Household Models for Endemic Diseases

Jonathon Pantelis

March 8, 2017

Thesis submitted for the degree of

Master of Philosophy

in

Applied Mathematics

at The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: ..... Date: .....

Acknowledgements

This project has been a journey, a journey that I often could not see the end of. Thankfully my supervisors, Josh Ross and Andrew Black, are absolutely amazing, patient people with brilliant minds. I honestly cannot thank them enough.

I would also like to thank my family and friends for listening to me talk about my research incessantly. I want to especially thank my mum, Kris, my dad, Jim, my brother, Will, and my girlfriend, Kiara, for putting up with me, and pushing me each and every step. They were the first people to hear of each and every success and failure.

I have never been very good at writing these 'thank you' things, but just know that if I ever talked to you about what this thesis contains, or if you are reading this now, I am grateful for you lending me your ears.

Abstract

The mathematical modelling of disease dynamics is now well-established, which allows us to better understand the processes affecting diseases in communities. Most of this work has focused on *epidemic* dynamics, in which we see a large outbreak of a disease but also its extinction within a short time-frame. Of interest in this thesis will be furthering the study of *endemic* diseases, which persist in populations over much longer time scales. This increased time scale means that it is important to account for demography, and in particular the replenishment of susceptible individuals through births and possibly waning immunity. Another feature that has been shown to be important for modelling disease dynamics is heterogeneity in the population, and in particular household structure. This is likely to be important for endemic diseases because of the close relationship between demography and household composition as these types of diseases persist in communities for periods long enough that assuming a static population structure is not realistic.

Hence we model both the disease and population dynamics as a continuous-time Markov chain, where the population of individuals is split into adults and children. These adults and children are assigned a household type, and within each household the number of adults is at most two, and the number of children is at most four. These households change through time as a consequence of demographic events such as births, deaths, children moving out of home, couples forming, separating, and migration. This demographic model is overlaid

with SEIR disease dynamics, where the rate of disease transmission is strongest within households, as opposed to between. We also develop a seasonally-forced model. The so-called Gillespie algorithm is used to simulate realisations of this process but simulation can be computationally expensive. Hence we also derive a deterministic approximation, valid in the limit of a large number of households. We use this deterministic approximation extensively to analyse the models dynamics.

For measles-like disease parameters, the period of the non-seasonally forced model is approximately two years, which agrees closely to the established biennial periodicity. In the model with seasonal forcing, the periodicity is more complicated, with annual periodicity in the deterministic approximation, and approximately 2-3 year periods for individual stochastic realisations.

The household model without seasonal forcing is used to investigate the case of a flu-like disease suddenly becoming fatal in children, adults, and both. We found that when the population has only a small proportion of susceptible individuals there is a higher chance of the disease being persistent when compared to the population with larger proportions of susceptible individuals. Significant change in household proportions are recorded initially during the first outbreak of the disease when the mortality rate increases during the infectious period of individuals.

In both the models with and without seasonal forcing, a measles-like disease fades out with higher probability in the household models compared to the homogeneous models. In each of the household models, within the first five years, we see periods of approximately 2-3 years where the proportion of realisations that fade out stay essentially constant for populations of greater than 225 thousand households in the seasonally-forced model, and above 250 thousand in the non-forced model. During these periods, the proportion of realisations that have not faded out is above 0.5, which is suggestive of a critical community size between 375-500

thousand individuals in both household models. However, the household model shows greater variability than the homogeneous models as a larger proportion of realisations fadeout over the course of 15 years.

The household models presented in this thesis allow a great deal of flexibility in parameterisation which means that many problems can be studied. However, as a consequence of the flexibility, the model's dynamics are evaluated at large computational expense, meaning approximations are necessary.

Chapter 1

Introduction

The study of infectious diseases has long been a topic of particular importance due to their impact. There are primarily two types of disease dynamics that are of interest to mathematicians. The first are *epidemics*, which is when there is an outbreak of a pathogen and its subsequent remission. However, many pathogens of concern are *endemic* to communities, meaning that following an initial outbreak, they form a recurring and persistent burden on health and well-being.

In order to understand the dynamics of diseases, we use mathematical models. The models we consider operate on the assumption that individuals can be classified as *susceptible*, *infectious* or *recovered* [1, 2]. A susceptible individual does not have the pathogen, but is able to contract the disease. An infectious individual is someone who has the pathogen in their system and is able to transmit the disease to susceptible individuals. Recovered individuals are those without the pathogen and are unable to contract the disease due to immunity.

Kermack and McKendrick [3] developed equations to model disease dynamics and it was this theory that has been studied and further developed into what is now known as the modern 'Susceptible-Infected-Recovered' (SIR) model. In the modern formulation of the model, a susceptible individual becomes infectious at

a rate of

$$\frac{\beta SI}{N-1} \tag{1.1}$$

where N is the total number of individuals in the population, S and I are the number of susceptible and infectious individuals in the population, respectively, and β is a parameter that quantifies how contagious the disease is as well as incorporating how frequently individuals within the population interact with one another. For example, in a town with people isolated from one another, a lower β value, compared with that in a busy city where people make contact with other people at higher rates, is expected. An infected individual recovers and hence becomes immune to the disease with a mean recovery time of $\frac{1}{\gamma}$, as the time between events is assumed to be distributed exponentially.

In the case of some diseases, such as measles and influenza, susceptible individuals that contract the infectious disease do not immediately become infectious [4]—that is, they have the pathogen but are not contagious—for some period of time; this is the *exposed* class (E). Hence, we make use of the SEIR model [1] throughout this thesis, which is similar to the SIR model. In this model, one of the two main differences is that when a susceptible individual contracts the disease, they become exposed. This occurs at the same rate as for the SIR model; that is to say that an individual becomes exposed at rate $\frac{\beta SI}{N-1}$. The second difference is that the exposed individual becomes infectious after a latent period with mean $\frac{1}{\sigma}$. The event that transitions an infectious individual into the recovered class is the same as in the SIR model. For many diseases, immunity is life-long and so there is no event that makes a recovered individual susceptible again. If immunity is not permanent, however, a recovered individual can become susceptible to the disease again after a period with mean $\frac{1}{\omega}$. Other variations of these compartmental models exist but are not detailed here [1].

These models are quite simple in construction and assume, for example, that

there is a fixed population size. That is, there are no events that correspond to new individuals entering the population and old ones leaving. An attempt to relax this particular assumption comes in the form of a simple demography framework where births and deaths are introduced. Here we detail two possible approaches to implement births and deaths. The first method is linking the two events such that whenever a death occurs, a birth happens at the same time [5]—which keeps the population size constant. The other method treats the two events separately where births occur at some rate and deaths occur at some other rate. These rates are often assumed to be equal [1, 6] to keep the population approximately equal over long periods of time.

These models with births and deaths are of particular importance to endemic diseases. An example of an endemic disease is measles prior to the introduction, in 1968, of mass vaccinations in England and Wales [2]. The number of weekly cases of measles peaks approximately every two years, but never entirely fades out between these peaks. The main reason why measles—and endemic diseases, in general—persists in particular communities is that the number of susceptible individuals in the population are replenished over time, as discussed in [7, 8]. Hence, when the birth rate is large enough, new susceptible individuals are born into the population at a rate sufficient for the pathogen of the disease find new hosts [1].

There are many diseases which are assumed to have life-long immunity [9, 8, 1], but many diseases do not have this property. A consequence of diseases without life-long immunity is that no one is ever immune to the disease forever, which means after some time (which in the case of the common cold is short, and is so long in the case of measles [10] that the disease was considered to have a life-long immunity for many years [11]) individuals are susceptible to contracting the disease again and the disease does not have an opportunity to die out in the population, assuming a large enough population size.

After an initial outbreak of measles, the disease will either fadeout or another outbreak will occur. The stochastic nature of disease transmission makes it difficult to determine whether the disease will fadeout or not, but Bartlett introduced the notion of a *critical community size* [12], which is a measure of how large a population must be in order for the disease to persist in the community more often than the disease fades out. A large population size sees the disease persist more frequently, compared to a small population size, as more births occur and hence there are more susceptible individuals for infectious individuals to transmit the disease to. In England, cities like Manchester, with populations of approximately 680,000 people, had a persistent problem with measles; however in smaller communities, such as Cardigan (under 4000 people), measles faded out, and only two epidemics arose between 1940 and 1956 [12]. Bartlett identified that the connection between cities and the geographic dispersion of the communities [13] also plays a role in determining the critical community size, and is now largely believed to be between 250,000 and 400,000 individuals [14].

The models presented for endemic diseases, even with simple demography, assume that the population mixes homogeneously. That is, each individual has equal probability of contacting any other individual in the population. This is not valid because people have social and work contacts that they interact with most frequently. To relax this assumption of homogeneous mixing, models with two levels of mixing are utilised [15, 16]. In these models, populations are composed of smaller populations, called *meta-populations* (also known as subpopulations). The idea is that the population is split into smaller subpopulations where infection occurs at different rates within a meta-population to between them. Ball, Mollison and Scalia-Tomba [17] generalised the concept of an infectious disease model with two levels of mixing. In these models, individuals are partitioned into non-overlapping groups that attempt to mimic how people interact within social settings (and hence not entirely homogeneously). A popular social struc-

ture used in models with two levels of mixing is grouping the individuals in the population into households. There are two disease transmission rates: a local transmission rate for the spread of disease within each of these households, and a global transmission rate for the spread of disease between these households, where the transmission rate within households is assumed to be at least as big as the between-households transmission rate.

Typically, now, the rate of transmission within a household k is given by

$$\frac{\beta S_k I_k}{n_k - 1},$$

where S_k and I_k are the number of susceptible and infected individuals, respectively, in the household k, n_k is the number of individuals in the household k, and β is a parameter that quantifies how contagious the disease is as well as incorporating how frequently individuals within households interact with one another. On the other hand, we have a global (or homogeneous) level of the disease infecting an individual in household k typically with rate,

$$\frac{\alpha S_k I^T}{N - 1},$$

where I^T is the total number of infectious individuals in the entire population, N is the total number of individuals in the population, and α is the parameter quantifying how rapidly the disease spreads among individuals that do not occupy the same household. As households are constructed in such a way that individuals that interact with one another more frequently are grouped together, the between-household transmission rate is smaller than the within-household transmission rate; that is, $\alpha \leq \beta$.

While there has been some progress in developing household models, they are mostly concerned with epidemic diseases, rather than endemic diseases [18]. Just

as the homogeneous models originally introduced did not define any demographic events, a majority of the household models lack the demography dynamics that are so important in modelling endemic diseases. Over the course of a number of decades, a length of time that is typical when discussing endemic diseases, real households change significantly: children in households will grow up and in most cases move out of home, possibly forming new partnerships and having children. The models described thus far do not allow for any change in household structure at all.

A simple way to address this lack of population dynamics is to introduce births and deaths in a similar manner to the homogeneous models. For example, as adopted in one of the models in Cross *et al* [8]. This allows the population of households to have individuals that die and are hence removed from the population and also introduce new susceptible individuals through birth events. By incorporating these events, simple population dynamics are replicated which in turn allows us to model endemic diseases as susceptible individuals are being replenished.

Although introducing births and deaths allows us to model endemic diseases, two models extend the demographic dynamics to include movement between households. The first is in Glass *et al* [19], which adopts a modified Reed-Frost model [16], with generation length determined arbitrarily, that allows for transmission of the disease within and between households; this stochastic, generation-based disease model is overlaid with a deterministic model for the population dynamics. The population is split into households of sizes ranging from 1 to 6. In these households, it is impossible to differentiate between children and adults; that is, there is no age structure. These households are able to undergo a number of population events, such as births, deaths, and movement between households. Specifically, a household with five or less individuals can experience a birth event, in which case the number of individuals in the household increases by one. There

are a number of different parametrisations for the birth rate. One formulation assumes that the larger household sizes have more births compared to smaller households, while another assumes the opposite: smaller households have a larger birth rate in comparison to the households with a larger size.

Death events can also occur in these households, where one individual is removed from the household. Just as there are different parametrisations for the birth rate, there are multiple formulations for the death rate also. Let $d(n)$ be the death rate for a household with n individuals in it, then three proposed forms of rates are $d(n) = nd_0$, $d = (1, 2, 2, 2, 2, 2)d_0$, and $d(n) = d_0$, where d_0 is the death event parameter. These rates correspond to the following interpretations: each individual has a constant death rate; the second formulation is similar to the first, except it is assumed that at most two individuals are ever at risk of death, which could correspond to an assumption that only adults are ever at risk of death; the last rate formulation is the case where each household is equally likely to have a death occur in it, independent of household size. The model is calibrated so that the births and deaths cancel out over the course of the year; that is, the expected population size is assumed fixed.

