nb = len([b for b in FreeBlist if b.l==place])
        thismlist = [M for M in Mlist if M.l==place]
        nm = len(thismlist)
        mb = sum([len([M.bac[i] for i in range(len(M.bac))
             if M.bac[i].int=='p']) for M in thismlist])  # phagosomal bacteria
        mc = sum([len([M.bac[i] for i in range(len(M.bac))
             if M.bac[i].int=='c']) for M in thismlist])  # cytosolic bacteria
        myrates.append(rho*nb*nm)              # phagocytosis rates
        myrates.append(mu*nb)                  # death rates
        myrates.append(mb*phi)                 # exit rates
        myrates.append(mc*beta)                # division rates
        myrates.append(mc*delta)               # rupture rates
        myrates.append(nb*gamma)               # migration rates
        myrates.append(tgf(T[place])*nu)       # macrophage becomes suppressed
        myrates.append(ifn(G[place])*nu)       # macrophage becomes activated
    return myrates

def writestuff(t,nevents,Mlist,FreeBlist,G,T):
    for place in locations:
        thismlist = [M for M in Mlist if M.l==place]  # no. macrophages in place
        for cohort_num in range(1,5):
            cofile.write(str(sum([len([M.bac[i] for i in range(len(M.bac)) if
                        (M.bac[i].int=='p' and M.bac[i].cohort==cohort_num)])
                            for M in thismlist]))+' ')
            cofile.write(str(sum([len([M.bac[i] for i in range(len(M.bac)) if
                        (M.bac[i].int=='c' and M.bac[i].cohort==cohort_num)])
                            for M in thismlist]))+' ')
    cofile.write('\n')
    cofile.flush()

def onestep(dt,t,FreeBlist,Mlist,G,T):
    ''' onestep of the tau-leaping algorithm '''
    rates = makerates(FreeBlist,Mlist,G,T)
    mevs = [dt*rate for rate in rates]  # mean numbers of events
    nevents = np.random.poisson(lam = mevs)
    for i,thisn in enumerate(nevents):
        place = locations[i // len(events)] # lung, liver, spleen, kidney, MLN
        if thisn > 0:
            FreeBlist,Mlist = events.get(i\%len(events))
                                        (thisn,place,FreeBlist,Mlist,t)
        alist = [M for M in Mlist if M.a==1]
        if len(alist)!=0:
            for place in locations:
                G[place] += dt*(len([M for M in alist if M.l==place])
                                            - G[place]*IFN_decay)
```

```
        slist = [M for M in Mlist if M.a==-1]
        if len(slist)!=0:
            for place in locations:
                T[place] += dt*(len([M for M in slist if M.l==place])
                                            - T[place]*TGF_decay)
    if int((t-dt/2)*10) != int((t+dt/2)*10):
        writestuff(t,nevents,Mlist,FreeBlist,G,T)
    return FreeBlist,Mlist,G,T,t+dt

def onerun(dt,thisN):
    FreeBlist = []
    Mlist = []
    for _ in range(thisN):
        FreeBlist.append(Bacterium('P'))
    for _ in range(int(M)):
        Mlist.append(Mphi('P'))
        Mlist.append(Mphi('L'))
    for _ in range(int(M/10)):
        Mlist.append(Mphi('S'))
        Mlist.append(Mphi('K'))
        Mlist.append(Mphi('MLN'))
    G = {}
    T = {}
    for place in locations:
        G[place]=0.0
        T[place]=0.0
    t = 0.0
    while t < tmax:
        FreeBlist,Mlist,G,T,t = onestep(dt,t,FreeBlist,Mlist,G,T)

# Running the algorithm
N = 10  # mean initial number of CFU
