

Stochastic Modelling of Vector-Borne Diseases with Household Structures

Noah Day

October, 2020

Supervisor: Dr Andrew Black

*Thesis submitted for the degree of
Honours of Mathematical Sciences
in*

Applied Mathematics

at The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Mathematical Sciences



THE UNIVERSITY
*of*ADELAIDE

Contents

Declaration	iv
Acknowledgements	v
Abstract	vi
1 Introduction	1
1.1 Vector-Borne Diseases	2
1.2 Epidemic Modelling	2
1.3 Thesis Outline	4
2 Technical Background	5
2.1 Continuous-Time Markov Chains	5
2.2 Path Integrals of Markov Chains	6
2.3 SIR Model	7
2.3.1 Generating the Q Matrix	9
2.4 Basic Reproduction Number, R_0	10
3 Household Model	11
3.1 Branching Process Approximation	12
4 Household Model with Vectors	15
4.1 SEIR-SEI Model	16
4.2 Next-Generation Matrices	20
4.3 Total Infection of a Household	21
4.4 Size-Biased Distribution, π_k	23
4.5 Biting Distribution	24
4.6 Results	27

5 Modelling the Movement of Hosts	30
5.1 Formulation	30
5.2 Methods and Results	32
6 Discussion	36
6.1 Limitations and Further Work	37
6.2 Conclusion	39
Bibliography	40

Declaration

Except where stated this thesis is, to the best of my knowledge, my own work and my supervisor has approved its submission.

Signed by student: Noah Day

Date: October 30, 2020

Signed by supervisor:

Date:

Acknowledgements

I would like to thank my supervisor, Dr Andrew Black, for the time and effort he has put into assisting me with the completion of this thesis. I would also like to thank my friends and family for their support throughout this year.

Abstract

Previous attempts to model the spread of infectious diseases have often assumed that each person in the population is equally susceptible to contracting the disease. However, this assumption may not hold for vector-borne diseases, such as dengue. The reason being that mosquitoes (who are the primary transmitters of the disease) are relatively restricted around the source of water, where they breed. In this thesis, a model is developed that has the potential to more accurately reflect the dynamics of a dengue infection, by confining vectors to households and allowing hosts to mix in the population. A compartmental continuous-time Markov chain model reflecting these dynamics was formulated, which was approximated by a two-type branching process which estimated host mixing with transmission functions. This approximation allowed for simple interpretation as it could be used to calculate epidemiological quantities such as household reproduction number and the early growth rate. The effect of the between household mixing of hosts on the expected number of households infected (per household) was investigated. It was found that each household could accommodate up to twice as many vectors when individuals reduce the amount of time spent in other households from approximately 7 hours to 2.5 hours per day before there is a chance of an epidemic. A second model was then created, this explicitly accounted for the effects of host movement between households. The results provide evidence that the branching process can successfully approximate the prominent dynamics of dengue transmission in the early stages of an epidemic. Such modelling has the potential to inform the further development of public health measures to prevent the spread of dengue in Australia and other countries, and may also have application to efforts to control other vector-borne diseases, such as chikungunya, yellow fever and Zika.

Chapter 1

Introduction

Dengue fever is one of the most common diseases in the world, with the World Health Organisation reporting that approximately 390 million new infections occur every year and approximately half of the world’s population are now classified as “at risk”, with the disease considered endemic in more than 100 countries [15, 28]. Although many of these countries are in tropical and subtropical regions (with South-East Asia seriously affected [16, 28]), the threat of an outbreak in other parts of the world, including Australia, is increasing. For example, local transmission was reported in France and Croatia for the first time in 2010, and in 2012, an outbreak of dengue in the Madeira islands resulted in over 2000 cases which then spread to mainland Portugal and subsequently to at least 10 other European countries [7]. In Australia, there have been a number of cases in Queensland, with two locally contracted cases in Townsville this year [16]. The spread of the disease into tropical areas of Australia is considered inevitable given the growth in tourist and business travel between Australia and South-East Asian countries, even when the travel restrictions that have been introduced in response to the COVID-19 pandemic are considered. For this reason, modelling the way in which dengue virus might spread across Australian cities (such as Townsville) may provide critical information to inform public health responses. Such modelling can, for example, help public health professionals to better understand the level of risk that Australia faces as a result of an outbreak as well as identify those critical factors that influence virus transmission and how an epidemic might best be prevented.

1.1 Vector-Borne Diseases

Vector-borne diseases are infections that are transmitted to a host - in this case humans - by blood-feeding arthropods, such as mosquitoes, ticks, and fleas [25]. Dengue virus is primarily transmitted by the *Aedes aegypti* mosquito, although given that male mosquitoes do not feed on human blood it is only the females that function as vectors [16, 28]. Transmission occurs when an infected mosquito bites a person (to feed on their blood) thereby allowing the virus to enter the bloodstream [28]. There is, however, an (extrinsic) incubation period before the mosquito is able to transmit the disease to another host, which is usually around 10 days [2]. Mosquitoes become infected from people who are viremic. Before becoming viremic, humans also experience a latent period (or intrinsic incubation period) of approximately one week [20].

Of the 390 million new infections of dengue that occur each year, approximately one in four – or 96 million people – will develop clinical symptoms [4]. During the febrile phase the infection is typically associated with a high fever ($>40^{\circ}\text{C}$) accompanied by symptoms such as headaches, joint pains, and nausea. The critical phase of the illness occurs 3 to 7 days after onset and is potentially fatal due to the effect of plasma leaking, fluid accumulation, respiratory distress, severe bleeding, and/or organ impairment [28]. There is currently no specific treatment for dengue fever. Maintaining body fluid volume is critical and analgesics are used to control muscle aches and pains. In 2015, the first vaccine for dengue was licensed and is now approved for administration in approximately 20 different countries [28]. It is however, predominantly used with those who live in endemic areas and who have had at least one previous infection and, as such, not considered to offer a viable control strategy for the transmission of dengue in Australia. Even though recovery from infection is believed to provide lifelong immunity, four distinct variations or serotypes of the virus have been identified. With recovery only providing immunity from the particular serotype involved, with subsequent infections increasing the risk of severe symptoms developing [28].

1.2 Epidemic Modelling

The modelling of epidemics has a long history, with the first presented in the 1920's by Reed and Frost, where they developed a chain-binomial model (which went unpublished). Since then, many extensions upon this model have been made, notably Greenwood [17]. These chain-binomial models are appropriate for infections that are

transmitted from one individual to another within a household [1]. These models are appropriate for within-household transmission dynamics, as they are able to produce final size distributions of the outbreak [33]. Hence, they have been used to identify optimal control strategies for diseases such as tuberculosis [1]. Such epidemic models typically categorise humans as either *susceptible* to the disease, *infected* by the disease or *recovered* from the disease (SIR). However, after an initial dengue virus infection both hosts and vectors enter an incubation period during which the disease cannot be transmitted. Thus, an additional *exposed* class is required for both humans and mosquitoes in the modelling of dengue transmission (i.e., a susceptible-exposed-infected-recovered model for humans and a susceptible-exposed-infected model for mosquitoes). Existing SEIR(human)-SEI(vector) models often assume that the population is well-mixed or homogeneous [2, 14, 20, 21, 23, 27, 30] and are predicated on the assumption that any one individual is as likely to infect any other. However, given that dengue virus is primarily transmitted through human-mosquito interactions and because mosquitoes do not generally travel large distances [29], this assumption may not apply. Rather it seems likely that there is a strong spatial component to transmission (i.e., those who live within close proximity of infected mosquito colonies will experience more interactions than those who do not). For this reason, new models that account for the dynamics of the early stages of the outbreak need to be developed. These would, for example, confine people and mosquitoes to households and consider the impact of host movement between households. Previous household modelling techniques have been used to model human-to-human transmitted infections such as influenza [6, 18, 19, 32]. Such stochastic modelling would capture some of the randomness associated with the possibility of the first infected individual further transmitting the disease to the susceptible population.

We propose to combine the SEIR-SEI compartmental and household models to develop a model designed for vector-borne diseases such as dengue. The construction of this model will provide epidemiological results such as the household reproduction number R_* , and the early rate of growth of the number of cases. As it is unlikely that an individual will be infected by multiple serotypes during the early stages of a dengue outbreak, it is assumed in the model that only a single dengue virus serotype will be present in the population. The parameters of the model are informed by studies on the 2014 outbreak of dengue fever in Guangzhou, China [2, 9, 35]. These parameters values are considered particularly useful as it is plausible that any outbreak of dengue in Australia will have originated from a South-East Asian region, including southern China.

1.3 Thesis Outline

In this thesis a stochastic process is used to model the spread of the pathogen across the newly exposed population. The aim is to demonstrate the effects of different levels of population mixing and varying densities of mosquito populations on dengue virus transmission in a large Australian population. The next chapter considers the stochastic processes used to model this disease spread, while also describing the methods used to quantify the threat of an outbreak. This is followed, in Chapter 3, by a formulation of an SIR household model and how this can be analysed as a branching process. This modelling is then extended in Chapter 4 to improve its applicability to a dengue outbreak in Australia. This includes introducing a vector population, considering the effects of a variety of household sizes and within household mixing. Chapter 5 proposes a novel model that explicitly accounts for how hosts move between households and, finally, Chapter 6 compares the models and presents a discussion of the results including the limitations of the model and possibilities for further research.

Chapter 2

Technical Background

This chapter begins with a brief summary of some common continuous-time Markov chain results that will be used in this thesis, as these are found in most textbooks [13, 24] no proofs or thorough explanations are included. This is done in order to define a basic SIR disease model (which assumes human-to-human transmission). It is then shown how these mathematical methods can be used to calculate some epidemiological quantities.

2.1 Continuous-Time Markov Chains

Firstly, we define the state space of the stochastic process $\{X(t)\}_{t \geq 0}$, \mathbb{S} , by the possible values of which $X(t)$ can take. Using the state space of the process, we say $\{X(t)\}_{t \geq 0}$ is a continuous-time Markov chain if for any $0 \leq s_0 < \dots < s_n < s$ and possible states i_0, \dots, i_n, i, j we have [13]

$$\mathbb{P}(X(t+s) = j | X(s) = i, X(s_n) = i_n, \dots, X(s_0) = i_0) = \mathbb{P}(X(t) = j | X(0) = i).$$

That is, provided we know the current state, the history of the process is irrelevant for predicting the future state of the process [13], making the process “memoryless”. This is formally known as the *Markov property*.

In discrete time, Markov chains can be described by their transition probabilities $p_{i,j}$ (the probability of jumping from i to j in one step). However, in continuous time there is no first time $t > 0$, hence the *transition function* $p_{i,j}(t)$ is defined as [13],

$$p_{i,j}(t) = \mathbb{P}(X(t) = j | X(0) = i).$$

When considering all values of $i, j \in \mathbb{S}$ we can arrange these values into the transition matrix $P(t)$ such that $P(t) = [p_{i,j}(t)]$. The rate at which the chain transitions from state i to j is determined by [13],

$$q_{i,j} = \lim_{h \rightarrow 0^+} \frac{p_{i,j}(h)}{h} \text{ for } j \neq i.$$

With the process' propensity to remain in its current state governed by,

$$q_{i,i} = \lim_{h \rightarrow 0^+} \frac{p_{i,i}(t) - 1}{h}.$$

These rates are ordered into the infinitesimal generator matrix, Q , which is defined as $Q = [q_{j,k}]$, where $q_{j,k}$ is the rate at which the process transitions from state j to state k in a single jump. Q will have the following properties:

- (i) $0 \leq -q_{i,i} < \infty$ for all i ,
- (ii) $q_{i,j} \geq 0 < \infty$ for all $i \neq j$,
- (iii) $\sum_{j \in S} q_{i,j} = 0$ for all i ,

provided $i, j \in \mathbb{S}$. The holding time of the process in state i ($\forall i \in \mathbb{S}$) will be determined by an exponential distribution with parameter $\sum_{j \neq i} q_{i,j}$, provided $j \in \mathbb{S}$.