There are also events where one individual leaves a household to move into a new household by themselves, and where someone who lives by themselves can move into another household, assuming that when moving into the new household there are no more than six individuals. The rates of these events also have multiple formulations. These include situations where larger households have more movement between households, and where smaller households have more movement, and also a formulation where everyone is equally likely to move out or into a new household. These events, where individuals leave and enter households, are calibrated over a year in order to keep the proportion of the household types constant.

The model in Glass *et al* [19] relaxes the assumption of a static population in a

way that is simple and flexible. The deterministic demographic dynamics lends itself to being efficiently analysed, which means that a thorough understanding of the behaviour of the model can be achieved. However, a deterministic framework for the demographic events does not allow any stochastic behaviour which is not necessarily a realistic assumption. Also, there are many diseases that impact children and adults differently, for example: measles, and this model is unable to distinguish between children and adults. These are two particular limitations of the model that we would like to remedy.

Another model that considers changing household structure is that of Geard *et al* [20]. The demography is modelled by a discrete-time individual-based model, where each time step corresponds to an arbitrarily picked number of days. What this means is that each individual in the population is characterised by their age, sex, and the household they belong to. Then each household can undergo birth and death events, in addition to couple formation and dissolution events, and children are able to move out of home. The probability of these events occurring are rather complicated as they rely on the sex and the specific age of the individuals. For example, women have a longer life expectancy and so the formulation for a woman dying in the population is different to that of a man dying. Further, realistically, only women can have children and the probability that a woman has a child changes depending on the age of the woman. As these probabilities rely on a large number of parameters, the demography is realistic compared to data.

This individual-based model is further extended to include disease dynamics by Geard *et al* [21]. The model is used to discuss a variety of scenarios. For example, comparing demographic traits between historic and modern years—such as community contacts between people of different ages, and household distribution—and discussing how these changes impact disease dynamics. They also investigate the effectiveness of some vaccination strategies; that is, the effect

of vaccinating based on what household an individual belongs to or vaccinating each individual with some probability ν regardless of their household.

Although the individual-based demographic model of Geard *et al* [21] is incredibly realistic, it relies upon simulations to analyse, which is very computationally expensive; they simulate only ten realisations per scenario modelled. Where Glass' model was simple yet efficient to analyse, Geard's is the opposite in that it is very realistic but inefficient to analyse. These two models occupy different ends of the complexity scale; our goal is to construct a model that fits somewhere between these—easier to analyse than the model in Geard *et al* [20, 21], but more detailed than in Glass *et al* [19]. That is, we introduce a model that incorporates age structure, which allows additional demographic events and more detailed transmission rates compared to [19]. On the other hand, we are able to derive a deterministic approximation which was not done in [20, 21], and so we can evaluate the dynamics more efficiently.

Our model can be used to investigate a number of problems. In particular, we explore epidemic fadeout and discuss the critical community size for measles in our model. We also introduce a flu-like disease into a population where individuals that are infectious have an increased mortality rate. From this, we learn how population dynamics are affected by varying levels of fatality in disease, and we are able to deduce some of the properties that our household model possesses.

1.1 Summary of chapters

Chapter 2 details the technical background required to understand the mathematics in this thesis, which includes definitions of continuous-time Markov chains and a selection of relevant properties and theorems pertaining to them. Chapter 2 introduces how we stochastically evaluate the model's dynamics using the

so-called Gillespie's algorithm [22], before introducing how we can derive a deterministic approximation using theory proved by Kurtz [23].

Chapter 3 develops a continuous-time Markov chain that models the demographic dynamics that will eventually feature alongside disease dynamics in Chapter 4. In Chapter 3, we not only define the events of the demography model, and their rates, but also detail how stochastic simulations are computed and the deterministic model is derived before finally parameterising the demography model in two different ways.

Chapter 4 extends the continuous-time Markov chain in Chapter 3 by introducing disease dynamics. In much the same way as in Chapter 3, we detail the events and their rates, how the stochastic simulations are computed and derive the deterministic approximation for the disease with demography model. We also introduce seasonal forcing into our model, and provide baseline values for the disease parameters.

Chapter 5 displays the results that our disease with demography household model yields. Here, the periodicity of the household model, both with and without seasonal forcing, is explored. We also investigate how a flu-like disease that increases the mortality rate of infectious individuals impacts on the population structure, and lastly, we look at the problem of epidemic fadeout, where we try to find links between the population size and the persistence of disease.

The last chapter, chapter 6, discusses the results and talks about the properties of the household models researched in this thesis. We discuss the design choices made, the limitations and assumptions, and future work that could be explored.

Chapter 2

Technical background

In this chapter, background on the theory that is utilised throughout this thesis is detailed. Continuous-time Markov chains are defined and some of their important properties discussed, as well as how one may analyse a system that is modelled as a continuous-time Markov chain.

2.1 Continuous-time Markov chains

Let $(X(t), t \geq 0)$ be a stochastic process which can be in any of a finite set of states, $\mathcal{S}$, which is referred to as the state space. The process $X(t)$ is called a *continuous-time Markov chain* (CTMC) if,

$$\mathbb{P}\left(X(t+s) = j | X(s) = i_s, X(u) = i_u, u \leq s\right) = \mathbb{P}\left(X(t+s) = j | X(s) = i_s\right)$$

for all $s, t \in [0, \infty)$ and all $i_s, i_u, j \in \mathcal{S}$. This expression—the Markov property—means that how the process evolves in the future, conditioned on the processes history, depends only upon the present and not on the past. This property is often referred to as the process being *memoryless*.

A CTMC is said to be *time-homogeneous* if,

$$\mathbb{P}\left(X(t+s)=j|X(s)=i\right)=\mathbb{P}\left(X(t)=j|X(0)=i\right),$$

for all $i,j\in\mathcal{S}$ and $s,t\in[0,\infty)$. Throughout this thesis, all CTMC are defined to be time-homogeneous. Hence, we can define,

$$p_{ij}(t)=\mathbb{P}\left(X(t+s)=j|X(s)=i\right),$$

for all $i,j\in\mathcal{S}$ and $s,t\in[0,\infty)$. This specifies the probability of the process moving from state i to state j in a time interval of length t. As $p_{ij}(t)$ is a probability, we must have $p_{ij}(t)\geq 0$ for all time intervals t, and for each $i\in\mathcal{S}$,

$$\sum_{j\in\mathcal{S}}p_{ij}(t)=1$$

because the process must be in one of the states for all times t.

A CTMC is usually defined by its infinitesimal rates. These rates are often presented as elements in a matrix known as the *infinitesimal generator*, or Q-matrix, denoted Q. Define the elements of the Q-matrix to be,

$$\begin{aligned}q_{ij}&=\lim_{h\to 0^+}\frac{p_{ij}(h)}{h}\quad\text{for } i,j\in\mathcal{S}, j\neq i,\\ q_{ii}&=\lim_{h\to 0^+}\frac{p_{ii}(h)-1}{h}\quad\text{for } i\in\mathcal{S}.\end{aligned}$$

The entries of this matrix, q_{ij}, $i\neq j$, are the instantaneous rates at which the process $X(t)$ moves from state i to state j, and $-q_{ii}$ is the instantaneous rate that the process $X(t)$ leaves state i. This can be shown by considering,

$$1-p_{ii}(h)=\sum_{j\neq i}p_{ij}(h).$$

Then,

$$\begin{aligned} -q_{ii} &= \lim_{h\to 0^+} \frac{1-p_{ii}(h)}{h} \\ &= \lim_{h\to 0^+} \sum_{j\neq i} \frac{p_{ij}(h)}{h} \\ &= \sum_{j\neq i} \lim_{h\to 0^+} \frac{p_{ij}(h)}{h} \\ &= \sum_{j\neq i} q_{ij}. \end{aligned}$$

Thus the diagonal elements of the Q-matrix, q_{ii}, are just the negative of the sum of the off-diagonal elements in row i and so the row sums to 0.

We have defined how the process transitions from state to state, but need to discuss the time between events. Let T_i be a random variable that takes the value of the time that the process $\{X(t) : t > 0\}$ spends in state i before first exiting. Then T_i is said to have the memoryless property if

$$\mathbb{P}\left(T_i > s+u | T_i > s\right) = \mathbb{P}\left(T_i > u\right) \quad \text{for } s, u \in \mathbb{R}^+.$$

For $s > 0$, the event $\{T_i > s\}$ is equivalent to $\{X(t) = i \text{ for } 0 < t < s\}$. Then let us consider the probability that the time spent in state i is greater than some time $s+u$, given the process has already spent time s in state i:

$$\begin{aligned} \mathbb{P}\left(T_i > s+u | T_i > s\right) &= \mathbb{P}\left(X(t) = i \text{ for } 0 < t < s+u | X(t) = i \text{ for } 0 < t < s\right) \\ &\quad \text{by the equivalent events defined above,} \\ &= \mathbb{P}\left(X(t) = i \text{ for } s < t < s+u | X(t) = i \text{ for } 0 < t < s\right) \\ &\quad \text{as the process is conditioned on } X(t) = i \text{ for } 0 < t < s, \\ &= \mathbb{P}\left(X(t) = i \text{ for } s < t < s+u | X(s) = i\right) \\ &\quad \text{by the Markov property,} \end{aligned}$$

$$= \mathbb{P}\left(X(t) = i \text{ for } 0 < t < u | X(0) = i\right)$$

by time homogeneity,

$$= \mathbb{P}\left(T_i > u\right)$$

This shows that the time in each state is memoryless, which implies that the time spent in each state is exponentially distributed because this is the only continuous distribution with such a property. Hence, as $-q_{ii}$ can be interpreted as the instantaneous rate of the process leaving state i, then the process leaves state i at an exponentially distributed time with rate $\sum_{j \neq i} q_{ij}$.
Of particular interest is the value $p_{ij}(t)$ as this tells us the probability of being in state j after t time, given the process started in state i. By considering, for $0 < u < t$, and $s > 0$, we have,

$$\begin{aligned}
p_{ij}(t) &= \mathbb{P}\left(X(t+s) = j | X(s) = i\right) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X(t+s) = j, X(s+u) = k | X(s) = i\right) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X(t+s) = j | X(s+u) = k, X(s) = i\right) \mathbb{P}\left(X(s+u) = k | X(s) = i\right) \\
&\text{by the memoryless property and time homogeneity} \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}\left(X(t) = j | X(u) = k\right) \mathbb{P}\left(X(u) = k | X(0) = i\right) \\
&= \sum_{k \in \mathcal{S}} p_{kj}(t-u) p_{ik}(u).
\end{aligned}$$

This final expression is known as the Chapman-Kolmogorov equation [24]. This equation is used directly to derive the Kolmogorov forward equations, which can be used to determine the term $p_{ij}(t)$. Consider,

$$p_{ij}(t+h) = \sum_{k \in \mathcal{S}} p_{ik}(t) p_{kj}(h).$$

Now let us take the derivative of $p_{ij}(t)$, in the first principles sense,

$$\begin{aligned}
\frac{\mathrm{d}p_{ij}(t)}{\mathrm{d}t} &= \lim_{h\to 0^+} \frac{p_{ij}(t+h) - p_{ij}(t)}{h} \\
&= \lim_{h\to 0^+} \frac{\left(\sum_{k\in\mathcal{S}} p_{ik}(t)p_{kj}(h)\right) - p_{ij}(t)}{h} \\
&= \lim_{h\to 0^+} \frac{\left(\sum_{\substack{k\neq j \\ k\in\mathcal{S}}} p_{ik}(t)p_{kj}(h)\right) + p_{ij}(t)p_{jj}(h) - p_{ij}(t)}{h} \\
&= \lim_{h\to 0^+} \left[\frac{\left(\sum_{\substack{k\neq j \\ k\in\mathcal{S}}} p_{ik}(t)p_{kj}(h)\right)}{h} - \frac{p_{ij}(t)\left(1 - p_{jj}(h)\right)}{h}\right] \\
&= \left(\sum_{\substack{k\neq j \\ k\in\mathcal{S}}} p_{ik}(t)q_{kj}\right) + p_{ij}q_{jj}, \text{ interchanging the limit and sum.}
\end{aligned}$$