M = 1e4  # initial number of macrophages in lung compartment

rho, phi, gamma, beta, delta, mu, nu = 0.01, 2, 1, 0.15, 0.001, 0.01, 0.01
IFN_decay, TGF_decay = 0.01, 0.1

# tau-leap step, end time (hours)
dt, tmax = 0.1, 24
locations=('P','L','S','K','MLN')
sizes={'P':0,'S':0.05,'L':0.11,'K':0.04,'MLN':0.8}
events={0:MeatBn,1:Bdien,2:Bescapen,
                    3:Bbirthn,4:Mruptn,5:Bmigraten,6:Mdownn,7:Mupn}

cofile = open('cohort.dat','w')
thisCFU = np.random.poisson(lam=N)
onerun(dt,thisCFU)
cofile.close()
```

References

ALLEN, L. (2003). *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Prentice Hall, Upper Saddle River, N.J. 10, 12, 34, 35, 174

ALLEN, L.A.H. (2013). Neutrophils: potential therapeutic targets in tularemia? *Frontiers in Cellular and Infection Microbiology*, **3**, 109. 4, 123

ALTHAUS, C.L. (2014). Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS Currents*, **6**. 129

ATHALE, C.A. & DEISBOECK, T.S. (2006). The effects of EGF-receptor density on multiscale tumor growth patterns. *Journal of Theoretical Biology*, **238**, 771-779. 99

BACCAM, P., BEAUCHEMIN, C., MACKEN, C.A., HAYDEN, F.G., & PEREL-SON, A.S. (2006). Kinetics of Influenza A Virus Infection in Humans. *Journal of Virology*, **80**, 7590–7599. 127

BARDOEL, B.W., KENNY, E.F., SOLLBERGER, G. & ZYCHLINSKY, A. (2014). The Balancing Act of Neutrophils. *Cell Host & Microbe*, **15**, 526–536. 1

BARR, J.N. & FEARNS, R. (2010). How RNA viruses maintain their genome integrity. *Journal of General Virology*, **91**, 1373–1387. 5

BAUER, A.L., BEAUCHEMIN, C.A.A. & PERELSON, A.S. (2009). Agent-based modeling of host–pathogen systems: The successes and challenges. *Information Sciences*, **179**, 1379–1389. 29

Best, K. & Perelson, A.S. (2018). Mathematical modeling of within-host Zika virus dynamics. *Immunological Reviews*, **286**, 81–96. 127, 135, 141

Bosio, C.M., Bielefeldt-Ohmann, H. & Belisle, J.T. (2007). Active Suppression of the Pulmonary Immune Response by *Francisella tularensis* Schu4. *The Journal of Immunology*, **178**, 4538–4547. 116

Brock, S. & Parmely, M. (2017). Complement C3 as a Prompt for Human Macrophage Death during Infection with *Francisella tularensis* Strain SCHU S4. *Infection and Immunity*, **85**, e00424–17. 36, 70, 76

Brown, S.P., Cornell, S.J., Sheppard, M., Grant, A.J., Maskell, D.J., Grenfell, B.T. & Mastroeni, P. (2006). Intracellular Demography and the Dynamics of Salmonella enterica Infections. *PLoS Biology*, **4**, e349. 80, 81, 109

Butler, K. & Stephens, M.A. (2017). The Distribution of a Sum of Independent Binomial Random Variables. *Methodology and Computing in Applied Probability*, **19**, 557–571. 74

Calain, P., Monroe, M.C. & Nichol, S.T. (1999). Ebola Virus Defective Interfering Particles and Persistent Infection. *Virology*, **262**, 114–128. 181

Cao, Y., Gillespie, D. & Petzold, L. (2006). Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, **124**, 044109. 23

Caswell, H. (2011). Perturbation analysis of continuous-time absorbing Markov chains. *Numerical Linear Algebra with Applications*, **18**, 901–917. 56

Center for Preparedness and Response (2017). Bioterrorism Agents/Diseases, https://emergency.cdc.gov/agent/agentlist-category.asp, [accessed 17-12-2018]. 2

Centers for Disease Control and Prevention (2017). Case Counts, https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html, [accessed 19-12-2018]. 4