It has been shown that transition function must satisfy [13],

$$P'(t) = P(t)Q,$$

which is known as Kolmogorov's forward equation. Hence, $P(t) = e^{Qt}$, and is solved by evaluating [13],

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}. \quad (2.1)$$

This form (e^{Qt}) is referred to as the *matrix exponential* and is calculated numerically by the `expokit` package [34].

2.2 Path Integrals of Markov Chains

For each of the Markov chains considered, the state space \mathbb{S} can be split into two classes, absorbing and transient [31]. Written technically as, $\mathbb{S} = A \cup C$, where A and C are the sets of all absorbing and transient states of \mathbb{S} respectively. State i is called *recurrent* if the chain re-visits state i almost surely, and called *transient*

otherwise [13]. Moreover, a state is *absorbing* if the chain will remain in the state almost surely [13]. Provided A is non-empty, $X(t)$ will enter A almost surely in finite time [24]. The function $f(\cdot) : \mathbb{S} \rightarrow [0, \infty)$ is defined such that $f(i)$ denotes the per-unit reward of being in state i . Strictly speaking, there exists a $\tau \in \mathbb{R}^+$ such that for $t \geq \tau$, $\mathbb{P}(X(t) \in A) = 1$. Given $X(0) \in C$, the total reward of the process Γ , can be determined by solving the path integral [31],

$$\Gamma = \int_0^\tau f(X(t))dt, \quad (2.2)$$

where $f : C \rightarrow [0, \infty)$ and $\tau = \inf\{t > 0 : X(t) \notin C\}$ is the first exit time of C (i.e. time at which the system enters an absorbing state). Define $e_i = \mathbb{E}[\Gamma | X(0) = i]$ to be the expected reward from starting the process in state i . Then, e_i can be calculated by solving the system of linear equations (by Proposition 2 [31]),

$$\sum_{j \in C} q_{i,j} e_j + f(i) = 0, i \in C. \quad (2.3)$$

Adding to this, if we take the Laplace-Steiltjes transform of distribution of Γ , $y_i(s) = \mathbb{E}[e^{s\Gamma} | X(0) = i]$, $y_i(s)$ will then be the solution to (Proposition 1 [31])

$$\sum_{j \in S} q_{i,j} y_j(s) = s f_i y_i(s), i \in A. \quad (2.4)$$

These two equations allow us to efficiently evaluate a number of epidemic quantities from a household model.

2.3 SIR Model

The early stages of a disease outbreak are known to be very unpredictable. This is due to the randomness associated with the first few infected individuals either successfully or unsuccessfully transmitting the disease to others in the population. This can result in there being a significant probability that the disease will become extinct relatively quickly, however, if transmission is successful it is likely that an epidemic will ensue. To model these outcomes, we first consider a compartmental *susceptible-infected-recovered* (SIR) continuous-time Markov chain model. A key strength of such a stochastic model (opposed to a deterministic model), is that the probability of a successful outbreak can be estimated, leading to additional

information that is vital to disease prevention strategies.

Event	Transition	Rate
Infection	$(S, I) \rightarrow (S - 1, I + 1)$	$\beta SI/N$
Recovery	$(S, I) \rightarrow (S, I - 1)$	γI

Table 2.1: Transitions and rates of an SIR model with parameters $\beta, \gamma \in \mathbb{R}_+$ and N representing the population size. The spread of the disease is assumed proportional to the number of infected I and the proportion of susceptible S/N , as is common throughout the literature. With recovery (or mortality) being directly proportional to the number of infected.

A compartmental continuous-time Markov chain $\{X(t)\}_{t \geq 0}$ is used to construct a SIR model. Individuals in this model transition from susceptible to infected classes at a rate proportional to the number of infected individuals in the population multiplied by the proportion of susceptible individuals, as infection requires an interaction between members of each of these classes. As a recovery (or mortality) event only requires an infected individual, the transition from infected to recovered occurs at a rate proportional to the number of infected individuals. We use β, γ as parameters for each transition, such that we can tune and adapt the model to specific disease. These transitions and rates are summarised in Table 2.1. The epidemiological interpretation of β is the rate at which an infected individual infects a susceptible individual, this parameter is defined as the product of the probability of transmission, p , and the number of contacts per day per individual, c . $1/\gamma$ can be simply interpreted as the expected duration of an individual's infection (as holding times in each class are exponentially distributed). Once an individual enters a recovered state they will remain there indefinitely. It should also be noted that this model is a closed system with a fixed population of N . We denote the number of individuals in a susceptible, infected or recovered state with S , I and R respectively. The SIR model will have a state space, $\mathbb{S} := \{(S, I) \in \mathbb{Z}_+^2 : S + I \leq N\}$. As each person must exist in one of these classes, the number of recovered individuals can be evaluated by the number of susceptible and infected people, $R = N - S - I$.

To assist in the visualisation of this model, we map the state space \mathbb{S} onto a two dimensional plane (as $\mathbb{S} \subset \mathbb{Z}_+^2$), with the number of susceptible and infected individuals on the x and y axes respectively. Figure 2.1 displays the transitions between states for a population size of four, with arrows representing the direction of each transition. Here, all states with zero infected individuals are absorbing, as there are no transitions leaving these states.

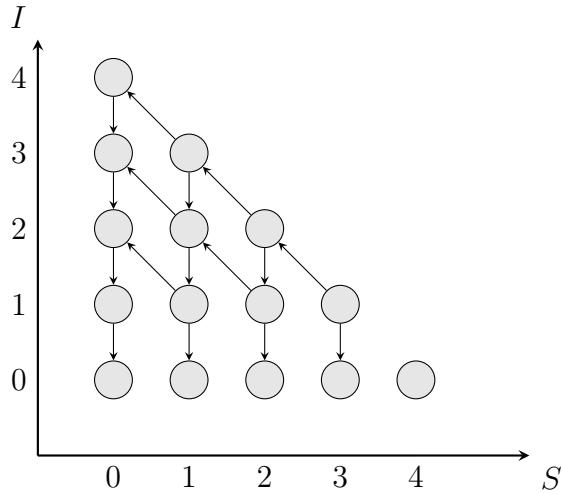


Figure 2.1: Transitions of an SIR model with a population of four individuals with the number of susceptible and infected individuals on the x and y axes respectively. Arrows denote the direction of transition.

2.3.1 Generating the Q Matrix

The Q matrix is generated by utilising MATLAB's `meshgrid` function, as it creates an ordered list of coordinates in two dimensions such that it encompasses the state space ($\mathbb{S} \subset \mathbb{Z}_+^2$). The coordinates (S, I) where the population size was exceeded were then removed, leaving a list of all states of \mathbb{S} . This sample space was ordered such that

$$\mathbb{S} = \{(0, 0), (1, 0), \dots, (N, 0), (0, 1), \dots, (N - 1, 1), \dots, (0, N)\}.$$

To generate Q , the script cycles through each state of \mathbb{S} and checks which events are possible (outlined in Table 2.1). The script will then search \mathbb{S} again to find the position of this corresponding state. After finding this transition, the rate of each event is entered into the corresponding position in the Q matrix. For example, for state $i = (3, 2)$ both events are possible. Hence, for $j = (2, 3)$ and $k = (3, 1)$, $q_{i,j} = \beta SI/N$ (i.e., $6\beta/5$) and $q_{i,k} = \gamma I$ (i.e., 2γ). After all the rates are input, the propensity to remain in each state ($-q_{i,i}$) is then determined by $q_{i,i} = -\sum_{j \neq i} q_{i,j}$ for $i, j \in \mathbb{S}$. The MATLAB function `sparse` was used to markedly reduce the storage required for this matrix, as it will only store non-zero elements of Q .

2.4 Basic Reproduction Number, R_0

The basic reproductive number, R_0 , has been described as being one of the most fundamental metrics for the study of disease dynamics [11]. R_0 is defined as the expected number of secondary cases per primary case in a completely susceptible population [12].

As disease outbreaks tend to either die out relatively quickly (only infecting a small proportion of the population) or a large proportion of the population contract the disease (an epidemic), these phenomena will be referred to as ‘minor’ and ‘major’ outbreaks respectively. In fact, it has been shown that when $R_0 < 1$ the disease will become more or less immediately extinct with a probability of 1 [12]. In the alternate case (where $R > 1$), there exists a strictly positive probability that the infection population tends to infinity (given an infinite population). In fact [12],

$$R_0 > 1 \iff P(\text{major outbreak}) > 0.$$

As mentioned in the previous section, the state space of a SIR model can be split into absorbing and transient states. By using the method outlined in Section 2.2, and redefining the function $f(i)$ to be the number of infected individuals present in state i . The total number of infections (over the course of the outbreak) can be evaluated, as for any finite N , the set of absorbing states A will be non-empty.

R_0 can be calculated by considering each individual having its own CTMC [32], such that $X(t) = 1$ if the individual is infectious and $X(t) = 0$ if it has recovered. Hence, its Q matrix will comprise of $q_{0,1} = \gamma$ and $q_{1,0} = -\gamma$. The rate of infection to another individual is then, $f(X(t)) = \beta X(t)$. By evaluating Eq. (2.3), we determine that $-\gamma e_1 = -\beta$. Hence, the expected number of individuals infected from this individual will be $R_0 = \beta/\gamma$. Technically, this method is evaluating the household reproduction number for households of size one, although this is applicable this case [32], and will be explored more in depth in Chapter 3.

Chapter 3

Household Model

The model in the previous chapter assumed that the population of humans mixes homogeneously. Meaning that each individual in this model has an equal likelihood of coming in contact with any other individual in the population. This is a considerable limitation due to the amount of time family members spend together, previous studies of diseases such as influenza have used this important factor to construct household transmission models [6, 18, 19, 32]. We suspect that additional information can be gathered by creating such a heterogeneous model, where there will exist two different levels of transmission. That is, an individual will have a heightened rate of transmission to their household members than those from other households. By extending the SIR model, we implement this heterogeneity by grouping all individuals into distinct households of size n . Within each household, the transitions will be consistent with the previously mentioned SIR model with a population size of n (although with different parameter values).

Hence the possible states a household can take will be

$$\mathbb{S} = \{(s, i) \in \mathbb{Z}_+^2 : s + i \leq n\}, \quad (3.1)$$

where s, i , represent the number of susceptible and infected individuals in a household. The number of households currently with the same configuration of susceptible/infected household members is denoted h_j , where j is a state in \mathbb{S} . The state of the entire population is denoted \mathbf{h} , which has entries h_j . The state space of \mathbf{h} will be $\hat{\mathbb{S}} := \{\mathbf{h} \in \mathbb{Z}_+^{|\mathbb{S}|} : \sum_{j \in \mathbb{S}} h_j \leq N_H\}$, where N_H is the total number of households in the population. From this CTMC, the population wide numbers of susceptible, infected and recovered individuals can be calculated by, $S := \sum_{j \in \mathbb{S}} s_j h_j$, $I := \sum_{j \in \mathbb{S}} i_j h_j$, $R := N - S - I$ respectively (where $N = nN_H$ is the number of individuals in the population).