This sum is the Kolmogorov forward equations [24],

$$\frac{\mathrm{d}p_{ij}(t)}{\mathrm{d}t} = \sum_{k\in\mathcal{S}} p_{ik}(t)q_{ik}. \tag{2.1}$$

2.1.1 Example of a CTMC

As this thesis is concerned with populations and disease dynamics, we contextualise all the above information into a relevant example to introduce and familiarise notation. Let us consider the SIR compartmental model, as first introduced in Chapter 1. The state space, $\mathcal{S}$, is the set of all possible combinations in which N individuals can be arranged into the three types: susceptible, infectious, and recovered. Because the population size, N, is constant, we can remove one dimension. That is,

$$\mathcal{S} = \{(s,i) : s,i \geq 0,\ s+i \leq N\},$$

and the number of recovered individuals is given by the equation $r = N - s - i$. The possible events are transmission and recovery, where a transmission event

has rate,

$$q_{(s,i),(s-1,i+1)} = \frac{\beta si}{N-1}$$

and a recovery event has rate,

$$q_{(s,i),(s,i-1)} = \gamma i.$$

Replenishment of susceptible individuals can be added by allowing birth and death events. To keep the population size fixed, the births and deaths are linked here. That is, when a death occurs, a new individual is born into the population as a susceptible. In a stable population, a child is born approximately whenever somebody dies. By linking the two events, we are effectively making a simple assumption that they happen simultaneously, which is only valid when the total number of individuals in the population is large. As a susceptible dying and being replaced by another susceptible effectively changes nothing, there are only two such death events, which have rates,

$$\begin{aligned} q_{(s,i),(s+1,i-1)} &= \mu i \\ q_{(s,i),(s+1,i)} &= \mu \left(N - s - i\right), \end{aligned}$$

where μ is the rate parameter for the death event, and hence $\frac{1}{\mu}$ is the average life expectancy. The reason that the births and deaths are linked is so that total population size is preserved, which simplifies analysis greatly. This will be discussed in more detail later.

2.2 Methods for evaluating the dynamics of the model

In order to gain insight from the models developed, we need to evaluate their dynamics. This section considers two methods that are used throughout this thesis. Ideally, the forward equations (Equation 2.1) would be used but finding the solution to these is not feasible for most of the models that we introduce later because the size of the state space, and hence the number of equations to solve, is simply too large. Instead, we introduce a simulation technique and a deterministic approximation that we use to evaluate the dynamics of our models.

2.2.1 Stochastic simulation

For a process, $X(t)$, realisations, or sample paths, can help to understand the behaviour of the process. Here, a method to exactly simulate these sample paths is described using the so-called Gillespie algorithm [22].

First, let $\mathcal{S} = \{1, ..., n\}$ be the finite state space of the process. Initialise the system by defining the parameters and setting the initial state of the process, say $X_{old} = i$, and setting the initial time $t_{old} = 0$.

Let $\mathbf{r} = (q_{1,2}, ..., q_{1,n}, q_{2,1}, ..., q_{n,n-1})$ be a vector with the rates of all the possible transitions in the process $X(t)$, and let Λ be the stoichiometric matrix such that each column of Λ is the change in the state of the process $X(t)$ corresponding to the event in $\mathbf{r}$. That is, the column vector Λ_j is the consequence of the event with corresponding rate in the jth component of $\mathbf{r}$, which we will now denote r_j.

The rate of the process $X(t)$ leaving the state i, which is denoted q_i, is calculated

by,

$$q_i = \sum_{j \neq i} q_{ij}.$$

As the CTMC will leave state i at an exponentially distributed time (with parameter q_i, the time until the process changes state is given by

$$\Delta t = -\frac{\log u_t}{q_i},$$

where u_t is a uniformly distributed random number between 0 and 1. Then,

$$t_{new} = t_{old} + \Delta t.$$

At this new time, the process leaves state i, and a transition to a new state must be chosen. To do this, we make use of another uniformly distributed random number between 0 and 1, u_e. Calculate $u_e q_i$ and let $l_e \in \mathbb{N}$ denote the index of the transition that is chosen. Calculate l_e by considering the index at which the cumulative sum of $\mathbf{r}$ first exceeds $u_e q_i$. That is,

$$l_e = \min \left\{ x \in \mathbb{N} : \sum_{k=1}^{x} r_k > u_e q_i \right\}.$$

The resulting state transition is given by the vector Λ_{l_e}, and hence,

$$X_{new} = X_{old} + \Lambda_{l_e}.$$

The process has undergone one transition and has moved forward in time by Δt. This method can be repeated until t_{new} reaches some endpoint, or some other criterion has been met.

2.2.2 Deterministic approximation

For small systems, running simulations to get realisations of the model is not very time consuming. However, for models with larger state spaces, simulations become more and more infeasible. In 1970, Kurtz published a paper [23] proving the theory on how to approximate the expected behaviour of a CTMC. In this section, we present the theory as relevant to our use in population and disease dynamics.

Let $\{X(t) = (x_1, ..., x_m) : t > 0\}$ be a CTMC with state space $\mathcal{S}$. Let the states of the CTMC have m variables; that is, there are m different 'types' of individuals in the population. Then let $\boldsymbol{x} = (x_1, ..., x_m) \in \mathcal{S}$, and hence $\boldsymbol{x} \in \mathbb{Z}^m$. Further, let N be the total population size; that is,

$$\sum_{k=1}^{m} x_k = N.$$

Now, let the non-zero vector $\boldsymbol{\eta} \in \mathbb{Z}^m$ be a transition vector that changes $\boldsymbol{x}$ such that after the event corresponding to $\boldsymbol{\eta}$, the new state is $\boldsymbol{x}_{new} = \boldsymbol{x}_{old} + \boldsymbol{\eta}$. The CTMC is said to be *density dependent* if for each event and transition $\boldsymbol{\eta}$,

$$q_{\boldsymbol{x},\boldsymbol{x}+\boldsymbol{\eta}} = Nw\left(\frac{1}{N}\boldsymbol{x}, \boldsymbol{\eta}\right),$$

where $w(\boldsymbol{\nu}, \mathbf{y})$ is a continuous function such that $\boldsymbol{\nu} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{Z}^m$.

If this property is satisfied, then define

$$F(\boldsymbol{\nu}) = \sum_{\boldsymbol{\eta}} \boldsymbol{\eta} w(\boldsymbol{\nu}, \boldsymbol{\eta}),$$

which is a vector where each component, say, k, corresponds to the total rate at which each of the different individual types transition into and out of the type k.

There are three mild technical conditions that need to be satisfied [23], which we do not detail, and all CTMCs in this thesis satisfy them. Hence, for every trajectory of $X(s, \boldsymbol{\nu}_0)$ with initial condition $X(0, \boldsymbol{\nu}_0) = \boldsymbol{\nu}_0$, where $0 \leq s \leq t$, and,

$$\frac{\partial}{\partial s} X(s, \boldsymbol{\nu}_0) = F\left(X(s, \boldsymbol{\nu}_0)\right),$$

then $\lim_{N\to\infty} \frac{1}{N} X(0) = \boldsymbol{\nu}_0$ implies for every $\delta > 0$,

$$\lim_{N\to\infty} P\left(\sup_{s\leq t} \left| \frac{1}{N} X(s) - X(s, \boldsymbol{\nu}_0) \right| > \delta \right) = 0.$$

This expression means that as the population size approaches infinity, the scaled stochastic model $X(t)$ converges, uniformly in probability, over finite time intervals to the deterministic approximation. This can be interpreted more simply as the solution to the differential equations that are constructed from the rates of the events and their transition vectors is in fact the expected behaviour of the CTMC as the population size N gets infinitely large.

This theory allows us to approximate the expected behaviour of a CTMC by constructing the differential equations by considering each of the events, their rates and their transitions, and then solving these differential equations. Figure 2.1 demonstrates the link between the deterministic trajectory and the average of the realisations for the model defined in Section 2.2.3. It can be seen that the average of the realisations follows the trajectory of the deterministic solution closely, even with only ten realisations.

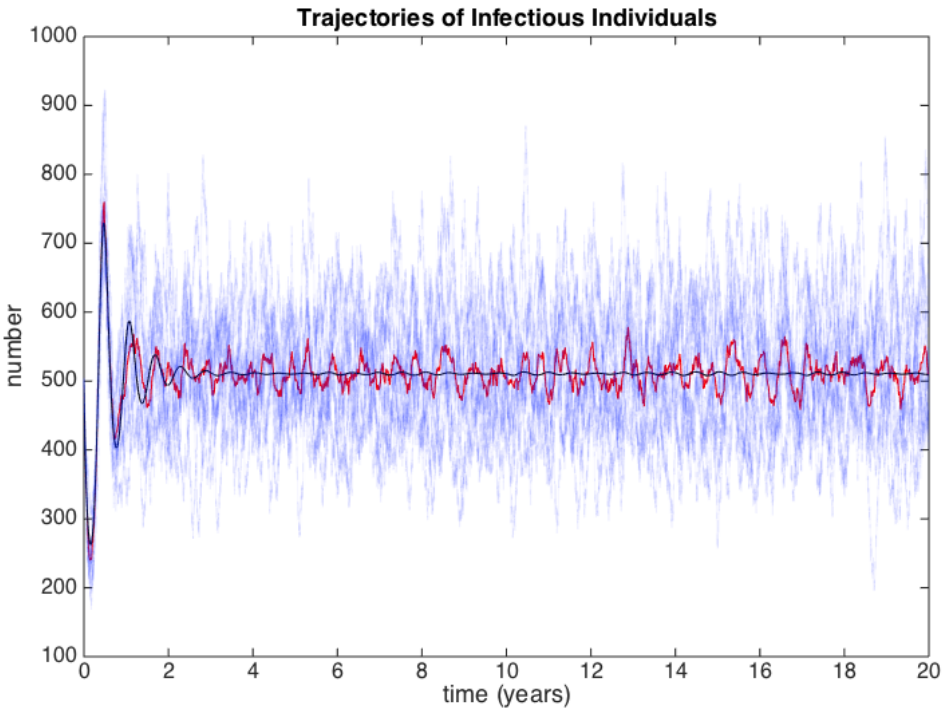


Figure 2.1: The red line is the average of ten realisations, and the solid black curve is the deterministic trajectory. The soft blue lines are each of the 20 realisations. As model is in years, the example parameters are $\beta = 584, \gamma = 36.5, \mu = 1/5$, with a population size of 100,000.

2.2.3 Example on the evaluation methods

A simple example is presented here that uses both techniques described previously. As the SIR model was defined earlier in Section 2.1.1, we will detail how to use Gillespie's algorithm, and how to construct the differential equations for the mean behaviour. Note that the population size stays constant such that $N = s + i + r$. By keeping the number of individuals in the population constant, less terms in the rates of the events must be recalculated at each time step. If N varies, then the rate corresponding to the transmission event requires three updated values (s, i, and N) but if N is held constant, only two need to be calculated at each time-step. The transitions and rates are summarised in Table 2.1.

Event	Transition	Rate
Transmission	$(s,i) \to (s-1, i+1)$	$\frac{\beta si}{N-1}$
Recovery	$(s,i) \to (s, i-1)$	γi
Birth and Death (I)	$(s,i) \to (s+1, i-1)$	μi
Birth and Death (R)	$(s,i) \to (s+1, i)$	$\mu (N-s-i)$

Table 2.1: Summary of the SIR model with demography.

One choice for the rate vector is,

$$\mathbf{r} = \left(\frac{\beta si}{N-1}, \gamma i, \mu i, \mu (N-s-i) \right)$$

and we can write the corresponding stoichiometric matrix as,

$$\Lambda = \begin{pmatrix} -1 & 0 & 1 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

where the jth column corresponds to the change in population numbers due to the jth event. For example, when a recovery event is chosen, this corresponds to the second event. The second column of Λ says that there is no change in the number of susceptible individuals, the number of infectious individuals changes by -1 and the number of recovered individuals changes by 1. With these things defined, stochastic simulation can now be carried using the steps outlined in Section 2.2.1.

Although this system is small and we can effectively simulate realisations, we construct the approximating deterministic differential equations for illustrative purposes. This method assumes that the total population, N, is limiting towards infinity, so we consider the density of the susceptible and infectious individuals.