REFERENCES

CHEN, D. (2007). Dose-Time-Response Cumulative Multinomial Generalized Linear Model. *Journal of Biopharmaceutical Statistics*, **17**, 173–185. 32

CHEN, H.Y., DI MASCIO, M., PERELSON, A.S., HO, D.D. & ZHANG, L. (2007). Determination of virus burst size *in vivo* using a single-cycle SIV in rhesus macaques. *Proceedings of the National Academy of Sciences*, **104**, 19079–19084. 80

CHRISTOPHER, LTC G.W., CIESLAK, LTC T.J., PAVLIN, J.A. & EITZEN, E.M. (1997). Biological Warfare: A Historical Perspective. *JAMA*, **278**, 412–417. 2

COWLEY, S. & ELKINS, K. (2011). Immunity to *Francisella*. *Frontiers in Microbiology*, **2**. 34

DAY, J., FRIEDMAN, A. & SCHLESINGER, L.S. (2011). Modeling the host response to inhalation anthrax. *Journal of Theoretical Biology*, **276**, 199–208. 80

D'ELIA, R.V., LAWS, T.R., CARTER, A., LUKASZEWSKI, R. & CLARK, G.C. (2013). Targeting the Rising DAMP during a *Francisella tularensis* infection. *Antimicrobial Agents and Chemotherapy*, **57**, 4222-4228. 4, 102

DI CRESCENZO, A., GIORNO, V., NOBILE, A. & RICCIARDI, L. (2008). A note on birth–death processes with catastrophes. *Statistics and Probability Letters*, **78**, 2248–2257. 34

DIXIT, N.M. & PERELSON, A.S. (2004). Multiplicity of Human Immunodeficiency Virus Infections in Lymphoid Tissue. *Journal of Virology*, **78**, 8942–8945. 93

DUVIGNEAU, S., DÜRR, R., LASKE, T., BACHMANN, M., DOSTERT, M., REICHL, U. & KIENLE, A. (2018). Mathematical modeling as a tool to improve influenza vaccine production processes. *IFAC-PapersOnLine*, **51**, 1–4. 128

Eigelsbach, H., Tulis, J., McGavran, M. & White, J. (1962). Live tularemia vaccine I. host-parasite relationship in monkeys vaccinated intracutaneously or aerogenically. *Journal of Bacteriology*, **84**, 1020–1027. xiii, 62, 63, 64, 65

Gani, J. & Swift, R.J. (2007). Death and Birth–Death and Immigration Processes with Catastrophes. *Journal of Statistical Theory and Practice* . **1**, 39–48. 85

Gillard, J.J., Laws, T.R., Lythe, G. & Molina-París, C. (2014). Modeling early events in *Francisella tularensis* pathogenesis. *Frontiers in Cellular and Infection Microbiology*, **4**, 169. xiv, 87, 92, 93, 104, 118

Gillespie, D. (2007). Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, **58**, 35–55. 22

Golovliov, I., Baranov, V., Krocova, Z., Kovarova, H. & Sjöstedt, A. (2003). An Attenuated Strain of the Facultative Intracellular Bacterium *Francisella tularensis* Can Escape the Phagosome of Monocytic Cells. *Infection and Immunity*, **71**, 5940–5950. 35, 39

Gómez-Corral, A. & López-García, M. (2012). Extinction times and size of the surviving species in a two-species competition process. *Journal of Mathematical Biology*, **64**, 255–289. 33

Gómez-Corral, A. & López-García, M. (2015). Lifetime and reproduction of a marked individual in a two-species competition process. *Applied Mathematics and Computation*, **264**, 223–245. 42

Gómez-Corral, A. & López-García, M. (2018). Perturbation analysis in finite LD-QBD processes and applications to epidemic models. *Numerical Linear Algebra with Applications*, e2160. 56

Gorochowski, T.E., Matyjaszkiewicz, A., Todd, T., Oak, N., Kowalska, K. & Reid, S. (2012). BSim: An Agent-Based Tool for Modeling Bacterial Populations in Systems and Synthetic Biology. *PloS One*, **7**, e42790. 29