Individuals in this model can infect members from their household or other households. The rate of between household transmission will be influenced by the number of between household contacts, i.e., the level of mixing of the population. To parametrise this we introduce α , which is defined as the proportion of time an individual spends in their household on average. It is assumed that each individual must be in a household at all times, hence $1 - \alpha$ denotes the proportion of time spent in other households, as such $0 \leq \alpha \leq 1$. When $\alpha = 1$, no between household mixing will occur, and $\alpha = 0$ will result in a homogeneous model. This is as individuals will not have a bias towards staying in any particular household, creating a well mixed population. A within-household infection occurs at rate $\beta \alpha s_i/n$ (per household) with between-household infections occurring at rate, $\beta(1 - \alpha)S_i/N$. Where i is the number of infected individuals in the considered household, S is the number of susceptible individuals in the population of size N , with infection parameter of β .

This model is an improvement on the originally proposed SIR one, although, there still exists an abundance of assumptions made. Currently, individuals exposed to the disease are instantly capable of transmitting the virus, as well as humans being able to directly infect other humans (which is not consistent with dengue transmission). Looking at the movement of individuals, it is still assumed that people come into contact with those outside their household homogeneously. Furthermore, it is assumed that the each household is of the same size and that the population size is fixed. Meaning that there are no births or deaths in the population. The final assumption is that every individual experiences the duration of each state with an identical exponential distribution.

3.1 Branching Process Approximation

Given a small infected population, the early stages of a continuous-time Markov chain model of this form can be approximated using a branching process. To do this, we make the assumption that each household in the process can only be externally infected once, which allows for each household to function independently (once infected). From which, a branching process can be defined, with the offspring distribution determining the number of households infected by a single household to the number. However, this approximation relies on the assumption of a very large population, where the proportion of infected households will not be significant. This is as the likelihood of a household being re-infected (externally) is very low, resulting in approximately independent infected households. Hence, results from this are only

valid when modelling the early stages of a potential outbreak with an insignificant initial infected population (in comparison to the susceptible population). This is as the branching process will generally either grow exponentially indefinitely or become extinct almost instantaneously, rather than the infectious population curtailing as the recovered population grows (as seen in the finite population CTMC). The difference being that the infection growth of the CTMC is reliant upon the number of both susceptible and infected individuals, whereas the branching process does not account for the change in the number of susceptible people in the population.

In this branching process approximation each household will still retain the same internal dynamics. We denote the SIR process of a single household as $\{\Psi(t)\}_{t \geq 0}$ which has the state space \mathbb{S} as defined in Eq. (3.1). However, now it is not simple to determine the value of R_0 , as there are two levels of mixing present. Although, we can use the branching process to quantify the expected number of households infected directly from a single infected household in a completely susceptible population, denoted R_* . This is known as the household reproduction number, and is of high interest, as it is the threshold parameter of the household model, and shares the same threshold of 1, as the previously mentioned R_0 [6].

We calculate R_* by determining the expected number of external infections produced from a single household CTMC $\Psi(t)$. Let $f(\Psi(t)) = \beta(1 - \alpha)i$ be the rate of external infection for branching process approximation (where i is the number of infected individuals in the household). This change in rate from the CTMC comes from the assumption of a large and almost completely susceptible population such that $S/N \approx 1$. By substituting into Eq. (2.2), the household reproduction number of the branching process approximation model is defined as

$$R_* = \mathbb{E} \left[\int_0^\infty f(\Psi(t)_1) dt \right], \quad (3.2)$$

where $\Psi(0)_1$ denotes a single infected individual initially in an otherwise susceptible household. We solve R_* in the same manner as R_0 , by evaluating the set of linear equations in Eq. (2.3). This is possible because $\Psi(t)$ is an independent homogeneous SIR CTMC with n individuals. This could potentially be calculated using the population wide CTMC, $\mathbf{h}(t)$, however, this is substantially more complicated as an external infection will not necessarily occur at a completely susceptible household.

The quantity R_* for a household model provides us a basic understanding of how the outbreak will play out, however, it does not inform the rate of growth. This is as a R_* value accompanied with a short period of individual infection will grow rapidly

compared to the same R_* with an extended infectious period, i.e. the shorter the time between infectious ‘generations’ the more rapid the growth.

As previously mentioned, the number of infected individuals in the early stages of an outbreak tend to grow exponentially e^{rt} . Here, r is commonly referred to as the early growth rate or Malthusian parameter (of a branching process approximation) [3]. Given that the branching process is supercritical ($R_* > 1$), r is the unique solution in $(0, \infty)$ of the equation [3],

$$\mathbb{E} \left[\int_0^\infty f(\Psi(t)_1) e^{-rt} dt \right] = 1, \quad (3.3)$$

(recall the external infection function $f(\Psi(t)_1) = \beta(1 - \alpha)i$. Which is again an expectation of a path integral. The exponential discounting term (e^{-rt}) can be thought as decreasing the size of $f(\Psi(t))$ (as $r > 0$) until it matches the infection growth curve. A numerical root finding tool was used to solve Eq. (3.3). The early growth rate can be used to determine the doubling time of the number of cases as it is more easily interpreted, this is calculated by $T_d = \ln(2)/r$ [6].

Chapter 4

Household Model with Vectors

In this chapter we extend the household model to include a vector population. The disease is now transmitted from host to vector (female mosquito) rather than host-to-host, as is seen with dengue transmission. We assume that these vectors can exist in three states: susceptible, exposed and infected. The exposed state is included to simulate the extrinsic incubation period mosquitoes infected with dengue experience. As mosquitoes cannot recover from dengue once infected, there is no need to implement a recovered state for the vector population [22]. However, due to a mosquito's lifespan being approximately 21 days, death will often occur [2]. To implement this dynamic, we assume that the death of a mosquito corresponds with a (susceptible) birth within the same household, resulting in the population size remaining constant. Moreover, as mosquitoes generally do not travel long distances it is assumed that mosquitoes cannot infect members of other households. Finally, as only female mosquitoes can function as vectors, males will not be considered.

Similarly to mosquitoes, humans also experience incubation periods from dengue, this leads to a SEIR(host)-SEI(vector) compartmental model. Numerous SEIR-SEI models have been used previously to capture the transmission of a single DENV serotype [2, 14, 20, 21, 23, 27, 30], however not with an explicit household structure. This previous work indicates that this classification will provide an accurate representation of homogeneous dynamics. By coupling these compartments with a heterogeneous household models (which have been previously used for human-human transmitted diseases) [6, 32], a more accurate model could be created.

4.1 SEIR-SEI Model

In order to construct a SEIR-SEI model, consider a population of N_H households. As we are primarily concerned with measuring the risk of an epidemic by utilising the quantities R_* and the early growth rate r , we can extrapolate the population wide dynamics from the process of a single household using a branching process approximation. A single SEIR-SEI household CTMC is represented by $\{\Psi(t)\}_{t \geq 0}$, but will now include a vector population. Let s_j, e_j, i_j denote the number of susceptible, exposed and infected hosts in household j , for $j = \{0, 1, \dots, N_H\}$. Similarly x_j, y_j will represent the number of susceptible and exposed vectors in household j . Hence, we define the state space of a household to be,

$$\mathbb{S} := \{(s, e, i, x, y) \in \mathbb{Z}_+^5 : s + e + i \leq n, x + y \leq m\}. \quad (4.1)$$

The fixed household sizes for both species allows us to determine the number of recovered hosts in each household by $r_j = n - s_j - e_j - i_j$, and the number of infected mosquitoes by $z_j = m_j - x_j - y_j$.

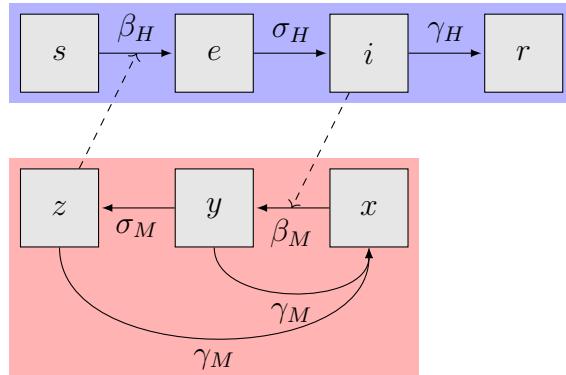


Figure 4.1: Diagram of a single SEIR-SEI household. Hosts and vectors transition from susceptible to exposed to infected states. Hosts then reach an infected state, whereas vectors may transition from an exposed or infected state to susceptible. This is to simulate the death of infected mosquitoes and birth of susceptible ones. Blue highlights the host classes and red for the vector classes. With dashed lines representing the influence a class size has on a transition rate as transmission can only occur from host-vector interactions.

Due to the introduction of the vector population, we remove the assumption that hosts can transmit the disease among each other, such that only host-vector transmissions exist. As the rate of transmission from host to a vector is not equal to

Parameter	Biological Meaning	Value	Source
a	Average daily biting rate	1	[2]
p_H	Transmission probability from human to mosquito	0.75	[2]
p_M	Transmission probability from mosquito to human	0.5	[2]
α	Proportion of time spent in primary household	0.7	-
σ_H^{-1}	Intrinsic incubation period	7	[9]
σ_M^{-1}	Extrinsic incubation period	10	[2]
γ_H^{-1}	Infectious period of a human	2/9	[35]
γ_M^{-1}	Average lifespan of mosquitoes	21	[2]
β	General infection rate	4/9	-

Table 4.1: Summary of parameters for the SEIR-SEI household model. These are found from a similar SEIR-SEI model of the 2014 Guangzhou dengue outbreak [21]. It is assumed that $\beta = 2\gamma_H$ such that there are significant probabilities of both major and minor outbreak events. The proportion of time spent in the primary household α will be tested to see the impact it has, hence there is no source for it.

that from vector to host, we introduce β_H , β_M to be the parameters associated with the rate of infection for each population. These parameters for host and mosquito infections are defined to as $\beta_H = \beta a p_v \alpha$ and $\beta_M = \beta a p_H \alpha$ respectively, where β is the general infection rate, a is the daily biting rate, p_M and p_H are the transmission probabilities from vector to host and vice versa. Lastly, α is the proportion of time hosts spend in their home, the values of these parameters are presented in Table 4.1.

Host infection occurs with a propensity equal to the number of infected mosquitoes in a household multiplied by β_H and the probability of selecting a susceptible host within the same household, s/n . Similarly, vector infection occurs with a rate equal to the product of β_M , the number of susceptible mosquitoes in the household x , and the proportion of infected hosts in the household i/n . Both exposed hosts and mosquitoes will transition to infected states with parameter σ_H , σ_M respectively. Infected individual hosts recover at a rate governed by γ_H . However, as mosquitoes cannot recover, and we assume that a death corresponds with a birth, exposed and infected mosquitoes transition to susceptible (to approximate this birth-death event) at rate γ_M . A detailed description of these transitions is displayed in Table 4.2. A single household's dynamics illustrated in Figure 4.1 with arrows representing transition of hosts (in blue) and vectors (in red), dashed lines are used to show the effect other classes have on transition rates.