Let $s^* = \frac{s}{N}$ and $i^* = \frac{i}{N}$ and,

$$\begin{aligned} F(s^*, i^*) &= \sum_{k=1}^{n} \eta_k w\left(\frac{1}{N}(s^*, i^*), \eta_k\right) \\ &= \beta s^* i^* (-1, 1)^T + \gamma i^* (0, -1)^T + \mu i^* (1, -1)^T + \mu(1 - s^* - i^*)(1, 0)^T \end{aligned}$$

where n is the total number of events, η_k is the kth column of Λ and $w\left(\frac{1}{N}(s^*, i^*), \eta_k\right)$ is the function such that the event k occurs at rate $Nw\left(\frac{1}{N}(s^*, i^*), \eta_k\right)$, and v^T is the transpose of some vector v. Then set $\frac{\mathrm{d}}{\mathrm{d}t}(s^*, i^*)^T = F(s^*, i^*)$. Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}(s^*, i^*)^T = \begin{pmatrix} -\beta s^* i^* + \mu(1 - s^*) \\ \beta s^* i^* - \mu i^* - \gamma i^* \end{pmatrix}.$$

Each differential equation is the flux in and out of the proportion of the respective type of individual. The solution to these differential equations is the approximate average of the stochastic realisations, see Figure 2.1. After only ten simulations, the average of the realisations follows the deterministic trajectory closely.

2.3 Linear stability analysis

In some the models we will consider, the dynamics tend to an equilibrium. To determine the stability of the equilibrium, a linear stability analysis is carried out. Consider the system of differential equations,

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = F(\boldsymbol{x}),$$

for a function $F : \mathbb{R}^m \to \mathbb{R}^m$. Let $\boldsymbol{x}_0$ be a fixed point; that is, $F(\boldsymbol{x}_0) = 0$. Then let J be the Jacobian of the system, such that,

$$J(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_m} \end{pmatrix}.$$

where $\boldsymbol{x} = (x_1, ..., x_m)$. Hence, the linearisation of the system can be expressed as,

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= F(\boldsymbol{x}_0) + J(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ &= J(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

because $\boldsymbol{x}_0$ is a fixed point. Define $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$, then,

$$\frac{\mathrm{d}\delta\boldsymbol{x}}{\mathrm{d}t} = J(\boldsymbol{x}_0)\delta\boldsymbol{x}.$$

Let $\lambda_1, .., \lambda_m$ be the eigenvalues of $J(\boldsymbol{x}_0)$. Then, by Themorem 8.12 in [25], the system is asymptotically stable if and only if $\mathrm{Re}\{\lambda_i\} < 0$ for $i = 1, ..., m$.

However, for the models in this thesis, analytically calculating the Jacobian is not always a viable course of action due to time constraints or complexity of the function F, so we make use of numerical algorithms. In this thesis, we make use of the *Adaptive Robust Numerical Differentiation* Matlab package by John D'Errico [26] which contains a numerical Jacobian function. We briefly outline the algorithm used below before further discussing the use of Jacobians in this thesis.

The function that numerically calculates the Jacobian, `jacobianest`, requires the input of the function f and the equilibrium point $\mathbf{x_0}$, and will output the estimate for $J(\mathbf{x_0})$ as well as a vector of estimated errors, each component corresponding

to the partial derivatives in the Jacobian. The function `jacobianest` numerically differentiates forward using Romberg Extrapolation [27], eventually evaluating at $\mathbf{x_0}$.

For some function $f(x)$, the Taylor series expansion about the point x_0 is given by,

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2}f''(x_0) + \frac{(x - x_0)^3}{6}f^{(3)}(x_0) + ..., \quad (2.2)$$

and hence, by rearranging the Equation (2.2), for $x = x_0 + \delta$, we can write,

$$f'(x_0) = \frac{f(x_0 + \delta) - f(x_0)}{\delta} - \frac{\delta}{2}f''(x_0) - \frac{\delta^2}{6}f^{(3)}(x_0) + ... \ . \quad (2.3)$$

For small values of δ, the terms with higher powers of δ in Equation (2.3) will become increasingly small and so we can truncate Equation (2.3) after the second derivative for small enough δ.

The idea behind the numerical method implemented in *jacobianest* is that if we take a number of approximations of Equation (2.3) with respect to different δ values, denoted $f'_\delta(x_0)$, first with some δ, then with $\frac{\delta}{2}$. A linear combination of these terms are added,

$$f'(x_0) = 2f'_{\frac{\delta}{2}}(x_0) - f'_\delta(x_0)$$

which cancels out the second derivative term in Equation (2.3). To improve the estimates, $\frac{\delta}{2}$ can be halved again, a different linear combination of these estimate terms are then utilised,

$$f'(x_0) = \frac{1}{3}f'_\delta(x_0) - 2f'_{\frac{\delta}{2}}(x_0) + \frac{8}{3}f'_{\frac{\delta}{4}}(x_0)$$

so that the second derivative terms cancel out. This process of halving δ can be

continued until the approximation is within some desired error tolerance.

An eigenvector of a square matrix is a non-zero vector that does not change 'direction' upon being multiplied by the matrix, but is instead a scalar multiple of the vector. That is, if $\mathbf{v}$ is an eigenvalue of the square matrix A, then,

$$A\mathbf{v} = \lambda\mathbf{v},$$

where λ is called the eigenvalue of matrix A corresponding to the right eigenvector $\mathbf{v}$. To find all of the eigenvalues of a square matrix, A, solve,

$$\det(A - \lambda \mathrm{I}) = 0, \tag{2.4}$$

where I is the identity matrix of the same size as A. Equation (2.4) is known as the characteristic equation of A. The matrix A will have the same number of eigenvalues as the number of rows (or columns, as they are equal) because the characteristic equation is a polynomial of degree equal to the number of rows in the matrix and, by the Fundamental Theorem of Algebra, will have as many solutions to Equation (2.4). That is, if A is a $m \times m$ matrix, A has m eigenvalues.

Let $\lambda_k = a_k + ib_k$, where $a_k, b_k \in \mathbb{R}$ for $k = 1, ..., m$, be the kth eigenvalue of $J(\mathbf{x_0})$, sorted such that $a_1 < \ldots < a_m$. If all the eigenvalues have real parts that are negative, then the process is stable. If just one eigenvalue has real part positive, then the system is unstable. Further, by considering the imaginary parts of the eigenvalues, we can determine the periodicity of the oscillations of the system [28]. The periodicity of the oscillations of the system with respect to the eigenvalue λ_k is $\frac{2\pi}{b_k}$, which is used to find the periodicity of peaks of infection in some of the models presented in this thesis.

Chapter 3

Demography model

In this chapter, we discuss how household models, introduced in Chapter 1, can be extended to include demography. That is, household structure is no longer assumed to stay static but can instead change over time. Disease dynamics are not included at this point as this model focuses on how the household types change over time.

3.1 Set-up

We model a population of households as a CTMC. We split a population of M households into 11 different household types. These household types are denoted by (i, j), where i is the number of adults and j is the number of children in a household. We define children as individuals that are dependent on adults (and not necessarily by their age). Considering Australian census data in previous years [29, 30], it becomes apparent that group households[1] make up less than 5% of the population. Hence we restrict the number of adults in a household to be at most 2. In order to keep this model practically feasible in terms of

[1]The Australian Bureau of Statistics defines group households as households that are made up of at least two individuals over the age of 15 that are not related and are not in a relationship of a romantic or parent-child kind.

computation, the maximum number of children in a household is restricted to be 4. This assumption is realistic because very few households have more than four children [29]. Hence, we enumerate the set of all household types as follows:

$$\mathcal{T} = \{(0,0),(1,0),(1,1),(1,2),(1,3),(1,4),(2,0),(2,1),(2,2),(2,3),(2,4)\}. \tag{3.1}$$

The empty household is included as one of the household types because it is then easy to increase the total number of households in the population by essentially building new empty households that the individuals in the population can move into.

Define $m_{(i,j)}$ to be the number of households in the population of type (i,j). The state space of the process is,

$$\mathcal{S}_{demo} = \Big\{\mathbf{m} = \big(m_{(0,0)}, m_{(1,0)}, ..., m_{(2,4)}\big) : \sum_{(i,j)\in\mathcal{T}} m_{(i,j)} = M, m_{(i,j)} \geq 0 \ \forall \ (i,j) \in \mathcal{T}\Big\}. \tag{3.2}$$

In order to complete the definition of the CTMC, we must define the events that can occur and how they alter the state of the system. The models defined in Glass *et al* [19] and in Geard *et al* [20] detail population events such as births, deaths and movement between households. The latter of these two models defines couple formation and separation events. We use the events defined in these as motivation for our own model; that is, we define birth, death, children moving out of home, couple formation, separation, migration, and a new empty household events. Below are descriptions of these demographic events and the rates at which they happen. The events are summarised by transitions and rates in Table 3.1 and a brief description of the parameters used in this table are summarised in Table 3.2.

i. Births

In a household of type $(2, j)$, where $j \leq 3$, the household can undergo a birth event at rate b_j, and the result is an additional $(2, j+1)$ household and the loss of a type $(2, j)$ household. In this model, the birth rate in a household depends on the number of children in a household because many couples choose not to have any children. However, once a family has one child they are more inclined to have a second. Indeed, of the households with children, the $(2, 2)$ household is the most common in Australia [29]. Note that households of type $(2, 4)$ cannot have a birth event.

ii. Deaths

In a household of type (i, j) with both $i, j \geq 1$, a child dies with rate jd_C, which means that a household of type (i, j) is lost and a household of type $(i, j-1)$ is gained. There are two possible scenarios for an adult death event. If $i = 2$, then an adult death event occurs at a rate $2d_A$ and the new household has type $(1, j)$. However, if $i = 1$ then the rate of an adult dying is d_A and the new household is of type $(0, 0)$. It is assumed children cannot live without an adult (as they are dependent on them) and hence when there are no adults in a household, it is simply removed from the population like an emigration event (still to be defined in vi).

iii. Child moving out

When a child leaves home, there are two possible avenues the child can take. If there are available empty households—that is $m_{(0,0)} > 0$—then a child may move into one, and hence become an adult. Alternatively, the child can leave the population. This second possibility can be explained by, for example, a child finishing their degree and pursuing a career in another city. A child in a

household of type (i, j), with $i, j \geq 1$, will move out of home with rate jm_{out}, where $\frac{1}{m_{out}}$ is the average age of a child when they move out of home. Then there is a probability of a child moving into an empty household or emigrating out of the system. The precise functional form of this probability will be discussed in more detail in Section 3.2, but the general objective of the function is to have a higher probability of emigrating when there are less empty households in the population. If the child moves into an empty household, the initial households (i, j) and $(0, 0)$ become $(i, j-1)$ and $(1, 0)$, respectively. On the other hand, if the child leaves the system, the initial household of type (i, j) becomes $(i, j-1)$; that is, no empty households are involved. Note also that in the second possible event, transitionally, it appears as if we have just removed a child from the household–in other words, this has the same affect as a child death event.

iv. Couple separation

A household of type $(2, j)$ undergoes couple separation at a rate of c_s. Similarly to the child moving out of home event, there are two possible outcomes to this couple dissolution event. One of the adults in the household will leave the household, taking no children with them, and will hence have to decide whether to move into an empty household or to leave the population. As before, the less empty households there are in the population, the more likely the adult will choose to leave the system. In the event that the adult doesn't leave the population, the households $(2, j)$ and $(0, 0)$ turn into $(1, j)$ and $(1, 0)$, respectively. Alternatively, in the case where the adult leaves the population, household $(2, j)$ becomes $(1, j)$. Again, in the second case described, it appears as if, transitionally, one of the adults has just been removed from the household—or in other words, an adult death event.

v. Couple formation

When $m_{(1,0)} \geq 1$, a household of type $(1, j)$ will combine with a $(1, 0)$ household at a rate $c_f m_{(1,0)}$ to form a new household of type $(2, j)$ and also leaves behind an empty household. It is assumed that at least one of the households has to be of type $(1, 0)$ to simplify the problem arising from the possibility of two households of, say, type $(1, 3)$ wanting to form. Our model limits the number of children to be at most four, but this formation would produce six children and further complicate the model if we did not make the aforementioned assumption. If there are no households of type $(1, 0)$, then the rate of couple formation is zero and the event cannot occur.

vi. Emigration and immigration

Lastly, we define the migration events. Let $p_{(i,j)}$ be the probability of a household type (i, j) leaving/entering the population given an emigration/immigration event has occurred. A household of type (i, j), for all j and $i \neq 0$, can leave the system with rate $p_{(i,j)}\phi$ and hence leave behind an empty household. We call this an emigration event. An empty household can become inhabited by an incoming family of type (i, j) at rate $p_{(i,j)}\xi$. However, just as with the couple separation and children moving out of home events, families choose to enter the population more if there are a lot of empty households. We call this an immigration event.

vii. New empty household

In most cities across the world, populations are increasing; this produces a demand for new houses to be built. By building new empty households we can simply relax the assumption of a constant number of total households. An alternative way to formulate this would be to not count the number of empty

households, but in doing so the rate of an immigration event would become difficult to define. Additionally, the rates of couple separation and children moving out of home both rely on the number of empty households. To remove the number of empty households from our state would further complicate the model, so we define the event where a new empty household is produced by adding a single $(0, 0)$ household to the population, at a rate gM, where M is the total number of households in the population initially. For $g > 0$, the number of households in the population is not held constant.