REFERENCES

HAAS, C.N., ROSE, J.B. & GERBA, C.P. (1999). *Quantitative microbial risk assessment*. Wiley, New York. 32

HALL, J.D., WOOLARD, M.D., GUNN, B.M., CRAVEN, R.R., TAFT-BENZ S., FRELINGER, J.A. *et al.* (2008). Infected-Host-Cell Repertoire and Cellular Response in the Lung following Inhalation of *Francisella tularensis* Schu S4, LVS, or U112. *Infection and Immunity*, **76**, 5843–5852. 3, 4, 105

HASELTINE, E.L., RAWLINGS, J.B. & YIN, J. (2005). Dynamics of viral infections: incorporating both the intracellular and extracellular levels. *Computers & Chemical Engineering*, **29**, 675–686. 128

HE, Q.M. (2014). *Fundamentals of Matrix-Analytic Methods*. Springer, New York. 12, 16, 37

HODGES, D.R., KLEIN, F., MAHLANDT, B.G., JONES JR, W.I., HAINES, B.W., RHIAN, M.A. & WALKER, J.S. (1965). Role of the Lymphatics in the Pathogenesis of Anthrax. *The Journal of Infectious Diseases*, 481–494. 116

HOEBE, K., JANSSEN, E. & BEUTLER, B. (2004). The interface between innate and adaptive immunity. *Nature Immunology*, **5**, 971. 2

HUANG, Y. & HAAS, C. (2009). Time-Dose-Response Models for Microbial Risk Assessment. *Risk Analysis*, **29**, 648–661. 31, 32

IWAMI, S., SATO, K., DE BOER, R.J., AIHARA, K., MIURA, T. & KOYANAGI, Y. (2012). Identifying viral parameters from *in vitro* cell cultures. *Frontiers in Microbiology*, **3**, 319. 130

IWAMI, S., HOLDER, B.P., BEAUCHEMIN, C.A.A., MORITA, S., TADA, T., SATO, K., IGARASHI, T. & MIURA, T. (2012). Quantification system for the viral dynamics of a highly pathogenic simian/human immunodeficiency virus based on an *in vitro* experiment and a mathematical model. *Retrovirology*, **9**, 18. 136, 149, 165

JONES, C., NAPIER, B., SAMPSON, T., LLEWELLYN, A., SCHROEDER, M. & WEISS, D. (2012). Subversion of Host Recognition and Defense Systems by *Francisella* spp. *Microbiology and Molecular Biology Reviews*, **76**, 383–404. xxii, 3, 46, 101, 106, 116, 117, 122

KARLIN, S. & TAVARÉ, S. (1982). Linear birth and death processes with killing. *Journal of Applied Probability*, **19**, 477–487. 34, 88, 89, 91

KIRBY, A.C., COLES, M.C. & KAYE, P.M. (2009). Alveolar Macrophages Transport Pathogens to Lung Draining Lymph Nodes. *The Journal of Immunology*, **183**, 1983–1989. 109

KNIPE, D.M. & HOWLEY, P.M. (2007). *Fields Virology*. Lippincott Williams & Wilkins, Philadelphia, PA. 5, 8, 130

KOELLE, K., FARRELL, A., BROOKE, C. & KE, R. (2018). Within-host infectious disease models accommodating cellular coinfection, with an application to influenza. *bioRxiv*. 128

KOLACZKOWSKA, E. & KUBES, P. (2013). Neutrophil recruitment and function in health and inflammation. *Nature Reviews Immunology*, **13**, 159. 2

LATOUCHE, G. & RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics, Philadelphia, PA. 12

LIAO, C.M., CHANG, C.F. & LIANG, H.M. (2005). A Probabilistic Transmission Dynamic Model to Assess Indoor Airborne Infection Risks. *Risk Analysis*, **25**, 1097–1107. 53

LINDEMANN, S., PENG, K., LONG, M., HUNT, J., APICELLA, M. & MONACK, D.M. (2011). *Francisella tularensis* Schu S4 O-Antigen and Capsule Biosynthesis Gene Mutants Induce Early Cell Death in Human Macrophages. *Infection and Immunity*, **79**, 581–594. xii, 36, 59, 60, 61

REFERENCES