Due to the vector population there now exists two different types of external (between household) infections, meaning an external transmission results in either a host or vector becoming exposed to the disease. The first case arises when a host

Event	Transition	Rate
Host Infection	$(s, e) \rightarrow (s - 1, e + 1)$	β_{HSz}/n
Host Latent Progression	$(e, i) \rightarrow (e - 1, i + 1)$	σ_{He}
Host Recovery	$(i) \rightarrow (i - 1)$	γ_{Hi}
Vector Infection	$(x, y) \rightarrow (x - 1, y + 1)$	β_{Mxi}/n
Vector Latent Progression	$(y, z) \rightarrow (y - 1, z + 1)$	σ_{My}
Vector Death (E)	$(x, y) \rightarrow (x + 1, y - 1)$	γ_{My}
Vector Death (I)	$(x, z) \rightarrow (x + 1, z - 1)$	γ_{Mz}

Table 4.2: Transitions and rates within a SEIR-SEI household model $\Psi(t)$. Both infection events have rates including the proportion of susceptible/infected hosts found in the household to represent the probability that a mosquito comes in contact with either class of host. It should be noted that the other unaffected classes are omitted.

from a completely susceptible household becomes exposed to the disease while visiting an infected household. The second circumstance occurs when an infected host visits a completely susceptible household and transmits the disease to a mosquito, causing a single mosquito infection. The former and latter phenomena will be referred to as *Type 1* and *Type 2* infections respectively.

The level of between household mixing is approximated by using the proportion of time an individual spends inside their primary house, α . As we assume a host is always in a household, the proportion of time spent visiting a household is $1 - \alpha$. To reiterate from Chapter 3, when $\alpha = 1$ no between household mixing will occur, and $\alpha = 0$ will create a homogeneous model. Therefore, the rate at which a single infected household ($\Psi(t)$) creates a new Type 1 infected household is,

$$f_1(\Psi(t)) = \beta(1 - \alpha)p_M z, \quad (4.2)$$

when considering a single household in a completely susceptible population (such as in the branching process approximation). That is, the rate that a single visitor is infected when visiting an infected household. A single infected household creates a Type 2 household at the rate

$$f_2(\Psi(t)) = \beta(1 - \alpha)p_H m \frac{i}{n}, \quad (4.3)$$

where i/n is the proportion of infected hosts in the infected household, and m is the number of susceptible mosquitoes in the susceptible household. Intuitively, this proportion appears as if it should be $i/(n + 1)$ (as there is a visitor in the household), however, on average, the increase of hosts in a household (visitors) will

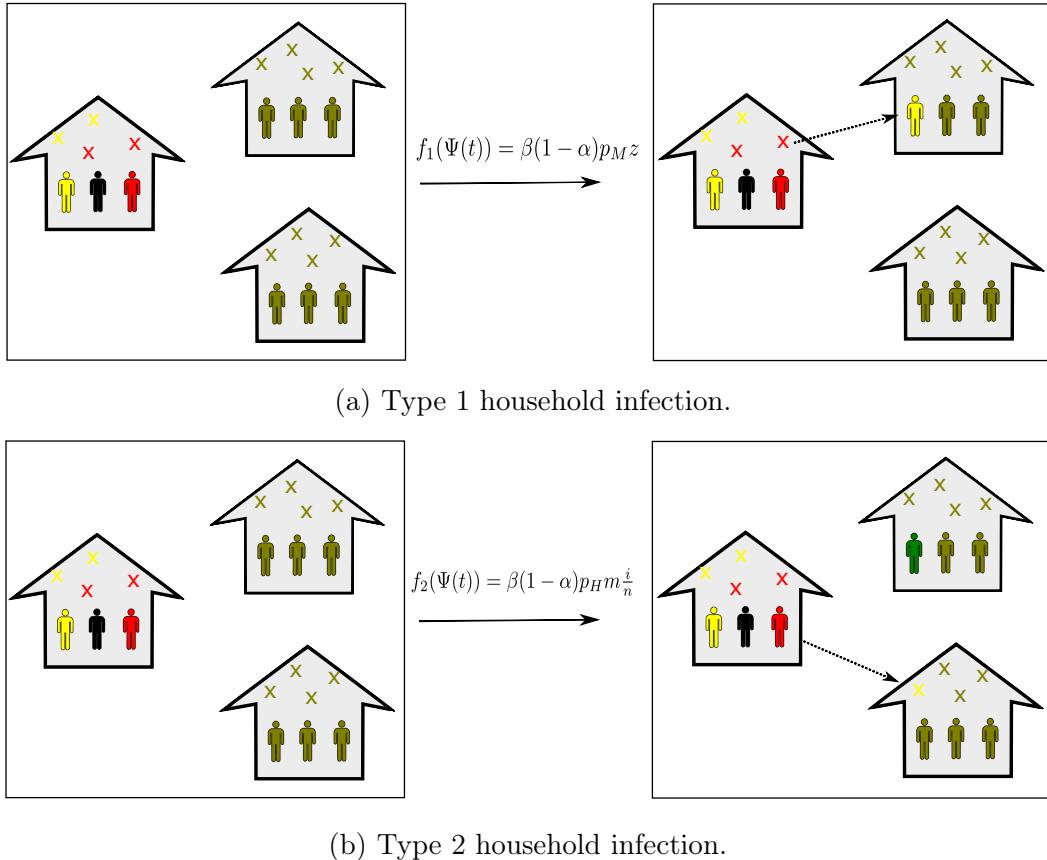


Figure 4.2: Between household infections, in a population of 3 households, with 3 hosts and 4 vectors in each. (a) A Type 1 infection occurs when a susceptible host visits and infected household and contracts the disease from a mosquito, which occurs at rate f_1 . (b) A Type 2 infection occurs when an infected host infects a mosquito while visiting a susceptible household, which occurs at rate f_2 . These examples comprise of 3 households, with green, yellow, red, and black representing susceptible, exposed, infected and recovered agents respectively.

be counteracted by the leavers of the same household. Thus, keeping the household size of n constant.

These events were made in order to approximate host movement between households, without the need to record the position of each host. A limitation of this heuristic is that it assumes that a household cannot be infected from another household more than once.

4.2 Next-Generation Matrices

Two different types of between household infections were introduced, this further complicates the meaning of R_* as we have to calculate the expected number of Type 1 and 2 infections resulting from a Type 1 or 2 infected household. Hence, a new method is now required to aggregate these results to determine R_* (as well as the early growth rate).

A matrix K is created such that the element in the i^{th} row and j^{th} column will correspond with the expected number of Type i external infections caused by a single Type j household in a completely susceptible population,

$$k_{i,j} = \mathbb{E} \left[\int_0^\infty f_i(\Psi(t)_j) dt \right], i, j \in \{1, 2\}, \quad (4.4)$$

where $\Psi(t)_j$ is the initial state of a single household CTMC which becomes infected by a Type j transmission. This matrix K is referred to as the next-generation matrix and is paired with a vector ϕ , where each element of ϕ is defined as the number of different types of external infection j .

Considering the branching process approximation framework, the m^{th} generation of ϕ is calculated by,

$$\phi_i^m = \sum_{i=1}^n k_{i,j} \phi_j^{m-1}, \quad (4.5)$$

or rather in matrix form $\phi^m = K\phi^{m-1}$ [12]. Although we refer to this as the m^{th} generation, implying a discrete-time branching process, as we are concerned with the total number of infections the time at which these occur does not matter. Furthermore, the m^{th} generation of ϕ can be recursively defined by

$$\phi^m = K\phi^{m-1} = K^2\phi^{m-2} = \dots = K^m\phi^0,$$

here ϕ^0 is the initial infection to the branching process's population. Hence, allowing us to calculate the number of Type 1 and 2 infections when using this approximation. To clarify the meaning of ϕ , under the current formulation of the model (where we consider two possible initial states),

$$\phi^0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

with Type 1, 2 initial infections being in the respective element of ϕ . That is, a household can either be of Type 1 or Type 2. Moreover, the total number of

household infections by ‘generation m ’ is defined as $\|\phi^m\|$.

Returning to the next-generation matrix, R_* can be calculated by [12],

$$R_* = \lim_{m \rightarrow \infty} \|K^m\|^{1/m}.$$

It has been shown that matrix K is positive definite, hence the overall reproduction number R_* will be its dominant eigenvalue (Theorem 7.3) [12]. Dominant meaning that R_* is greater than or equal to the absolute value of every other eigenvalue of K . This value is the total number of households infected by a single infected individual, regardless of type. The proportion of Type 1/Type 2 infections are found by normalising the corresponding eigenvector [12].

Using the same next-generation method, we define the early growth matrix $G(r)$ such that,

$$g_{i,j} = \mathbb{E} \left[\int_0^\infty f_i(\Psi(t)_j) e^{-rt} dt \right], \quad i, j \in \{1, 2\}, \quad (4.6)$$

in order to calculate the early growth rate r . The exponential discounting term (e^{-rt}) can be thought as decreasing the magnitude of $f_i(\Psi(t)_j)$ as $r > 0$ (with $r = 0$ when $R_* \leq 1$). The discounting term essentially increases the rate of recovery by r until the reproductive threshold of 1 is met. Thus, the corresponding early growth rate of each element of $G(r)$ is found by solving for r when the dominant eigenvalue of G is equal to one, $\Lambda(r) = 1$. Finding the solution for r is not possible analytically, so a numerical root finding tool (MATLAB’s function `f-zero`) is used instead.

4.3 Total Infection of a Household

Within household dynamics are also important to understanding the spread of the disease. For example, if we compare two scenarios, the first being where no within household transmission occurs and the second being when the entire household becomes infected. Given the R_* values are the same for these situations, the latter situation will result in n (the number of hosts per household) times more host infections than the former case. Which could result in quite a considerable difference in the expected number of mortalities in a population. To investigate this, we determine the total number of hosts that become infected before the household enters an absorbing state (where no exposed or infected hosts/mosquitoes exist). As every infected host reaches a recovered state, the absorbing states of each household consists of hosts in susceptible or recovered classes with all mosquitoes being in the

susceptible class.

To determine this total infection distribution of $\Psi(t)$ some more theory is required. We introduce a *hitting time* to be the time at which the process will enter a particular state. The hitting time of subset $A \subset \mathbb{S}$ (where A is the set of all absorbing states) is determined by

$$H^A = \inf\{n \geq 0 : \Psi(t) \in A\}$$

provided the infimum of the empty set will be infinity. Then, the probability of hitting A given $\Psi(t)$ is in state i at time t will be [24]

$$h_i^A = \mathbb{P}_i(H^A < \infty).$$

Let \mathbf{h}^A be a vector of hitting probabilities such that $h^A(i) = h_i^A : i \in \mathbb{S}$, then \mathbf{h}^A will be the minimal non-negative solution to the system of linear equations [24]

$$h_i^a = \begin{cases} 1 & i \in a, \\ \sum_{j \in \mathbb{S}} p_{ij} h_j^a & i \notin a, \end{cases} \quad (4.7)$$

where h_i^A is the hitting probability of $\Psi(t)$ in state i hitting an absorbing state a and p_{ij} is the jump chain of the model. Using this, the probability of k infected hosts (over the disease's lifespan) is equivalent to the hitting probability of the process reaching an absorbing state with k recovered hosts. As all of these states will be absorbing, the household infection distribution will then be the solution of the system of linear equations in Eq. (4.7) for each of the states. As there are two types of infection this calculation must be performed for both initial cases.