Event	**Transition**	**Rate**	**Exclusions**
Birth	$(2,j) \to (2,j+1)$	$b_j m_{(2,j)}$	$j \neq 4$
Adult Death (2)	$(2,j) \to (1,j)$	$2d_A m_{(2,j)}$	
Adult Death (1)	$(1,j) \to (0,0)$	$d_A m_{(1,j)}$	
Death of Child	$(i,j) \to (i,j-1)$	$jd_C m_{(i,j)}$	
Child Moving Out	$(i,j),\ (0,0) \to (i,j-1),(1,0)$	$jm_{out} m_{(i,j)} f\left(\frac{m_{(0,0)}}{M}\right)$	$j \neq 0$
Child Moving Out (emigrate)	$(i,j) \to (i,j-1)$	$jm_{out} m_{(i,j)} \left(1 - f\left(\frac{m_{(0,0)}}{M}\right)\right)$	$j \neq 0$
Couple Formation	$(1,j),\ (1,0) \to (2,j),\ (0,0)$	$\frac{c_f}{M} m_{(1,j)} m_{(1,0)}$	
Couple Separation	$(2,j),\ (0,0) \to (1,j),\ (1,0)$	$c_s m_{(2,j)} f\left(\frac{m_{(0,0)}}{M}\right)$	
Couple Separation (emigrate)	$(2,j),\ (0,0) \to (1,j),\ (1,0)$	$c_s m_{(2,j)} \left(1 - f\left(\frac{m_{(0,0)}}{M}\right)\right)$	
Emigration	$(i,j) \to (0,0)$	$p_{(i,j)} \phi m_{(i,j)}$	
Immigration	$(0,0) \to (i,j)$	$p_{(i,j)} \xi f\left(\frac{m_{(0,0)}}{M}\right) m_{(0,0)}$	
New Empty Household	$\varnothing \to (0,0)$	gM	

Table 3.1: Summary of the demographic events.

Parameter	Quick Description
b_j	birth rate in a household with j children
d_C	death rate of a child
d_A	death rate of an adult
m_{out}	rate of a child moving out
c_f	rate of couple formation parameter
c_s	rate of couple separation parameter
ϕ	rate of emigration parameter
ξ	rate of immigration parameter
$p_{(i,j)}$	probability of emigration/immigration out of/into household type (i, j)

Table 3.2: A summary of the parameters and a brief description of the demography event to which it is associated.

3.2 Preference function

Recall the child moving out and couple separation and immigration events, and the as yet undefined functional form of the probability of leaving the system. Let $f : [0, 1] \to [0, 1]$ be a function that depends on the proportion of empty households, $\frac{m_{(0,0)}}{M}$. Let the event L be when an individual leaves their household through either a child moving out or a couple separation event, and let the event E be when that individual chooses to stay in the population and hence move into an available empty household. We have an interpretation of this function f such that,

$$f\left(\frac{m_{(0,0)}}{M}\right) = \mathbb{P}\left(E \mid L, m_{(0,0)}\right)$$

This leads to the requirement that $f(0) = 0$, which says that if there are no empty households available, there is 0 probability that a home-leaver will move into an

empty household. Further, it should be noted that this function does not imply a choice but rather reflects a variety of factors in whether an individual leaves the population or stays. Small numbers of empty households means it is more difficult to find an available home to move into. As a consequence, there will also be more competition in order to secure a house as all the home-leavers are competing against one another for very few available empty households. From a macroscopic viewpoint, the less empty households there are in the population, the more likely one would see a home-leaver (whether that be a child moving out or an adult leaving their partner or family looking for a new city to live in) leave the population entirely as securing a new home to live in the same city becomes more difficult than if there were a large number of empty households.

One example of such a function is,

$$f\left(\frac{m_{(0,0)}}{M}\right) = C\left(1 - \left(1 - \frac{m_{(0,0)}}{M}\right)^k\right). \tag{3.3}$$

where C is the maximum probability of an individual moving into an empty household. That is $f(0) = 0$ and $f(1) = C$. We set $C = 0.95$, which has the intuitive interpretation that in a population with sufficient empty households such that an individual could move into an empty household if he/she so desired, they would move into an empty household 95% of the time.

There is still the question of what power, k, is utilised in Equation (3.3). Four possible values are displayed in Figure 3.1. For each value of k, the shape is similar, except for the steepness of the slope for small proportions of empty households. Census data suggests that empty households make up approximately 10–15% of all households [29] so for a "typical" population, low proportions of empty households should still produce moderate to high probabilities of choosing to stay in the population. We arbitrarily choose $k = 9$, and set $C = 0.95$.

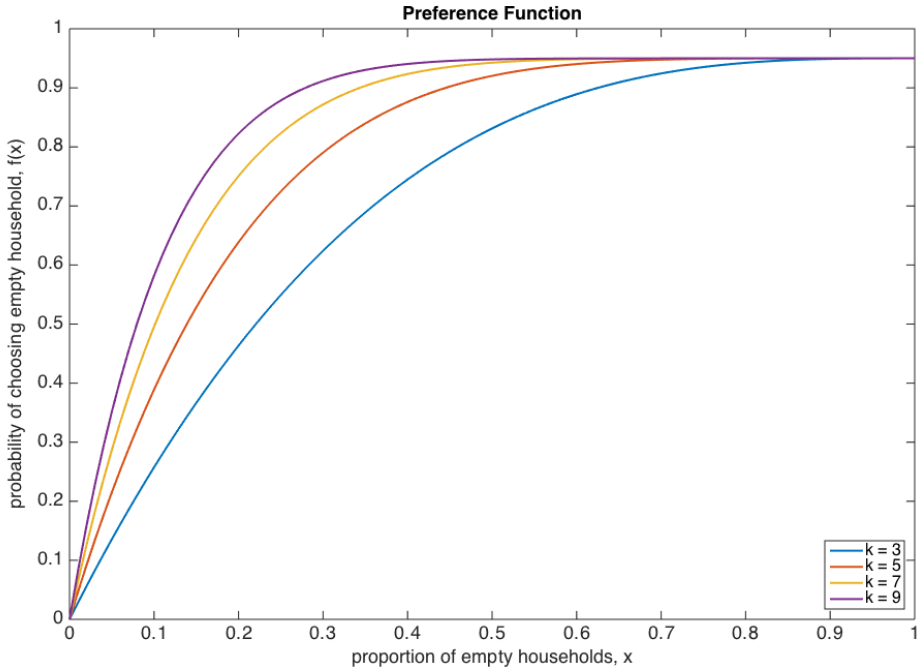


Figure 3.1: Graph of Equation (3.3) for k values specified in the legend and $C = 0.95$.

3.3 Methods for evaluating the population dynamics

In this section, we discuss the theory and the methods that we use to analyse our stochastic model. The two methods include a stochastic simulation method and a deterministic approximation for the expected behaviour of our stochastic model.

3.3.1 Stochastic simulation

To simulate the demographic model, refer to Section 2.2.1 for the general methodology. However, in order to properly explain the procedure, the generic variables used should be defined. The state space of the system we are simulating is $\mathcal{S}_{demo}$ (recall Equation (3.2)), where the initial condition is taken to be the distribution

of households as seen in 2001 census data [29]. More detail on how this data is utilised is discussed in Section 3.4

In total, there are 60 unique events. Let $\mathbf{r}$ be a vector of length 60 with components corresponding to the rate of each of these events. Then Λ is the stoichiometric matrix of size 11×60, where column c is the net change in the number of households of each type corresponding to the event with rate in the cth component of $\mathbf{r}$. With this information, the system can be simulated using the so-called Gillespie algorithm.

We plot a realisation against the deterministic trajectory (which will be defined in Section 3.3.2), and discuss the relationship between the two methods in detail in Section 3.3.3.

3.3.2 Deterministic approximation

Stochastic simulation can be computationally expensive, which will be discussed briefly in Section 3.3.3 and so theory proved by Kurtz [23] becomes pivotal to evaluate the model efficiently. The result allows us to approximate the expected behaviour of a density dependent CTMC deterministically, which can improve the speed with which we can analyse the model. In order to use the theory proved by Kurtz, we require our model to be density dependent, as defined in Section 2.2.2. In addition to the density dependent rates, the theory looks at the limit as the number of households approaches infinity.

Recall Equation (3.2), where $\mathbf{m}$ is the vector with the number of each household type in the population, and let $\mathbf{z}$ be the transition vector that changes $\mathbf{m}$ due to the event type k. To show the model is density dependent, all of the rates for the event types k, as described in Section 2.2.2, must be able to be put in the form,

$$q^{(k)}_{\mathbf{m},\mathbf{m}+\mathbf{z}} = Mw\left(\frac{1}{M}\mathbf{m}, \mathbf{z}\right),$$

where M is the total number of households in the population and $w(\boldsymbol{x}, \mathbf{z})$ is a continuous function such that $\boldsymbol{x} = \frac{1}{M}\boldsymbol{m} \in \mathbb{R}^{11}$ and $\mathbf{z} \in \mathbb{Z}^{11}$.
In order to satisfy the above condition, w must be defined for each event type. Once it is known that the CTMC is density dependent, the results detailed by Kurtz in [23] can be applied. Denote,

$$\boldsymbol{x} = (x_{(0,0)}, x_{(1,0)}, \ldots, x_{(2,4)}).$$

Then for a household of type (i, j) and the preference function f, set,

$$w = \begin{cases} b_j x_{(i,j)} & \text{births, where } b_j = 0 \text{ if } i \neq 2 \\ d_C x_{(i,j)} & \text{child death} \\ d_A x_{(i,j)} & \text{adult death} \\ \phi x_{(i,j)} & \text{emigration} \\ \xi x_{(0,0)} f(x_{(0,0)}) & \text{immigration} \\ m_{out} f\left(x_{(0,0)}\right) x_{(i,j)} & \text{moving out} \\ m_{out}\left(1 - f\left(x_{(0,0)}\right)\right) x_{(i,j)} & \text{moving out (emigrate)} \\ c_s f\left(x_{(0,0)}\right) x_{(i,j)} & \text{couple separation, where } c_s = 0 \text{ if } i \neq 2 \\ c_s\left(1 - f\left(x_{(0,0)}\right)\right) x_{(i,j)} & \text{couple separation, where } c_s = 0 \text{ if } i \neq 2 \text{ (emigrate)} \\ c_f x_{(1,0)} x_{(i,j)} & \text{couple formation, where } c_f = 0 \text{ if } i \neq 1 \\ g & \text{new empty household.} \end{cases}$$

Hence the CTMC is density dependent and so a system of differential equations, as described in [23], can be constructed. These differential equations, in the most intuitive sense, measure the rate at which each event type changes the proportion of each household type. These rates have different functional forms, so we split

the events up into subsets. Each subset has events with similar forms of rates which allows us to construct our differential equations in a concise manner.

We solve the system of differential equations numerically because finding an analytic solution is infeasible for this model. In order to evaluate the system efficiently, we need a compact representation of the differential equations. That is, we construct the system of differential equations as a product of matrices and vectors, where the matrices depend on the subsets of events just introduced. We will define the subsets shortly.

If all the rates were linear, the system of differential equations would be a product of a matrix and a vector. That is, for some matrix A,

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}.$$

However, because the rates are not all linear, we must use a different form to specify the system of differential equations.