López-García, M., King, M.F. & Noakes, C. (2019). A multi-compartment SIS stochastic model with zonal ventilation for the spread of nosocomial infections: detection, outbreak management and infection control. *Risk Analysis*, **Under Review**. xxi, 53, 65, 67

Lowrie, D.B., Aber, V.R. & Carrol, M.E.W. (1979). Division and Death Rates of *Salmonella typhimurium* Inside Macrophages: Use of Penicillin as a Probe. *Microbiology*, **110**, 409–419. xxii, 122

Madelain, V., Oestereich, L., Graw, F., Nguyen, T.H., De Lamballerie, X., Mentré, F., Günther, S. & Guedj, J. (2015). Ebola virus dynamics in mice treated with favipiravir. *Antiviral Research*, **123**, 70–77. 129

Madelain, V., Baize, S., Jacquot, F., Reynard, S., Fizet, A., Barron, S., Solas, C., Lacarelle, B., Carbonnelle, C., Mentré, F., et al. (2018). Ebola viral dynamics in nonhuman primates provides insights into virus immuno-pathogenesis and antiviral strategies. *Nature Communications*, **9**, 4013. 129

Marino, S. & Kirschner, D.E. (2004). The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *Journal of Theoretical Biology*, **227**, 463–486. xxii, 122

Mastroeni, P., Grant, A., Restif, O. & Maskell, D. (2009). A dynamic view of the spread and intracellular distribution of *Salmonella enterica*. *Nature Reviews Microbiology*, **7**, 73. 80

McElroy, A.K., Mühlberger, E. & Munoz-Fontela, C. (2018). Immune barriers of Ebola virus infection. *Current Opinion in Virology*, **28**, 152–160. 5

Messaoudi, I., Amarasinghe, G.K. & Basler, C.F. (2015). Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. *Nature Reviews Microbiology*, **13**, 663. xi, 6

MITTLER, J.E., SULZER, B., NEUMANN, A.U. & PERELSON, A.S. (1998). Influence of delayed viral production on viral dynamics in HIV-1 infected patients. *Mathematical Biosciences*, **152**, 143–163. 135

MOSSER, D.M. & EDWARDS, J.P. (2008). Exploring the full spectrum of macrophage activation. *Nature Reviews Immunology*, **8**, 958. 100

MURPHY, K. & WEAVER, C. (2016). *Janeway's Immunobiology*. Garland Science, New York. 1

MURRAY, P.J. & WYNN, T.A. (2011). Protective and pathogenic functions of macrophage subsets. *Nature Reviews Immunology*, **11**, 723. 1

NELSON, P.W., GILCHRIST, M.A., COOMBS, D., HYMAN, J.M. & PERELSON, A.S. (2004). An age-structured model of HIV infection that allows for variations in the production rate of viral particles and the death rate of productively infected cells. *Mathematical Biosciences and Engineering*, **1**, 267–288. 136

NOAKES, C. & SLEIGH, P. (2009). Mathematical models for assessing the role of airflow on the risk of airborne infection in hospital wards. *Journal of the Royal Society Interface*, rsif20090305. xxi, 53, 54, 58, 65, 66, 67

NOWAK, M.A. & MAY, R.M. (2000). *Virus dynamics: mathematical principles of immunology and virology*. Oxford University Press, Oxford, UK. 127, 170

NGUYEN, V.K., BINDER, S.C., BOIANELLI, A., MEYER-HERMANN, M. & HERNANDEZ-VARGAS, E.A. (2015). Ebola virus infection modeling and identifiability problems. *Frontiers in Microbiology*, **6**, 257. 129, 146, 176

OSOGAMI, T. & HARCHOL-BALTER, M. (2006). Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluations*, **63**, 524–552. 16, 17

OYSTON, P., SJÖSTEDT, A. & TITBALL, R. (2004). Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nature Reviews Microbiology*, **2**, 967–978. xi, 3, 71, 80

REFERENCES