The resulting distributions from each type of external infection are displayed in Figures 4.3a and 4.3b. It is noticeable that the most likely outcomes are: no further infections or complete infection. This suggests that once an additional host becomes infected there is a high chance that the entire host population will become infected (in the household). We should note that these results are dependent on the choice of parameters values.

To demonstrate the importance of reducing the infection within a household we calculate R_* when only a single host per household can be infected (with 4 hosts and 5 vectors per household). It was found that when only a single household member could become infected R_* was 2.0786 compared with the unrestricted scenario that resulted in a R_* of 3.7637. This scenario was implemented by altering the internal infection rate from mosquitoes to humans such that it was approximately

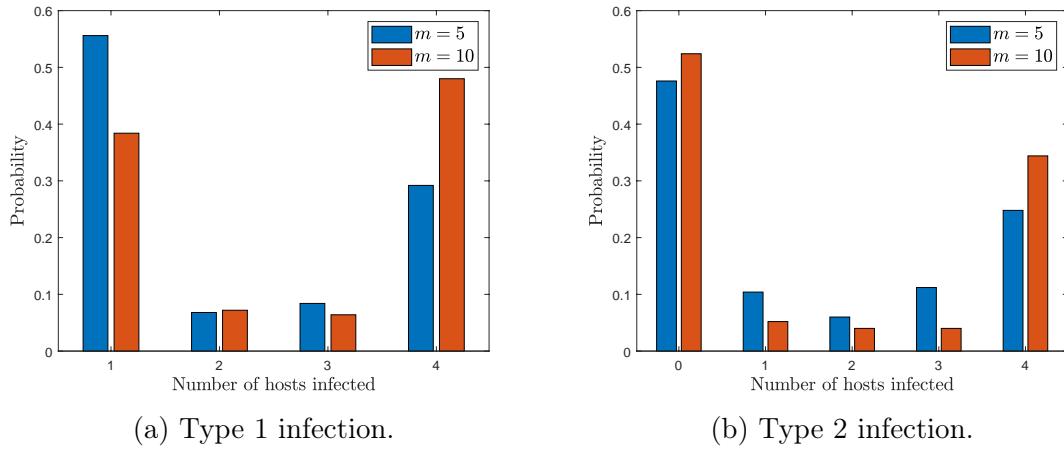


Figure 4.3: Total infection distributions of hosts in households with Type 1 and Type 2 infections from the SEIR-SEI model. Comparing the distributions with vector populations of five and ten (with four hosts in all households). Both types of infection tend to die out before a further infection can occur or result in all the hosts of the household being infected.

0. This allowed for the same mixing of between households (hence, external infection rates), but only a single host per household be infected. Suggesting that when within-household transmission is uncontrolled, not only will the expected number of households infected (per household), but so will the number of infected hosts in the population.

4.4 Size-Biased Distribution, π_k

This model is then extended to incorporate a distribution of household sizes to increase the realism of the model. Using census data, the proportion of households in Australia which are of size k (n_k) was obtained [26]. Meaning that a randomly sampled individual from the population will have a household of size k with probability π_k . This is calculated by [6],

$$\pi_k = \frac{k n_k}{\sum_j j n_j}, \quad (4.8)$$

where n_k is the proportion of the population that has a household size of k members. Therefore, the household size of an external infection is determined by sampling from the distribution, π_k . The difference of these distributions is shown in Figure 4.4, from here we see the bias of an external infection occurring in a larger household.

We now update R_* and r to incorporate the size-biased distribution. A K matrix is calculated for each possible household size (hosts) n . We then find the eigenvalue

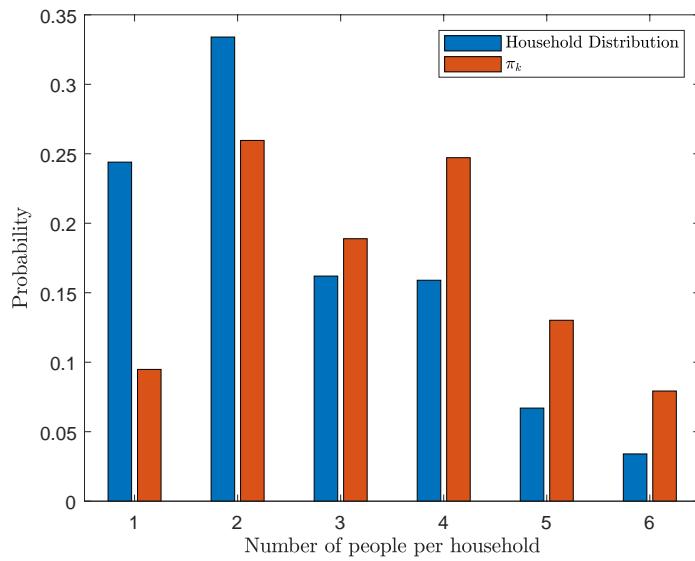


Figure 4.4: Comparison of the household size distribution with the size-biased distribution. Using household size data from the Australian census in 2016 [26], the probability of selecting someone from a larger household is significantly greater than the proportion of these household sizes. Displaying that the size-biased distribution will produce different results, than the proportion of houses.

of each K matrix and denote this denoted R_*^k (the R_* from a household with k hosts). From this, the overall $R_* = \sum_k \pi_k R_*^k$. Similarly, we calculate the average early growth rate by finding r for each host size r_k and multiplying by the size-biased distribution $r = \sum_k \pi_k r_k$.

4.5 Biting Distribution

In the current model, it is assumed that an infected visitor can only be bitten by a single mosquito. This assumption is now removed by defining a discrete distribution for the number of mosquito infections caused by an infectious visitor. To calculate this we consider the number of mosquito infections in a susceptible household with a single infected host (the visitor). As the visits are relatively short compared with the lifespan of a mosquito we do not consider the probability that an infected mosquito recovers during these visits.

Using the transition function $P(t)$ we calculate the probability that a completely susceptible household $\Psi(t)$ with an infected visitor ($n+1$ hosts) will have k exposed mosquitoes over the duration of the visit is calculated. Although, describing the population mixing with a single parameter α causes some limitations. The major

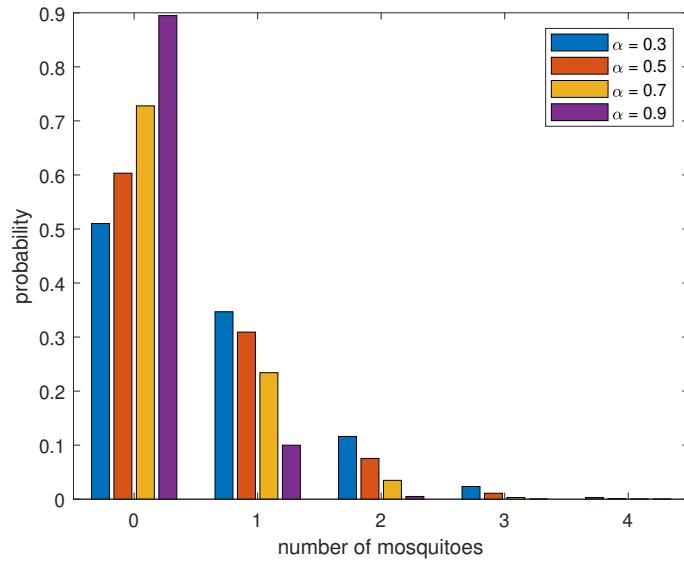


Figure 4.5: Biting distributions for four values of α found from the SEIR-SEI model with a household for n hosts and one infected visitor with five mosquitoes per household. As duration of a visit ($1 - \alpha$ days) reduces the biting distribution is shifted towards the left. Meaning that longer visits will cause more infections on average.

problem being that this parameter controls both the length of a visit and number of visits per day. This impacts calculations as, infecting a single mosquito in two households will result in a R_* twice as large as infecting two mosquitoes in the same household. It is assumed that the duration of a visit will be $1 - \alpha$ days per day, for simplicity. Implying that each host will visit one household on average per day.

Four different levels of mixing were plotted against each other (see Figure 4.5). As the expected duration of a visit increases the number of mosquito infections increase. For example, a visiting duration of 12 hours ($\alpha = 0.5$) approximately results in 0.5 mosquito infections on average. We notice that these probability distributions resemble that of a Poisson distribution. Hence, a Poisson distribution is fitted to the biting distributions in Figure 4.6. We find that the biting distribution can be reasonably approximated by a Poisson PMF with parameter $1 - \alpha$ for $0 \leq a \leq 1$. This is due to the exponential distribution of the time of each infection in the household. The difference between these distributions is caused from the mosquito infection parameter β_M , as when an β_M value results in the vector transmission rate $\beta_{Mxi}/(n + 1) \approx 1$, the biting distribution will be exactly Poisson with parameter $1 - \alpha$. This is noted as the implementation of a Poisson distribution will markedly reduce computational power as well as being much simpler to implement.

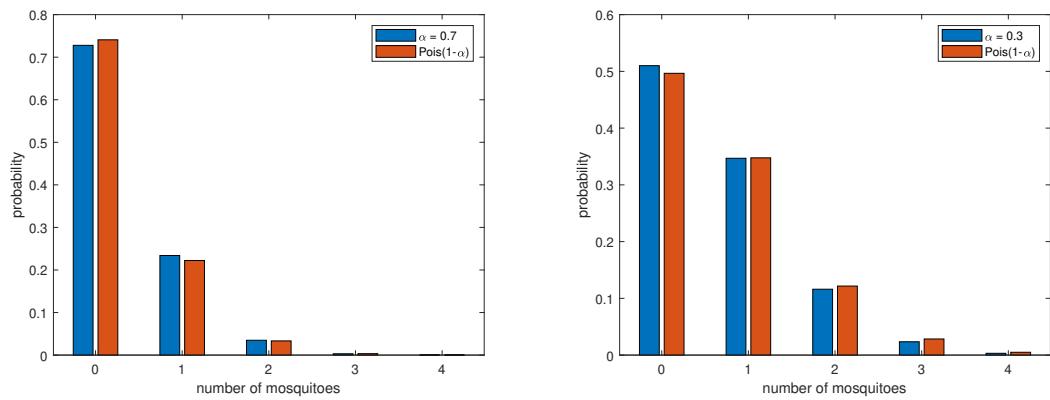


Figure 4.6: Fitting a Poisson distribution to biting distributions. There is a relationship between a biting distribution with parameter α and a Poisson distribution with parameter $1-\alpha$. It should be noted that this relationship holds for all $\alpha \in (0, 1)$ with the other selected parameter values, but only two cases were shown as examples.

4.6 Results

After incorporating the previously mentioned extensions into the SEIR-SEI model, some results were produced, to determine the effectiveness of decreasing the size of mosquito populations. This was achieved by calculating R_* as a function of the number of vectors per household, m (see Figure 4.7a). We find that there is a positive-linear relationship between R_* and the number of vectors per household, and that reducing the mosquito population will result in the population being safe from an epidemic (as $R_* \leq 1$). By altering the mixing between households ($1 - \alpha$) we find that the gradient of the line reduces drastically. For example, if the proportion of time hosts spend at home is increased from 0.7 to 0.9, households can accommodate twice as many vectors before a risk of a major outbreak.