Let $\mathcal{E}$ be the set of all event types; that is, births, adult deaths, child deaths, children moving out of home, couple formation and separation, immigration and emigration. Of these types of events, births, both adult and child deaths, and emigration have linear rates, and so these events make up the set we denote $\mathcal{E}_{lin}$. Let $\mathcal{E}_{dec}$ be the set of events consisting of moving out of home and couple separation, as these events both have the same formulation for the rates where there is some probability that an individual stays or leaves the population. Let $\mathcal{E}_{im}$ be the set containing the immigration event, and $\mathcal{E}_{cf}$ be the set containing couple formation. Hence we have,

$$\mathcal{E} = \mathcal{E}_{lin} \cup \mathcal{E}_{dec} \cup \mathcal{E}_{im} \cup \mathcal{E}_{cf}.$$

For each of these subsets, a matrix will be constructed to formulate the system

of differential equations described by Kurtz. The system of differential equations can be written as,

$$\frac{d\boldsymbol{x}}{dt} = \left(F_{lin} + f\left(x_{(0,0)}\right) F_1 + \left(1 - f\left(x_{(0,0)}\right)\right) F_2 + x_{(1,0)} F_3\right) \boldsymbol{x} + G,$$

where $G = (g, 0, ..., 0)^T$ which increases the proportion of empty households at a constant rate g, and F_{lin}, F_1, F_2, and F_3 are defined in Appendix A.

3.3.3 Numerical verification of evaluation methods

The system of differential equations just derived is quite obviously non-linear and hence is more complicated to solve than standard linear differential equations. We content ourselves with numerical solutions using the Matlab function `ode15s`.

The theory proved by Kurtz [23], as introduced in Section 2.2.2, outlines how a system of differential equations can approximate the expected behaviour of the stochastic model as the population size gets larger. At the beginning of Section 3.3.2, it was shown that the CTMC satisfies the conditions of Kurtz's theorem and hence a system of differential equations were constructed.

Figure 3.2 displays the deterministic trajectories of each of the eleven household types with a realisation of the stochastic model on the same axes. It is noted that the stochastic trajectories follow roughly the same path as the deterministic as expected.

To further show that the two models are consistent, 50 simulations were run and averaged for an initial population of 5.55×10^5 households (1.13×10^6 individuals) and hence the average trajectories for each of the household types are plotted on the same axes as the deterministic solution, as seen in Figure 3.3. The difference between the deterministic and stochastic trajectories are quite small. After only 50 realisations, we see the average of the simulations much more closely

matched compared to those in Figure 3.2. This shows satisfactory validation of our approximation using Kurtz's results.

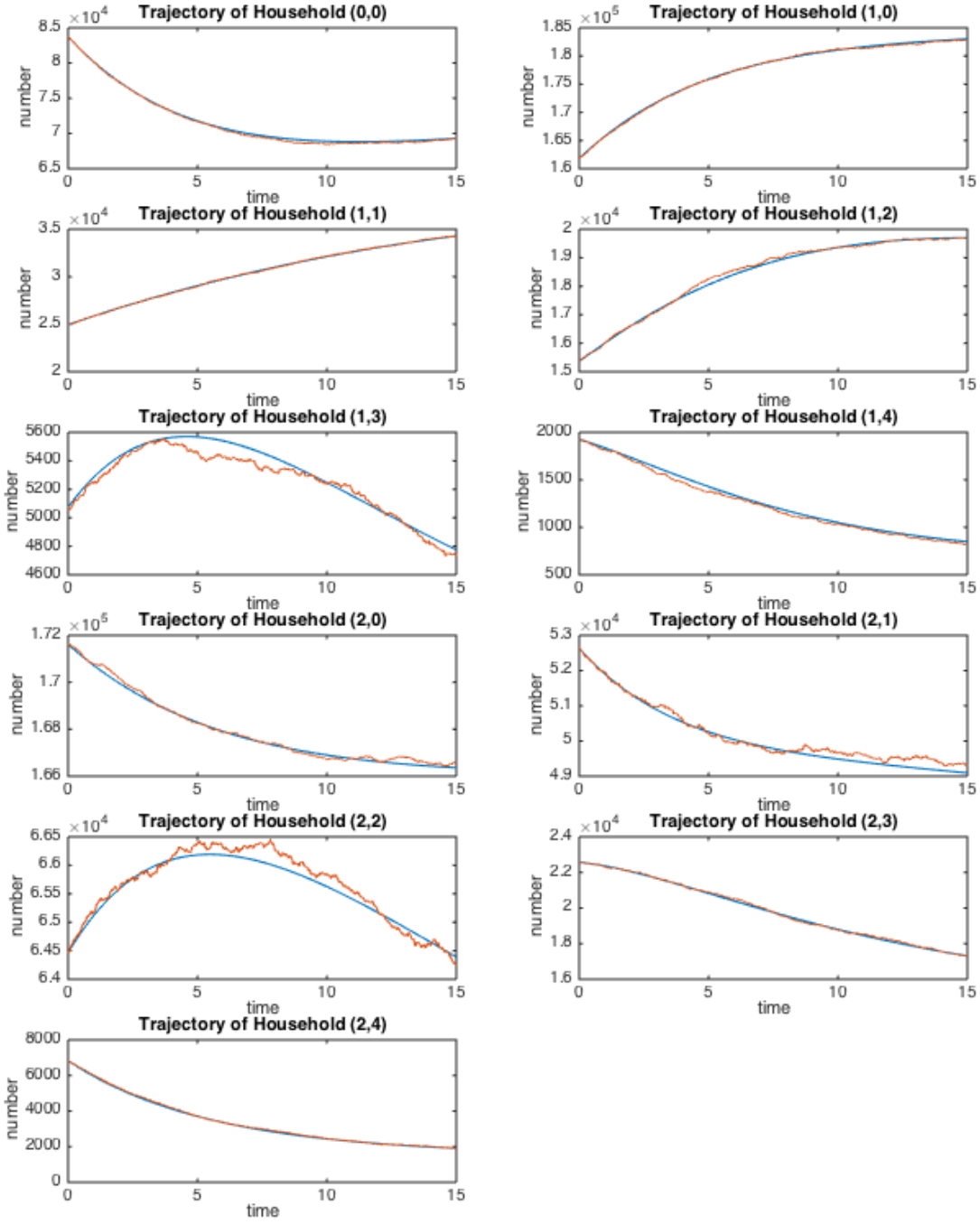


Figure 3.2: The trajectories of a single realisation and the deterministic solution of the number of households of each type, over fifteen years, on the same axes, with an initial population of 5.55×10^5 households (1.13×10^6 individuals). Parameters as in Table 3.3.

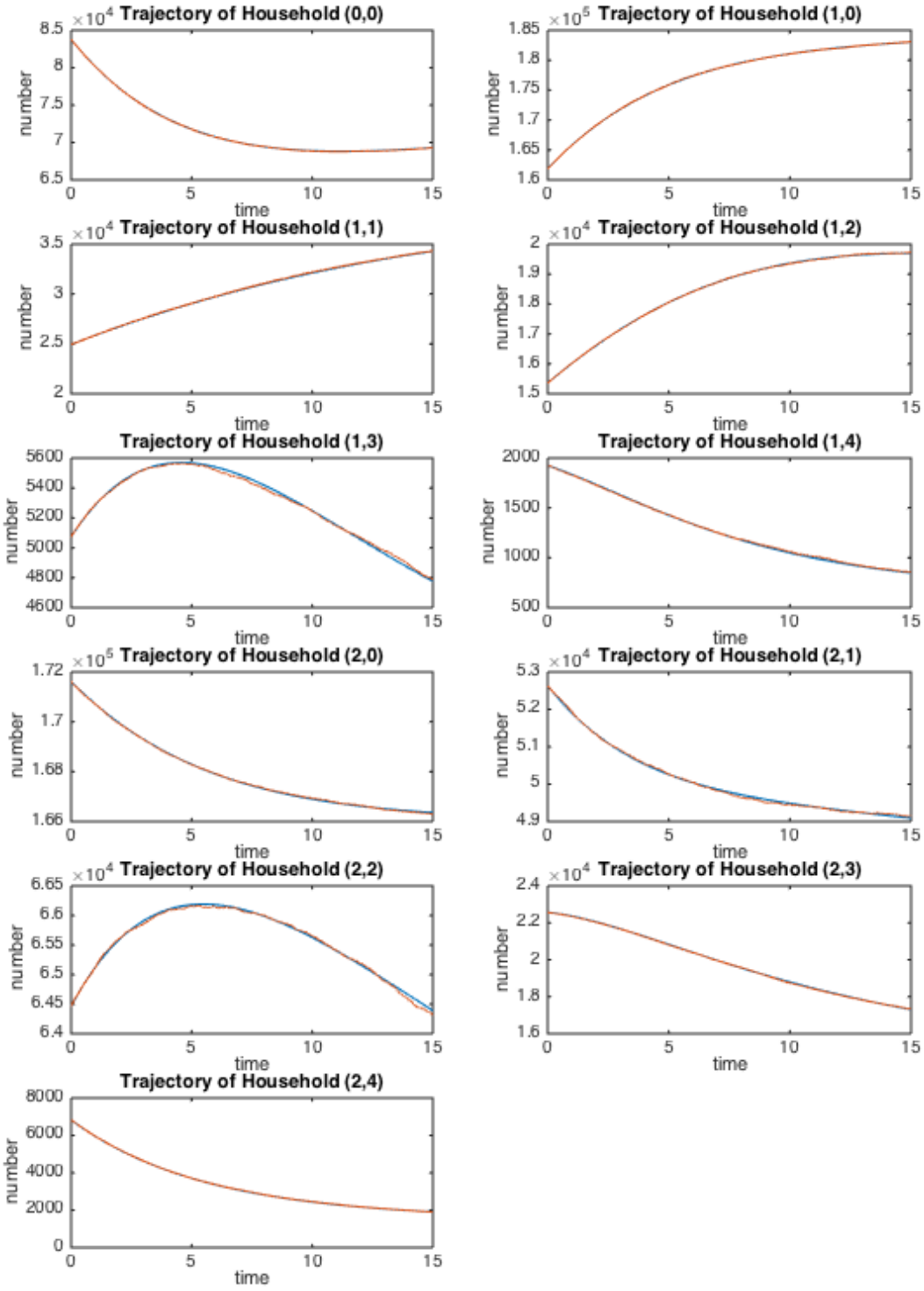


Figure 3.3: The trajectories of the average of 50 realisations and the deterministic solution of the same axes, on an initial population of 5.55×10^5 households (1.13×10^6 individuals). Parameters as in Table 3.3.

3.4 Parameterisation

In this section, the model is parameterised using household composition data from the years 2001, 2006 and 2011 [29], shown in Figure 3.4. We present two different parameterisations using this data that corresponds to a population with a fixed number of households and also a population with a growing number of households.

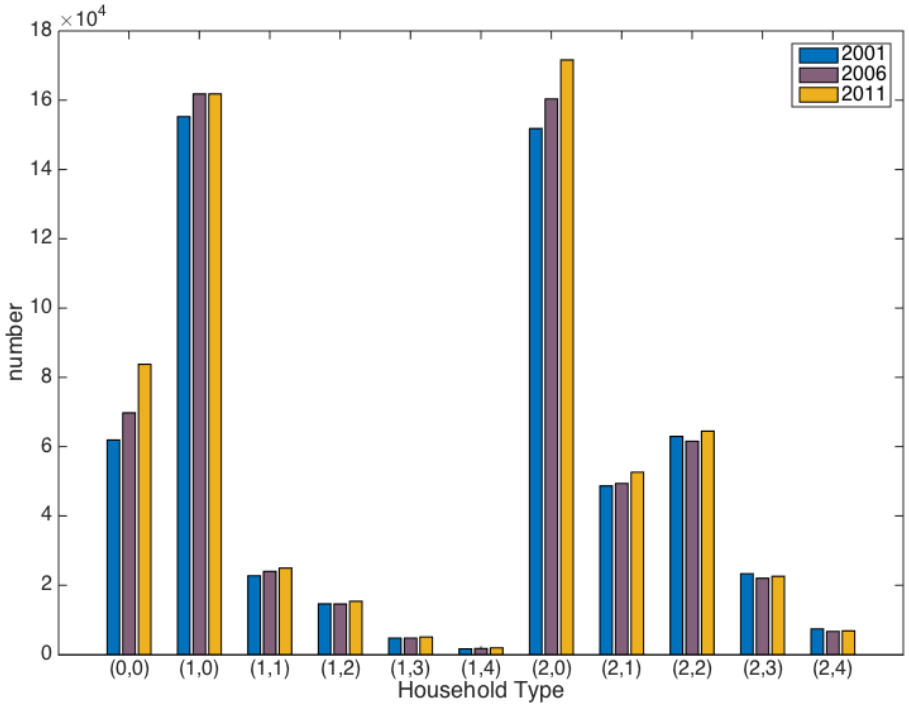


Figure 3.4: The number of each household type in South Australia, from the 2001, 2006, 2011 census data.