OYSTON, P. (2008). *Francisella tularensis*: unravelling the secrets of an intracellular pathogen. *Journal of Medical Microbiology*, **57**, 921–930. 2

PARKER, S. & WILLIAMSON, S. (2016). Visual assessment of contaminant impacts in multizone buildings. *Building and Environment*, **102**, 39–53. 54

PEARSON, J.E., KRAPIVSKY, P. & PERELSON, A.S. (2011). Stochastic Theory of Early Viral Infection: Continuous versus Burst Production of Virions. *PLoS Computational Biology*, **7**, e1001058. 80

PERELSON, A.S. (2002). Modelling viral and immune system dynamics. *Nature Reviews Immunology* , **2**, 28 127

PINILLA, L.T., HOLDER, B.P., ABED, Y., BOIVIN, G. & BEAUCHEMIN, C.A.A. (2012). The H275Y Neuraminidase Mutation of the Pandemic A/H1N1 Influenza Virus Lengthens the Eclipse Phase and Reduces Viral Output of Infected Cells, Potentially Compromising Fitness in Ferrets. *Journal of Virology*, **86**, 10651–10660. 135, 176

PORNILLOS, O., GARRUS, J.E. & SUNDQUIST, W.I. (2002). Mechanisms of enveloped RNA virus budding. *Trends in Cell Biology* , **12**, 569–579. 79

PRESCOTT, J.B., MARZI, A., SAFRONETZ, D., ROBERTSON, S.J., FELDMANN, H. & BEST, S.M. (2017). Immunobiology of Ebola and Lassa virus infections. *Nature Reviews Immunology*, **17**, 195. 4, 6

PUJOL, J., EISENBERG, J., HAAS, C. & KOOPMAN, J. (2009). The Effect of Ongoing Exposure Dynamics in Dose Response Relationships. *PLoS Computational Biology*, **5**, e1000399. 32

RACHAH, A. & TORRES, D.F.M. (2015). Mathematical Modelling, Simulation, and Optimal Control of the 2014 Ebola Outbreak in West Africa. *Discrete Dynamics in Nature and Society*, **2015**. 129

RAILSBACK, S.F. & VOLKER G. (2012). *Agent-Based and Individual-Based Modeling*. Princeton University Press, Princeton, N.J. 103

RENSHAW, E. (2011). *Stochastic Population Processes*. Oxford University Press, Oxford, UK. 12, 20, 90

RIBEIRO, R.M., LO, A. & PERELSON, A.S. (2002). Dynamics of hepatitis B virus infection. *Microbes and Infection*, **4**, 829–835. 127

ROSSMAN, J.S. & LAMB, R.A. (2011). Influenza virus assembly and budding. *Virology* , **411**, 229–236. 79

SALTELLI, A., TATANTOLA, S., CAMPOLONGO, F. & RATTO, M. (2004). *Sensitivity Analysis in Practice*. John Wiley & Sons Ltd, Chichester, UK. 27

SANTIC, M., PAVOKOVIC, G., JONES, S., ASARE, R. & KWAIK, Y.A. (2010). Regulation of apoptosis and anti-apoptosis signalling by *Francisella tularensis*. *Microbes and Infection*, **12**, 126–134. 36

SASLAW, S., EIGELSBACH, H., PRIOR, J., WILSON, H. & CARHART, S. (1961). Tularemia Vaccine Study II. Respiratory Challenge. *Archives of Internal Medicine*, **107**, 702–714. 62, 72

SAWYER, W., DANGERFIELD, H., HOGGE, A. & CROZIER, D. (1966). Antibiotic Prophylaxis and Therapy of Airborne Tularemia. *Bacteriological Reviews*, **30**, 542–550. 62, 72

SCHWARTZ, J.T., BARKER, J.H., KAUFMAN, J., FAYRAM, D.C., McCRACKEN, J.M. & ALLEN, LA.H (2012). *Francisella tularensis* Inhibits the Intrinsic and Extrinsic Pathways to Delay Constitutive Apoptosis and Prolong Human Neutrophil Lifespan. *The Journal of Immunology*, **201**, 1102863. 4, 122

SFDPH (2008). Tularemia. *San Francisco Department of Public Health (SFDPH)*. 52

SHABRAM, P. & AGUILAR-CORDOVA, E. (2000). Multiplicity of Infection/Multiplicity of Confusion. *Molecular Therapy*, **2**, 420–421. 93

REFERENCES