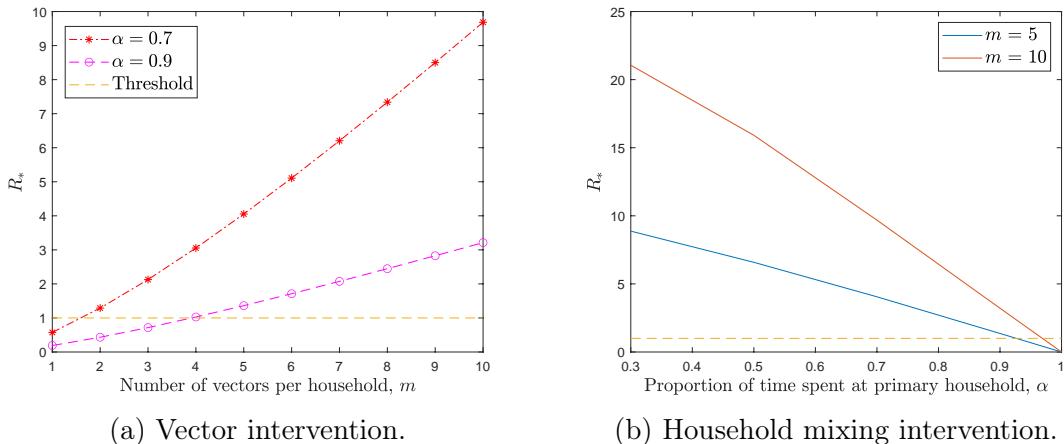


Figure 4.7: Intervention by mosquito reduction, or by reducing the level of mixing of the population. Household reproduction number, R_* , as a function of the numbers of vectors (in each household) m , and the between household mixing α from the SEIR-SEI model. (a) Reproductive thresholds are exceeded for $m = 2$ and 4 for α values of 0.7 and 0.9 respectively. (b) The basic reproduction number threshold exists when hosts spend 93% and 98% of their time at home, for the respective vector sizes of 5 and 10.

To investigate the effect of α further, we plot the values of R_* resulting from values of $\alpha \in [0.3, 1]$ (see Figure 4.7b). As α is increased the level of mixing of the population is reduced, this results in the considerable decrease of R_* . A limit of α occurs when $\alpha = 1$, as hosts in the population cannot interact with vectors from other households. This results in the infection being restricted to the initially infected household, resulting in $R_* = 0$. Given 5 vectors per household, if the proportion of time spent at home is increased to over 93% there will be zero chance of a major outbreak. Whereas, when there are 10 vectors per household this threshold

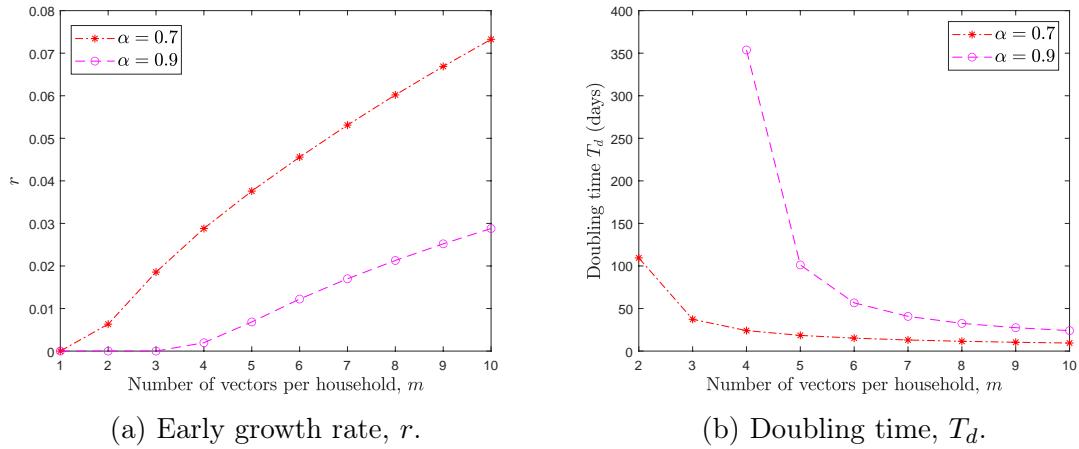


Figure 4.8: Intervention by mosquito population reduction on the rate of growth of infection. Early growth rate, r , as a function of vectors per household m , with two values of α from the SEIR-SEI model. (a) The difference in rate of growth for levels of mixing ($1 - \alpha$) of 0.3 and 0.1. (b) Interpretation of this difference by doubling times, $T_d = \ln(2)/r$. The decrease in mixing ($\alpha = 0.9$) caused the doubling time to increase from 18 days to 101 days (given 5 mosquitoes per household).

is increased to approximately 98%. Meaning that the proportion of time hosts spend outside their main residence must be reduced approximately three times (0.07 to 0.02). This indicates that a combination of both intervention techniques will provide the most effective decrease of R_* .

The rate of growth of the infection in the population is also of concern, as it provides public health officials a time frame in which they must implement control strategies. We compare the early growth rate, r for two values of α (0.7 and 0.9) (see Figure 4.8a). Similarly to Figure 4.7a, we see that the threshold ($r > 0$) is broken at two and four vectors per household for α values of 0.7 and 0.9 respectively. Moreover, the rate of change per vector is significantly reduced when limited movement is implemented ($\alpha = 0.9$). The significance of this difference is displayed in Figure 4.8b, as it allows additional time to intervene with the potential epidemic. Here, we find that for a population of households with 5 vectors each, if the proportion of time hosts spend outside their household is reduced by 3 times (0.3 to 0.1) the doubling time becomes five times greater (101 days rather than approximately 18 days). This significantly reduces the rate of growth of the infection, and should be considered as a short term strategy to provide additional time to implement a more long term intervention. It should be noted that doubling times that correspond to $r = 0$ are undefined, hence could not be included in Figure 4.8b, due to $T_d = \ln(2)/r$.

Using a Gillespie simulation of the SEIR-SEI model, we calculate the *offspring*

distribution which is the probability that k households of Type i are infected by a household of Type j . This is to illustrate the difference in the impact of household transmission types. As we can see from Figure 4.9, regardless of initial conditions,

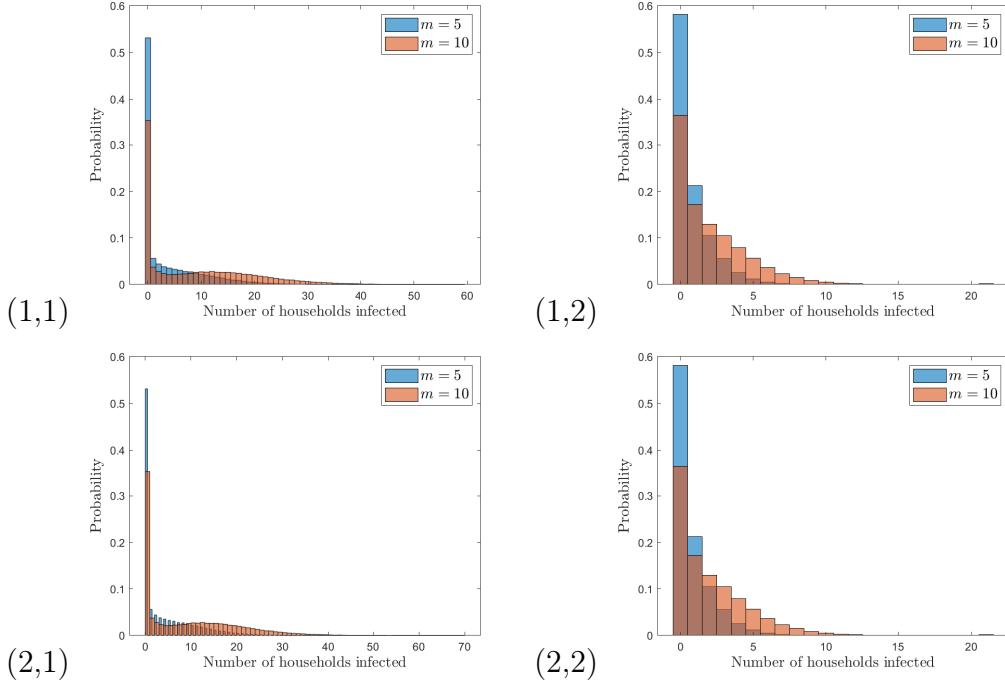


Figure 4.9: Offspring distributions for each combination of types, comparing vector sizes of five and ten. The plot in position (i, j) displays the number of secondary infected households of Type j that arise from an initial household of Type i .

Type 1 secondary infections appear rampant opposed to Type 2. Therefore, this suggests that the majority of between household infections occur when a susceptible visiting host becomes infected. This is further, reinforced by the analytic results of the proportion of Type 1, 2 infections. This was calculated by normalising corresponding eigenvector of R_* . We found that approximately 66% of external infections consisted of a host returning after being exposed visiting an infected household.

Chapter 5

Modelling the Movement of Hosts

The previous SEIR-SEI household model approximated the movement of hosts in a population by the use of transmission functions (f_1, f_2) in conjunction with the biting distribution. To extend this model, we assign rates at which hosts can travel from their primary household to a secondary household and vice versa. This is still a stochastic model, however, a generator matrix cannot be easily defined due to the state space now being considerably larger. This is as the location of each host at every epoch must be known. Moreover, each host will now have an additional $N_H - 1$ movement events that can occur, where N_H is the number of households in the model. The motivation for this model comes from the understanding that it will provide us with more accurate dynamics of disease spread between households, from which the results can be used to assess the strengths and weaknesses of the branching process approximation.

This model removes various other assumptions made previously. Now, household sizes decrease when a household member visits another, there is also the possibility of multiple hosts visiting or leaving any of the households at any point of time. A previously infected household can now be further infected via external transmission making infected households dependent, which breaks any branching process definition. This chapter will identify whether the assumptions made by the branching process accurately portray the between household transmission dynamics.

5.1 Formulation

This model functions similarly to the SEIR-SEI household CTMC, however, due to the inclusion of host ‘movement’ events the state space increases substantially as each host also exists in a compartment denoting the household which they are

in. Similarly to the SEIR compartments, there now is additional N_H compartments to incorporate the host movement. Each host has a rate at which they travel to each secondary household in the population τ and a rate at which they return home μ . Secondary households are defined as any household in the population of which they do not primarily reside. Lastly, hosts can only move between their primary household and a secondary household (which occur instantaneously).

In the previous model it was assumed that the level of mixing could be described by a single parameter α , however, there are two factors that play into the level of mixing experienced. Firstly, the time a host spends outside their primary household on average $(1 - \alpha)$ as well as the number of households a host visits per day. The associated parameter with the rate of host movement in the model is denoted by λ .

The propensity each individual has to return home is calculated by the product of the rate of household movements per day (per individual) and the proportion of time spent in the primary household (α). Using this, the propensity each individual has to return home is,

$$\mu = \lambda\alpha. \quad (5.1)$$

Similarly, the propensity an individual has to travel to a secondary household will be determined by $\lambda(1 - \alpha)$, making the rate of travel to each particular secondary household being,

$$\tau = \frac{\lambda(1 - \alpha)}{N_H - 1}, \quad (5.2)$$

as there are $N_H - 1$ secondary households. This is as the the product of exponential distributions with parameter $\lambda(1 - \alpha)/(N_H - 1)$ will have the parameter of $\lambda(1 - \alpha)$.

To implement this model, each host will now have individualised rates dependent on the sizes of each class in the household at which the host is currently present. We summarise the disease transitions for each host in household j in Table 5.1. The mosquito dynamics for this model will remain the same as in the SEIR-SEI model (see Chapter 4), the only slight alteration being that the number of hosts per household (found in the exposure rate), which is now variable depending on visitors and leavers of that household. Meaning that the rates update more often than previously. Finally, the assumption that mosquitoes cannot move between households remains.