We use a least squares method to parametrise the demographic model using the deterministic approximation. This is because we only require a fit that allows realistic movement between households, in addition to keeping the proportions of each of the households resembling the census data in Figure 3.4. When the disease dynamics are introduced, these features will be more important rather than a robust parameterisation as the demography model is on a much slower time scale compared to the disease dynamics.

Data concerning the structure of households in cities around Australia are available and are utilised. Recall that a group household is a household containing two or more unrelated individuals over the age of 15 that are not in a relationship of a romantic or parent-child kind. As the proportion of group households is so small in comparison to the other household types, we neglect them and only consider households that are represented by our model. We parameterise our model according to South Australian census data [29]. Note that in the data, there are a total of 6.1×10^5 households (1.2×10^6 individuals) in the year 2011, which is a large enough population size such that the deterministic approximation to the household model is very good, as seen in Section 3.3.3.

To use the data, we need to group the data into the household types detailed earlier in Chapter 3. We let $\mathbf{v}^{(2001)}$, $\mathbf{v}^{(2006)}$ and $\mathbf{v}^{(2011)}$ be the vectors containing the number of households of each household type. Recall the set $\mathcal{T}$ (Equation (3.1)), then $v_{(i,j)}^{(2001)}$ is the number of the households of type (i, j). Then $\mathbf{v}^{(2001)} = \left(v_{(0,0)}^{(2001)}, v_{(1,0)}^{(2001)}, ..., v_{(2,4)}^{(2001)}\right)$, and similarly for $\mathbf{v}^{(2006)}$ and $\mathbf{v}^{(2011)}$. Further, let,

$$\mathbf{v}_p^{(2001)} = \frac{1}{\sum_{(i,j)\in\mathcal{T}} v_{(i,j)}^{(2001)}} \mathbf{v}^{(2001)},$$

and similarly for $\mathbf{v}_p^{(2006)}$ and $\mathbf{v}_p^{(2011)}$; these vectors represent the proportion of each household.

There are two sets of parameters that are considered for the demography dynamics throughout this thesis: a household model with a fixed number of households and a model where the number of households is growing. The model with the fixed number of households is used to compare with existing models with fixed population sizes and a stable population structure, and this is done by taking an average of the census data over the ten years as the composition of the population is not changing significantly. However, the model with the growing number of households can be used to investigate a constantly changing population structure,

and so we fit the parameters to the number of households in the population over the ten years, utilising the fact that the population is indeed growing. The model is parameterised (to determine the parameters used in Table 3.2) by solving the following optimisation problems.

3.4.1 Parameterising the stable model

Let $\boldsymbol{x}$ be a vector of length 21, such that, $\boldsymbol{x} = (\mathbf{b}, d_C, d_A, m_{out}, c_f, c_s, \phi, \xi, \mathbf{p})$, where $\mathbf{b} = (b_0, ..., b_3)$ and $\mathbf{p} = \left(p_{(1,0)}, p_{(1,1)}, ..., p_{(2,4)}\right)$. Let $\mathbf{m}^*$ be the number of households of each type in equilibrium. Then define the optimisation problem by the following,

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f_s(\boldsymbol{x}) = \left| \mathbf{m}^* - \frac{1}{2}\left(\mathbf{v}_p^{(2006)} + \mathbf{v}_p^{(2011)}\right) \right|^2 \\
\text{subject to} \quad & \sum_{(i,j)\in\mathcal{T}} p_{(i,j)} = 1, \\
& \left(10^{-2}, 10^{-2}, 10^{-2}, 10^{-2}\right) \leq \mathbf{b} \leq (0.5, 0.5, 0.5, 0.5), \\
& \frac{1}{90} \leq d_C \leq \frac{1}{50}, \\
& \frac{1}{75} \leq d_A \leq \frac{1}{40}, \\
& \frac{1}{30} \leq m_{out} \leq \frac{1}{18}, \\
& 10^{-3} \leq c_s \leq 0.05, \\
& 10^{-3} \leq c_f \leq 0.085, \\
& 10^{-3} \leq \phi \leq 0.04, \\
& 10^{-3} \leq \xi \leq 0.06, \\
& L \leq \mathbf{p} \leq U,
\end{aligned}
$$

where

$$L = (0.2, 0.04, 0.04, 0.04, 0.04, 0.2, 0.05, 0.05, 0.05, 0.05) \text{ and,}$$

$$U = (0.5, 0.3, 0.2, 0.2, 0.2, 0.5, 0.2, 0.15, 0.15, 0.1).$$

Using the `fmincon` Matlab function, the parameters that minimise $f_s(\boldsymbol{x})$ are summarised in Table 3.3. These parameters are then utilised in solving the differential equations established in Section 3.3.2, where the proportion of each household can be compared to the census data in Figure 3.5. Additionally, the figure also shows that in the long term the distribution of the household types is very similar to that of the current population structure, albeit with a slight increase in $(1, 0)$ households. It should be noted that the model was also fitted assuming $b_0 = b_1 = b_2 = b_3$, to test whether the extra flexibility of additional parameters is necessary, but this method did not produce as close a fit as the method where the birth rate depends on household type.

Parameter	**Value**
b	$(0.0158, 0.1263, 0.0292, 0.0100)$
d_C	0.0111
d_A	0.0133
m_{out}	0.0333
c_f	0.0850
c_s	10^{-3}
ϕ	0.0057
ξ	0.060
p	$(0.20, 0.04, 0.04, 0.04, 0.04,$ $0.20, 0.19, 0.05, 0.15, 0.05)$

Table 3.3: A summary of the parameters that minimise the function $f_s(\boldsymbol{x})$.

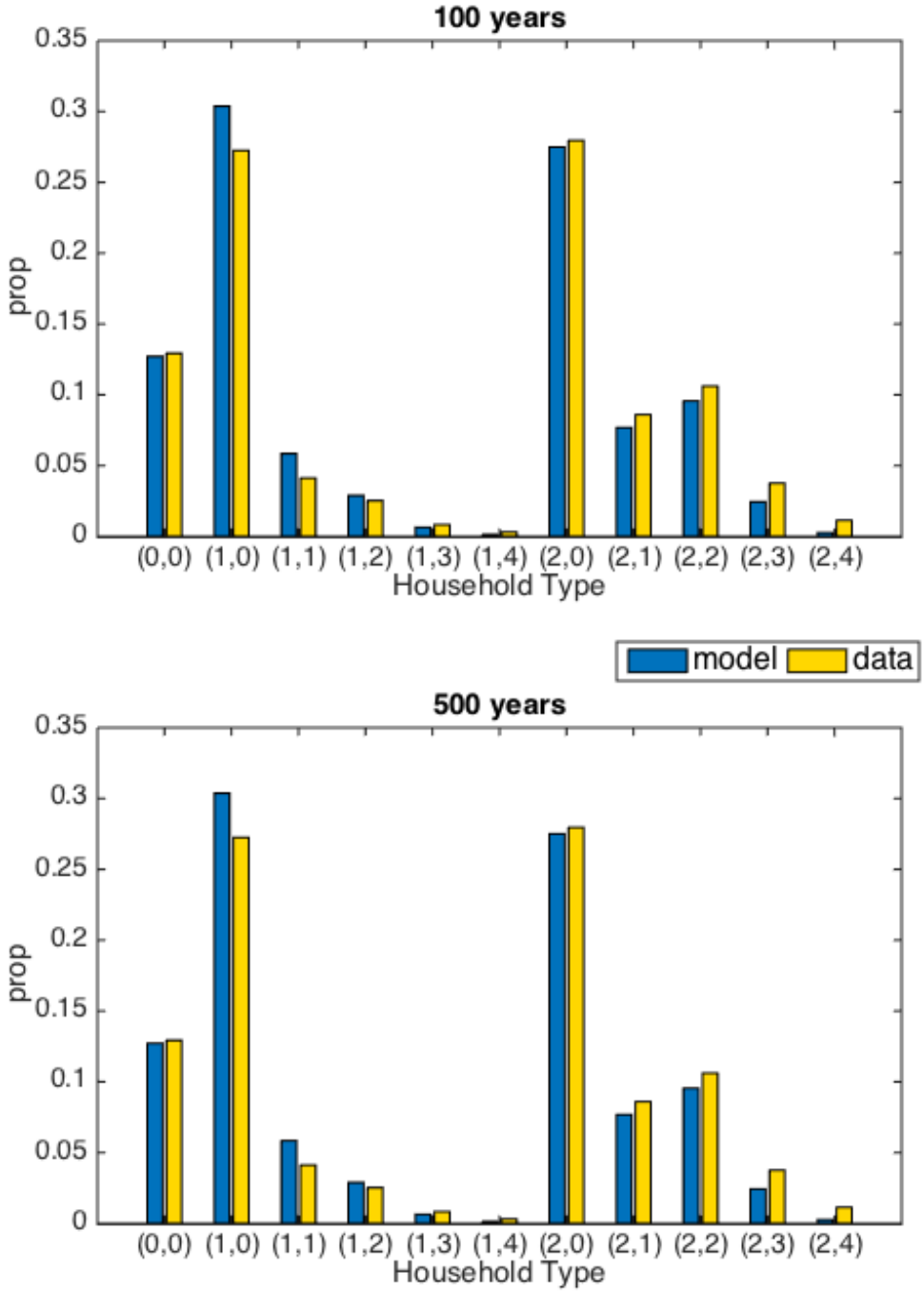


Figure 3.5: Two graphs that demonstrate the fit of the parameters, in Table 3.3, by forecasting the proportion of each household type over 100 years and 500 years, and comparing against the average of the real data from 2006 and 2011. As this parameter set is for a stable population, the assumption that the population composition is not changing is seen to be valid as the proportions of each household type are not changing significantly.

3.4.2 Parameterising the growing model

Let $\boldsymbol{x}$ be a vector of length 22, such that, $\boldsymbol{x} = (\mathbf{b}, d_C, d_A, m_{out}, c_f, c_s, \phi, \xi, g, \mathbf{p})$, where $\mathbf{b} = (b_0, ..., b_3)$, $\mathbf{p} = \left(p_{(1,0)}, p_{(1,1)}, ..., p_{(2,4)}\right)$, and g is the rate at which a new empty household is added, per household in the population. Let $\mathbf{m}(t)$ be the number of households of each type after t years. Then define the optimisation problem by the following,

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f_g(\boldsymbol{x}) = \left|\mathbf{m}(5) - \mathbf{v}^{(2006)}\right|^2 + \left|\mathbf{m}(10) - \mathbf{v}^{(2011)}\right|^2 \\
\text{subject to} \quad & \sum_{(i,j)\in\mathcal{T}} p_{(i,j)} = 1, \\
& \left(10^{-2}, 10^{-2}, 10^{-2}, 10^{-2}\right) \leq \mathbf{b} \leq (0.5, 0.5, 0.5, 0.5)\,, \\
& \frac{1}{100} \leq d_C \leq \frac{1}{60}, \\
& \frac{1}{100} \leq d_A \leq \frac{1}{50}, \\
& \frac{1}{35} \leq m_{out} \leq \frac{1}{18}, \\
& 10^{-3} \leq c_s \leq 0.05, \\
& 10^{-3} \leq c_f \leq 0.085, \\
& 10^{-3} \leq \phi \leq 0.05, \\
& 10^{-3} \leq \xi \leq 0.05, \\
& 0 \leq g \leq 1, \\
& L \leq \mathbf{p} \leq U,
\end{aligned}
$$

where

$$L = (0.2, 0.04, 0.04, 0.04, 0.04, 0.2, 0.05, 0.05, 0.05, 0.05) \text{ and,}$$

$$U = (0.5, 0.3, 0.2, 0.2, 0.2, 0.5, 0.2, 0.15, 0.15, 0.1).$$

Using the `fmincon` Matlab function, the parameters that minimise $f_g(\boldsymbol{x})$ are summarised in Table 3.4. These parameters are utilised and compared to the census data in Figure 3.6. We consider endemic diseases, and hence we look at the long-term population structure. Figure 3.6 also shows the number of each of the households that the model produces in 100 years times, and how the proportion of each household compares to the current data. This parameter set means that 4095.5 households are added to the population each year, meaning in just over 100 years the population will approximately double. Despite the increase of additional households, the proportion of each of the households does not change significantly. In fact, if anything it shows a decrease in families with children and increase in families without, which is the current trend, which can be seen in the Figure 3.4.