SHAPIRO, D. & SCHWARTZ, D. (2002). Exposure of Laboratory Workers to *Francisella tularensis* despite a Bioterrorism pPocedure. *Journal of Clinical Microbiology*, **40**, 2278–2281. 52

SIDORENKO, Y. & REICHL, U. (2004). Structured model of influenza virus replication in MDCK cells. *Biotechnology and Bioengineering*, **88**, 1–14. 127

SOMPAYRAC, L. (2008). *How the Immune System Works*. Blackwell Publishing, Malden, MA. 1

SMITHER, S.J., LEAR-ROONEY, C., BIGGINS, J., PETTITT, J., LEVER, M.S. & OLINGER JR, G.G. (2013). Comparison of the plaque assay and 50% tissue culture infectious dose assay as methods for measuring filovirus infectivity. *Journal of Virological Methods*, **193**, 565–571. 130

SOBOL, I. (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments*, **1**, 407–414. 27, 28

SRIVASTAVA, R., YOU, L., SUMMERS, J. & YIN, J. (2002). Stochastic vs. Deterministic Modeling of Intracellular Viral Kinetics. *Journal of Theoretical Biology*, **218**, 309–321. 127

STROBER, W. (2015). Trypan Blue Exclusion Test of Cell Viability. *Current Protocols in Immunology*, **111**, A3–B. 131

TAO, W., GAN, T., GUO M., XU Y. & ZHONG, J. (2017). Novel Stable Ebola Virus Minigenome Replicon Reveals Remarkable Stability of the Viral Genome. *Journal of Virology*, **91**, e01316–17. 8

TAYLOR, H. & KARLIN, S. (1998). *An Introduction To Stochastic Modeling*. Academic Press, San Diego, CA. 10, 12

TONI, T., WELCH, D., STRELKOWA, N., IPSEN, A. & STUMPF, M. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, **6**, 187–202. 25, 27

TRINCHIERI, G. (2003). Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nature Reviews Immunology*, **3**, 133. 2

UNDERHILL, D.M. & GOODRIDGE, H.S. (2012). Information processing during phagocytosis. *Nature Reviews Immunology*, **12**, 492. 1

VAN DEN DRIESSCHE, P. & WATMOUGH, J. (2008). Further Notes on the Basic Reproduction Number. *Mathematical Epidemiology*, 159–178 170

VAN DOORN, E.A. & ZEIFMAN, A.I. (2005). Birth–death processes with killing. *Statistics & Probability Letters*, **72**, 33–42. 97

WAKELAND, W., MACOVSKY, L. & AN, G. (2007). A Hybrid Simulation Model for Studying Acute Inflammatory Response. *Proceedings of the 2007 Spring Simulation Multiconference-Volume 2*, 39–46. 99

WHITE, J., ROONEY, J., PRICKETT, P., DERRENBACHER, E., BEARD, C. & GRIFFITH, W. (1964). Pathogenesis of Experimental Respiratory Tularemia in Monkeys. *The Journal of Infectious Diseases*, **114**, 277–283. xiii, 62, 63, 64, 65

WICKSTRUM, J., BOKHARI, S., FISCHER, J., PINSON, D., YEH, H.W., HORVAT, R. & PARMELY, M. (2009). *Francisella tularensis* Induces Extensive Caspase-3 Activation and Apoptotic Cell Death in the Tissues of Infected Mice. *Infection and immunity*, **77**, 4827–4836. 36

WOOD, R., EGAN, J.R. & HALL, I. (2014). A dose and time response Markov model for the in-host dynamics of infection with intracellular bacteria following inhalation: with application to *Francisella tularensis*. *Journal of The Royal Society Interface*, **11**. xii, xiii, xiv, 32, 33, 34, 35, 36, 46, 47, 48, 49, 56, 59, 61, 62, 63, 64, 67, 69, 71, 72, 73, 74, 76, 77, 106

YUAN, Y. & ALLEN, L.J.S. (2011). Stochastic models for virus and immune system dynamics. *Mathematical Biosciences*, **234**, 84–94. 135

REFERENCES

ZHANG, X.Y., TRAME, M., LESKO, L. & SCHMIDT, S. (2015). Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT: Pharmacometrics & Systems Pharmacology*, **4**, 69–79. 26