A major draw back of this model is that infected households are not independent. Meaning that, the previous techniques for evaluating R_* and r have now

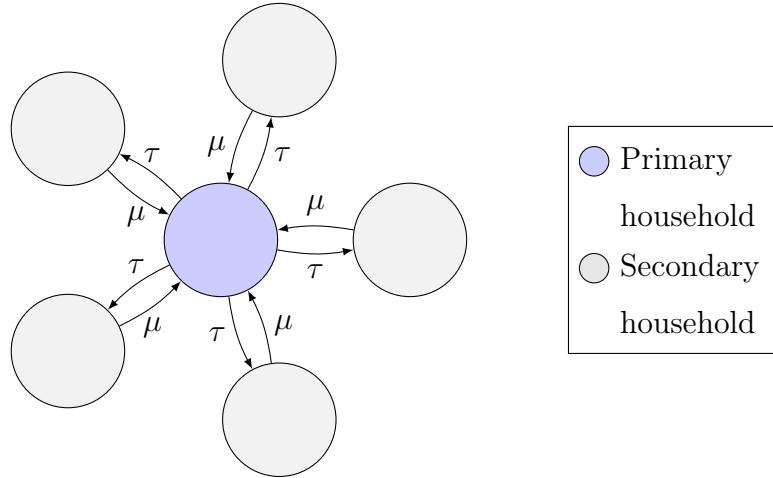


Figure 5.1: Movement transition diagram of a host in the movement model. Here we consider 6 households, the host has the same propensity (τ) to travel to each other household. A visitor must return to their primary household before visiting another secondary household. The duration of each visit will be exponential distributed with mean $1/\mu$.

Event	Transition	Rate
Host Infection	$s \rightarrow e$	$\beta_H z / n_j$
Latent Progression	$e \rightarrow i$	σ_H
Recovery	$i \rightarrow r$	γ_H

Table 5.1: Transitions of an individual in household j , N_j is the number of hosts currently in household j .

become redundant. These quantities will now need to be calculated through simulation exclusively, leading to a major increase computational power. Moreover, this movement model is substantially slower to run as substantially more events occur each day.

5.2 Methods and Results

To test whether the movement simulation is behaving as expected. Firstly, the proportion of time hosts spend in their primary household is extracted, this is then compared to the proportion set by α . The comparison in Figure 5.2 demonstrates that the simulation results lie almost exactly with $\alpha = 0.7$. To validate this result if we consider a population of two households such that state 0 is the primary household and state 1 is the secondary. Here, there will only be two transitions for a host (to and from the primary household), hence the Q matrix will have elements

$q_{0,1} = \tau$ and $q_{1,0} = \mu$ (with the negative counterparts on the diagonal). By taking the Kolmogorov forwards equation, $dP/dt = QP$, it was found that [13]

$$p_{0,0}(t) = \frac{\tau}{\mu + \tau} + \frac{\mu}{\mu + \tau} e^{-(\mu + \tau)t}.$$

That is, the probability of being in state 1 after time t given the process starts in state 1. As t tends to infinity, $\lim_{t \rightarrow \infty} p_{0,0}(t) = \tau/(\mu + \tau)$. After substituting in the values of $\mu = \lambda\alpha$ and $\tau = \lambda(1 - \alpha)$, the probability of a host being in their primary household is found to be α . By the definition of $\tau = \lambda(1 - \alpha)/(N_H - 1)$ this result remains true for any value of $N_H \geq 2$ or $\lambda > 0$.

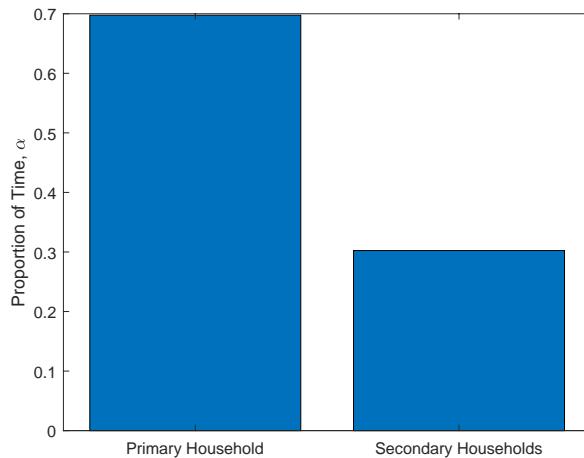


Figure 5.2: Proportion of time hosts of the simulation spend in their primary household compared with secondary households. Hosts in the simulation spend 70% of their time in their household, which is equivalent to $\alpha = 0.7$. This suggests that the implementation and calculation of the transition parameters are accurate.

Using this same classification that a host is in a primary or secondary household, we can determine the value of λ . It is assumed that each host will leave their primary home to visit a secondary household and return home each day (on average). As the duration of time spent in any household is exponentially distributed, the mean duration in the primary and secondary household will be $(\lambda(1 - \alpha))^{-1}$ and $1/(\lambda\alpha)$ respectively. As we assume a host will return to their primary household in one day, λ will be the solution of,

$$\frac{1}{\lambda(1 - \alpha)} + \frac{1}{\lambda\alpha} = 1,$$

hence, $\lambda = (\alpha - \alpha^2)^{-1}$. For example, when we assume $\alpha = 0.7$, $\lambda = 100/21 = 5.01$.

To demonstrate the difference in transition rates, the distribution of a single

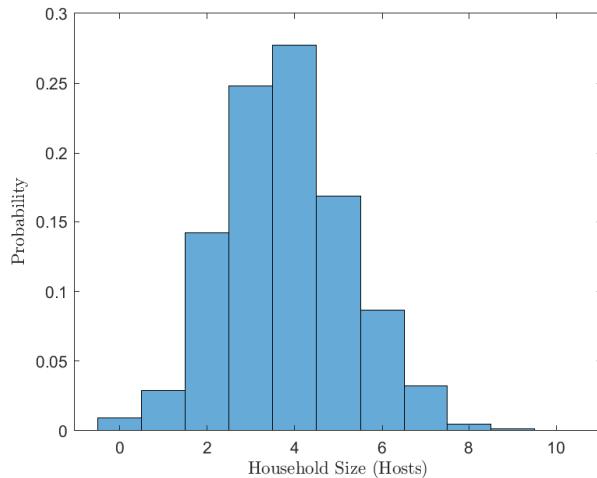


Figure 5.3: The distribution of a single household’s size with $\lambda = 5.01$. As host movement has been implemented, household sizes fluctuate in accordance to λ . The distribution of household size is approximately symmetric with a mean of the initial household size $n = 4$.

household’s size over the course of a simulation is presented in Figure 5.3. This shows the variance of household size experienced, with households now able to be empty or twice their original size, although this occurs rarely. The spread of the distribution is directly affected by the value of λ , in this example $\lambda = 5.01$, corresponding with two movements per host per day.

The early growth rate of the movement model was found by fitting an exponential curve to the number of infections the simulation produced. Figure 5.4 compares the early growth rates of the movement model and the branching process with a $\alpha = 0.7$. The discrepancy between these models is believed to be the result of a small population size (of 1000 hosts) selected when simulating the movement model. This is due to the infectious population quickly becoming a significant proportion of the total population, resulting in the rate of growth to curtail. The population was limited by the computation power available, as each simulation took well over an hour to complete. If this is in-fact the cause of the difference, these results suggest that the branching process may be capturing the important between-household dynamics and that these two models are not too dissimilar.

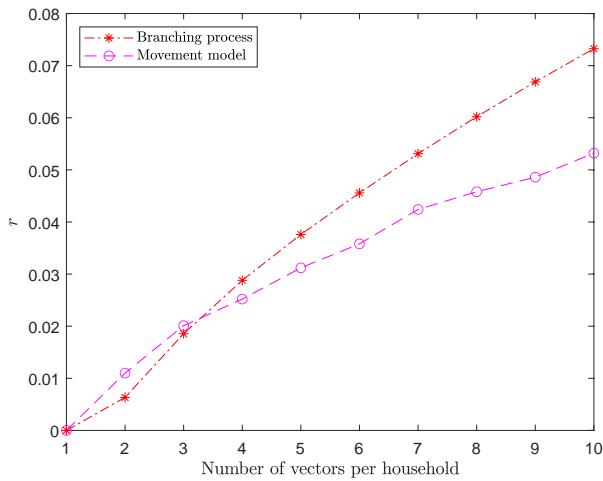


Figure 5.4: Comparison of the movement model with the branching process with $\alpha = 0.7$. The early growth rate, r , is calculated while parametrising for the number of mosquitoes in both the movement model and branching process approximation. These lines suggest similarity between the models, as the disparity is likely caused by the small population size of 1000 hosts.

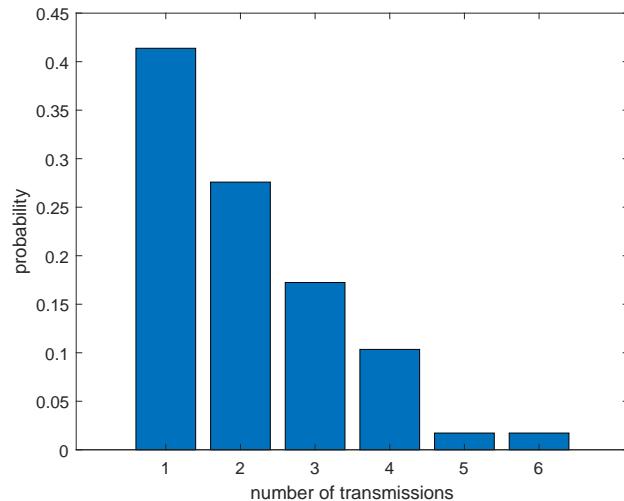


Figure 5.5: Number of transmissions per infected person given they successfully transmit the disease. This is approximately a Poisson distribution, explained by the counting process of the exponential times for each infection.

In Figure 5.5 the probability of the number of transmissions per infected host is shown, given the individually successfully transmits the disease at least once. This approximately takes the form of a Poisson distribution, which is expected as the event in question is a counting process of events that occur at times that are exponentially distributed.

Chapter 6

Discussion

Understanding how infectious disease is transmitted across a population is a key component of effective disease control. In this thesis, continuous-time Markov chains and branching processes were used to construct a household model that approximates the spatial component of how a vector-borne virus such as dengue, is likely to spread in an Australian city. Based on the assumption that each infected household will function independently, a branching process was created. It was shown that by evaluating path integrals of a single household CTMC the value of the household reproduction number R_* and the early growth rate could be calculated. The model was then extended to consider the effects of an incubation period for both host and vector populations, as well as household size and external infection (biting) distributions. Using transmission parameters informed from studies on the recent dengue outbreak in Guangzhou, China [2, 9, 35], and household sizes from Australian census data [26], the resulting model was then used to determine the effect of a single infected host on a large susceptible Australian population. After introducing parameters relating to the level of mixing and the number of mosquitoes present in each household it was determined that when hosts spend an average of 30% of their time in other households ($\alpha = 0.7$), 2 vectors per household would result in a non-zero probability of a major outbreak occurring. However, if the hosts only spent 10% of their time in secondary households, up to twice as many vectors could be accommodated before the same threshold is crossed. By comparing the rate at which the infected population grows (from which the doubling time was calculated), we were able to show that when α increased from 0.7 to 0.9 (for 5 vectors per household) the doubling time increased from approximately 18 days to 101 days. The modelling also demonstrated that approximately 66% of external transmissions occurred from a host becoming infected whilst visiting an infected household, with the complimen-

tary percentage accounting for the infection of mosquitoes in susceptible households by infected visiting hosts. Hence, this thesis provides evidence that public health measures that result in a reduction of movement between households should significantly slow the spread of dengue. This would provide public health officials with additional time to implement more long term interventions and thereby potentially prevent an epidemic from occurring.

To test the underlying assumptions of the model, a new model was then developed that better accounted for the transmission between households by considering the explicit movement of hosts. To ensure that the level of mixing α from the branching process model was consistent with the new model some assumptions had to be made. Hosts must return home after visiting a household and the proportion of time a host spent outside their primary household is determined by the value of α . A level of correspondence was found by comparing the early growth rates of both models. although the additional complexity of calculating R_* that arises from removing the independence of infected households meant that comparisons of R_* could not easily be made. The results suggest that the branching process captured the main dynamics seen in the early stages of an outbreak.

6.1 Limitations and Further Work

A major limitation of the models is that their accuracy has not been tested against outbreak data. This would allow us to establish whether we have (or have not) fully captured the key underlying dynamics of vector-borne disease outbreaks. For example, the duration of each state was assumed to be exponentially distributed, as is natural with a CTMC construction. However, this has yet to be substantiated with data. Further modelling using dengue data would allow us to test this assumption and further refine the model as necessary (e.g., this might be achieved by incorporating additional infected classes to manipulate the overall distribution of infectious period as seen in Black *et al.* [6]). Another assumption made in the modelling was that mosquito population sizes per household remained constant and that they do not move between households. Given that mosquitoes are known to be more prevalent in highly populated regions. [28], modelling the effects of travelling mosquito populations may improve accuracy. This would involve the creation of a comprehensive spatial model, such that vectors could only bite households near their location. By adapting work from Binder *et al.* [5] another spatial component could be added by allowing this movement of mosquitoes to nearby households

(which could also be applied to hosts). The host movement model could be easily extended in accordance to census data by assigning varying numbers of hosts per household. The short lifespan of the mosquito perhaps suggests that the mosquito population size will change across the early stages of an outbreak and it may be useful to incorporate these temporal effects into any new modelling. Moreover, it has been demonstrated that mosquito activity rates peak during sunrise and sunset, with mosquitoes lying dormant at other times of the day. By accounting for the particular times of day when population mixing occurs it would be expected that the overall accuracy of the model would increase.

There are a number of other ways in which the framework of the branching process approximation and the host movement model might be extended. Kong *et al.* [21], for example, have extended an SEIR-SEI model to include the reproduction process of mosquitoes, which resulted in the creation of three additional states. Furthermore, modelling that accounts for the spread of the three other serotypes of dengue could provide further insight into how outbreaks occur given that infection from multiple serotypes increases the risk of severe dengue [28]. By adapting household structure and SEIR-SEI methodologies outlined in House and Keeling [18] and Dantas *et al.* [10] respectively, a deterministic model (with the same transitions) could be used to approximate a CTMC of the entire population. This may provide further insight into the longer-term dynamics of an outbreak, as a deterministic model would not assume an infinite population. Such a model would be substantially more efficient than the stochastic modelling due to the dynamics being described by a set of coupled ordinary differential equations. These results would show the average path of the CTMC. Providing information regarding the total proportion of the population infected as well as whether the disease will become endemic. Hence, a combination of stochastic and deterministic modelling would allow for further analysis of epidemics and the effects of interventions. The final major extension would be to define a second level grouping beyond households to model transmission seen in environments such as schools and workplaces, creating a network of mixing [8].

The suitability of the host movement model for use with large populations could also be improved by implementing the simulation in a language such as C++ rather than MATLAB, as efficiency is currently an issue for the model. Extensions upon this model could result in a substantial increase in computational power required to run the resulting simulations which may only produce a limited return. Hence, further research should be invested in the branching process approximations for the early

stages of an outbreak, and a deterministic model for later stages. From a public health perspective, it may well be that the current branching process modelling is sufficient to support the implementation of appropriate epidemic control strategies.

6.2 Conclusion

Given the increasing threat that dengue presents to northern Australia, accurate modelling of how transmission can lead to an epidemic is critical to the development of effective prevention strategies. This thesis demonstrates how a branching process approximation can be used to extend previous work and create a model that more accurately determines the likelihood and potential severity of such an outbreak. A number of possible extensions have been identified to further target vector-borne disease dynamics. An extended model that accounted for host movement between households was also created and appears promising, although much more computationally expensive. More work is required to form true correspondence between the two models. However, the evidence suggests that the branching process model accurately approximates between household host movement and that it is suitable to model the early stages of a dengue epidemic.

Bibliography

- [1] S. Akhtar, T. E. Carpenter, and S. K. Rathi. A chain-binomial model for intra-household spread of Mycobacterium tuberculosis in a low socio-economic setting in Pakistan. *Epidemiol Infect.*, 135(1):27–33, 2007.
- [2] M. Andraud, N. Hens, C. Marais, and P. Beutels. Dynamic Epidemiological Models for Dengue Transmission: A Systematic Review of Structural Approaches. *PLoS ONE*, 7(11):1–14, 2012.
- [3] F. Ball, D. Morrison, and G. Scalia-Tomba. Epidemics with two levels of mixing. *Ann App Prob*, 7(1):46–89, 1997.
- [4] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. Wint, C. P. Simmons, T. W. Scott, J. J. Farrar, and S. I. Hay. The global distribution and burden of dengue. *Nature*, 496(7446):504–507, 2013.
- [5] B. J. Binder, J. V. Ross, and M. Simpson. A hybrid model for studying spatial aspects of infectious diseases. *The ANZIAM Journal*, 54(1-2):37–49, 2012.
- [6] A. J. Black, T. House, M. J. Keeling, and J. V. Ross. Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza. *J. R. Soc. Interface*, 10:20121019, 2013.
- [7] O. J. Brady, P. W. Gething, S. Bhatt, J. P. Messina, J. S. Brownstein, and et al. Refining the global spetial limits of dengue virus transmission by evidence based consensus. *PLOS Neglected Tropical Diseases*, 6(8):e1760, 2012.
- [8] T. Britton, T. Kypraios, and P. O'Neill. Inference for Epidemics with Three Levels of Mixing: Methodology and Application to a Measles Outbreak. *Scandinavian Journal of Statistics*, 38:578–599, 2011.

- [9] M. Chan and M.A Johansson. The Incubation Periods of Dengue Viruses. *PLoS ONE*, 7(11):e50972, 2012.
- [10] E. Dantas, M. Tosin, and Americo Cunha Jr. Calibration of a SEIR-SEI epidemic model to describe the Zika virus outbreak in Brazil. *Applied Mathematics and Computation*, 338:249–259, 2018.
- [11] P. L. Delameter, E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen. Complexity of the Basic Reproduction Number (R_0). *Emerg Infect Dis.*, 25(1):1–4, 2019.
- [12] O. Diekmann, J. H. Heesterbeek, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, Princeton, New Jersey, 2013.
- [13] R. Durrett. *Essentials of Stochastic Processes*. Springer International Publishing, New York, 2012.
- [14] R.A. Erickson, S.M. Presley, L.J.S. Allen, K.R. Long, and S.B Cox. A dengue model with a dynamic Aedes albopictus vector population. *Ecological Modelling*, 221(24):2899–2908, 2010.
- [15] Centers for Disease Control and Prevention. Dengue. <https://www.cdc.gov/dengue/about/index.html>, 2019. Accessed: 2020-10-13.
- [16] Queensland Government. Dengue: virus, fever and mosquitoes. <https://www.health.qld.gov.au/clinical-practice/guidelines-procedures/diseases-infection/diseases/mosquito-borne/dengue/virus-fever>, 2020. Accessed: 2020-10-02.
- [17] M. Greenwood. The statistical measure of infectiousness. *Journal of Hygiene*, 31:336–351, 2011.
- [18] T. House and M. J. Keeling. Deterministic epidemic models with explicit household structure. *Math. Biosci.*, 213(1):29–39, 2000.
- [19] T. House and M. J. Keeling. Household structure and infectious disease transmission. *Epidemiol Infect*, 137(5):654–661, 2008.
- [20] M.A. Johansson, J. Hombach, and D.A.T Cummings. Models of the impact of dengue vaccines: A review of current research and potential approaches. *Vaccine*, 29(35):5860–5868, 2011.

- [21] L. Kong, J. Wang, and et. al. Modeling the Heterogeneity of Dengue Transmission in a City. *Int J Environ Res Public Health*, 15(6):1128, 2018.
- [22] Nature. Dengue Transmission. <https://www.nature.com/scitable/topicpage/dengue-transmission-22399758/#:~:text=No%20once%20infected%20with%20dengue,%2D%20to%20four%2Dweek%20period>, 2014. Accessed: 2020-10-08.
- [23] E. Newton and P. Reiter. A Model of the Transmission of Dengue Fever with an Evaluation of the Impact of Ultra-Low Volume (ULV) Insecticide Applications on Dengue Epidemics. *The American Journal of Tropical Medicine and Hygiene*, 47(6):709–720, 1992.
- [24] J. R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, 1997.
- [25] County of Los Angeles Department of Public Health Acute Communicable Disease Control. Acute Communicable Disease Control. <http://publichealth.lacounty.gov/acd/VectorDengue.htm>. Accessed: 2020-10-26.
- [26] Australian Bureau of Statistics. Census of Population and Housing. <https://www.abs.gov.au/census>, 2016. Accessed: 2020-07-20.
- [27] M. Oki, T. Sunahara, M. Hashizume, and T. Yamamoto. Optimal timing of insecticide fogging to minimize dengue cases: Modeling dengue transmission among various seasonalities and transmission intensities. *PLOS Neglected Tropical Diseases*, 5(10):1–7, 2011.
- [28] World Health Organisation. Dengue and severe dengue. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>, 2020. Accessed: 2020-10-14.
- [29] World Health Organisation. Dengue control. <https://www.who.int/denguecontrol/mosquito/en/>, 2020. Accessed: 2020-10-13.
- [30] M. Otero, D.H. Barmak, C.O. Dorso, H.G. Solari, and M.A. Natiello. Modeling dengue outbreaks. *Math. Biosci.*, 232(2):87–95, 2011.
- [31] P. K. Pollet and V. T. Stefanov. Path Integrals for Continuous-time Markov Chains. *Journal of Applied Probability*, 39(4):901–904, 2002.
- [32] J. V. Ross, T. House, and M. J. Keeling. Calculation of Disease Dynamics in a Population of Households. *PLoS ONE*, 5(3):1–9, 2010.

- [33] G. Scalia-Tomba. Asymptotic Final Size Distribution of the Multitype Reed-Frost Proces. *Journal of Applied Probability*, 23(3):563–584, 1986.
- [34] R. B. Sidje. EXPOKIT. A software package for computing matrix exponentials. *ACM Trans. Math. Softw.*, 24(1):130–156.
- [35] D.W. Vaughn, S. Green, S. Kalayanarooj, B.L. Innis, S. Nimmannitya, S. Sun-tayakorn, T.P. Endy, B. Raengsakulrach, A.L. Rothman, and F.A. et al. Ennis. Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J. Infect. Dis.*, 181(1):2–9, 2000.