Parameter	**Value**
$\mathbf{b}$	$(0.0100, 0.0996, 0.0399, 0.0320)$
d_C	0.0144
d_A	0.0100
m_{out}	0.0291
c_f	0.085
c_s	10^{-3}
ϕ	10^{-3}
ξ	0.05
g	0.0067
$\mathbf{p}$	$(0.20, 0.04, 0.04, 0.04, 0.04,$ $0.20, 0.19, 0.05, 0.15, 0.05)$

Table 3.4: A summary of the parameters that minimise the function $f_g(\boldsymbol{x})$.

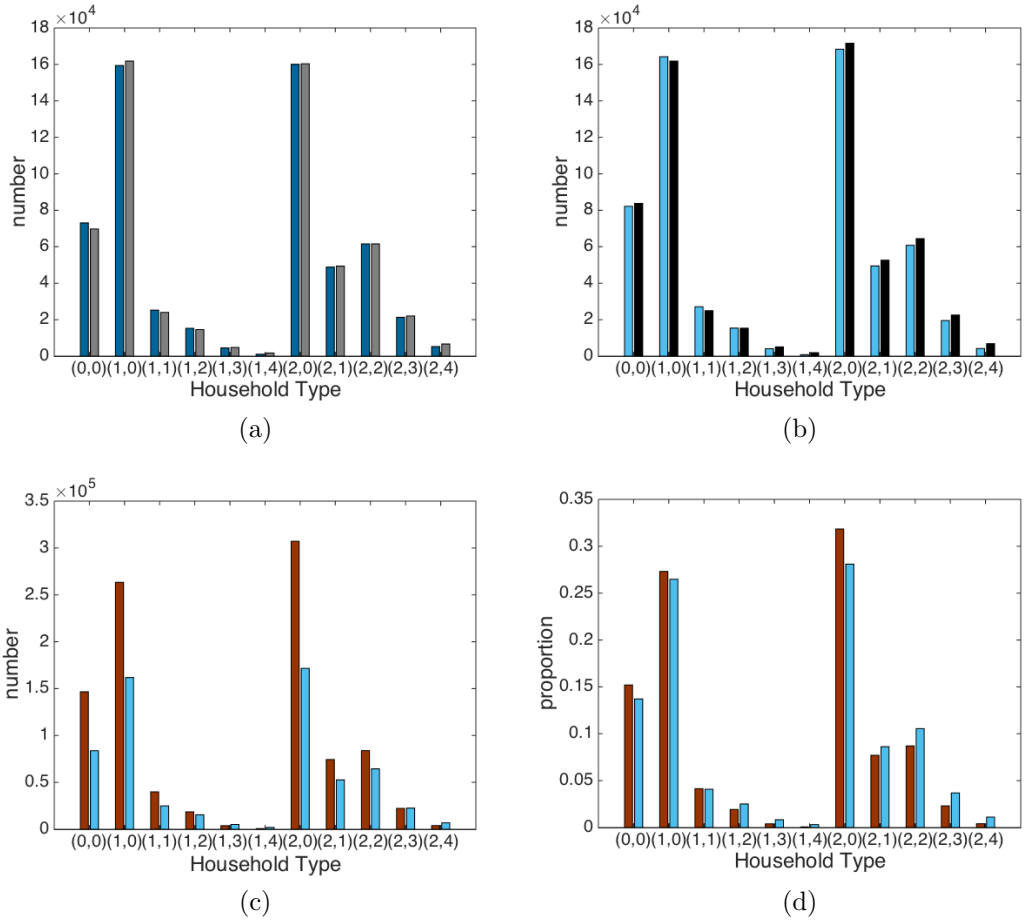


Figure 3.6: Four plots that demonstrate how the parameters in Table 3.4 fit against data. Plot (a) displays the parameterised model after 5 years (dark blue) compared against real 2006 census household composition data (grey). Plot (b) displays the parameterised model after 10 years (light blue) compared against real 2011 census household composition data (black). Plots (c) and (d) display the number of households and the proportion, respectively, of the parameterised model after 100 years (brown) compared against the the parameterised model after 10 years (light blue).

Throughout this chapter, we have developed a CTMC that models population dynamics. In Chapter 4, we extend the CTMC by adding disease dynamics by defining events and rates for the transmission of the disease, moving from the exposed class to the infectious class, recovering from the disease, and losing immunity to the disease.

Chapter 4

Demography with disease model

In this chapter we extend the model in Chapter 3 to include disease dynamics. Recall that the population is made up of M households, each of which is split into one of eleven different household types,

$$\{(0,0)\} \cup \{(i,j) : \ i \in \{1,2\}, \ j \in \{0,1,2,3,4\}\},$$

where i is the number of adults in a household and j is the number of children in a household. We let $m_{(i,j)}$ be the number of households of type (i,j) in the population.

For the disease dynamics, we adopt a SEIR compartmental model. That is, each individual can be in one of the four compartments: susceptible (S), exposed (E), infectious (I) or recovered (R). If an individual is susceptible then they have not contracted the disease but are able to. An exposed individual has the pathogen in their system but is not yet infectious and hence is not able to transmit the disease to susceptible individuals. Infectious individuals are those that are able to infect susceptible individuals, and recovered individuals are those that are currently

immune to contracting the disease and are unable to transmit the pathogen. We represent each household of type (i, j) by the vector

$$H^{(i,j)}_{(S_a,E_a,I_a,S_c,E_c,I_c)} = \left(S_a^{(i,j)}, E_a^{(i,j)}, I_a^{(i,j)}, S_c^{(i,j)}, E_c^{(i,j)}, I_c^{(i,j)}\right),$$

where in a household of type (i, j), $S_a^{(i,j)}$ and $S_c^{(i,j)}$ are the number of susceptible adults and children, respectively, $E_a^{(i,j)}$ and $E_c^{(i,j)}$ are the number of exposed adults and children, respectively, $I_a^{(i,j)}$ and $I_c^{(i,j)}$ are the number of infectious adults and children, respectively. Lastly, although not needed to define the state of the process, define $R_a^{(i,j)}$ and $R_c^{(i,j)}$ to be the number of recovered adults and children, respectively. Henceforth, we will omit the superscript (i, j) to denote what household we are referring to, instead making it clear by prefacing statements with the household type. The number of recovered adults and children in each household are not needed in describing the state because in each household type, there is a fixed number of adults and children. That is, in a household of type (i, j),

$$S_a + E_a + I_a + R_a = i \tag{4.1}$$

$$S_c + E_c + I_c + R_c = j. \tag{4.2}$$

Configurations

In Chapter 3 the households were split into different types. Now, however, the individuals within these households can take on a number of different disease statuses. As such, for each household type (i, j), there are a number of different disease configurations that the inhabitants of the households can be in. By considering all possible combinations for adults, and then for children, we can calculate the total number of configurations that are required for each household type. That is, we can calculate the number of different ways i adults can be

compartmentalised from the set $\{S, E, I, R\}$ with repetition allowed, and then calculate the number of ways j children can be compartmentalised and then we multiply those two numbers together to get the total number of configurations. Hence, we calculate below the number of different configurations that groups of 1, 2, 3, or 4 individuals can take.

If we first consider a group of individuals of size 1, then there are 4 possible combinations $\{S, E, I, R\}$. In a group of size 2, there are 10 possible combinations $\{SS, SE, SI, SR, EE, EI, ER, II, IR, RR\}$. We can determine this for any group size k by considering the number of combinations with repetition from the set of four possible disease statuses $\{S, E, I, R\}$. If we let C_k be the number of configurations that k individuals can take. Then, using the k-multicombination formula,

$$C_k = \left(\binom{4}{k}\right) = \binom{4+k-1}{k} = \frac{(3+k)!}{3!k!}. \tag{4.3}$$

We summarise in Table 4.1 the number of disease configurations for each household type. In total, there are 981 different household configurations.

Household Type	**Combinations**
$(0, 0)$	1
$(1, 0)$	4
$(1, 1)$	16
$(1, 2)$	40
$(1, 3)$	80
$(1, 4)$	140
$(2, 0)$	10
$(2, 1)$	40
$(2, 2)$	100
$(2, 3)$	200
$(2, 4)$	350

Table 4.1: The number of disease configurations for each different household type.

In order to simplify notation, let $\mathbf{v} = (S_a, E_a, I_a, S_c, E_c, I_c)$, and let $\mathbf{e}_i$ be a vector of length 6 consisting of all zeros except for a 1 in the ith component. Further, let $m_{\mathbf{v}}^{(i,j)}$ be the number of households of type $H_{\mathbf{v}}^{(i,j)}$ in the population. Also, let $H^{(0,0)}$ denote an empty household and $m^{(0,0)}$ be the number of empty households in the population. We recall the set $\mathcal{T}$ (Equation (3.1)) and hence write the state space of the disease model by ordering our household configurations by their household type (as in $\mathcal{T}$), and then within each of these types we order the configurations lexicographically. Then we can define the state space of our disease dynamics model, $\mathcal{S}_{dis}$, as

$$\mathcal{S}_{dis} = \Bigg\{ \mathbf{m} = \left(m^{(0,0)}, m_{(1,0,\ldots,0)}^{(1,0)}, m_{(0,1,0,\ldots,0)}^{(1,0)}, \ldots, m_{(0,\ldots,0)}^{(2,4)} \right) : \sum_{(i,j)\in\mathcal{T}} \sum_{\mathbf{v}\in C_{(i,j)}} m_{\mathbf{v}}^{(i,j)} = M,\ m_{\mathbf{v}}^{(i,j)} \geq 0 \ \forall\ (i,j) \in \mathcal{T},\ \mathbf{v} \in C_{(i,j)} \Bigg\}, \tag{4.4}$$

where $C_{(i,j)}$ is the set of all the disease combinations for the household type (i, j).

4.1 Events

In this section we detail each of the events that can happen to our population of households. We split the events into two different categories: *disease dynamics* and *demography dynamics*. First, the disease dynamics are defined. We then redefine the population dynamics, which are summarised in Table 3.1, so that they align with the new formulation of the state space. Also, note that because the demographic events occur on a slower time scale compared to the disease dynamics, all rates (disease and demographic) have been chosen to be defined in terms of years for convenience.

4.1.1 Disease events

While the events for a SEIR disease model are well established, our model can distinguish between adults and children which in turn alters some of these events and rates. We define infection, recovery and waning immunity events, in addition to the two different types of transmission events corresponding to within- and between-household transmission. We define these events, and their associated transition rates, for a household with configuration $H_{\mathbf{v}}^{(i,j)}$. Recall that in a household of type (i, j),

$$R_a = i - S_a - E_a - I_a \text{ and } R_c = j - S_c - E_c - I_c.$$

i. Exposed becoming infectious

An exposed adult will become infectious after a latent period with mean $\frac{1}{\sigma}$ years, so the transition

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}-\mathbf{e}_2+\mathbf{e}_3}^{(i,j)}$$

occurs at rate $\sigma E_a m_{\mathbf{v}}^{(i,j)}$. With rate $\sigma E_c m_{\mathbf{v}}^{(i,j)}$, we can similarly define for children the transition for an exposed child becoming infectious,

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}-\mathbf{e}_5+\mathbf{e}_6}^{(i,j)}.$$

ii. Recovery

An infectious adult will recover from the disease after an infectious period with mean $\frac{1}{\gamma}$ years. Hence we see an adult recovery event in the household type $H_{\mathbf{v}}^{(i,j)}$ at rate $\gamma I_a m_{\mathbf{v}}^{(i,j)}$, with transition

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}-\mathbf{e}_3}^{(i,j)}.$$

Similarly, we see a child recovery event in the household type $H_{\mathbf{v}}^{(i,j)}$ at rate $\gamma I_c m_{\mathbf{v}}^{(i,j)}$, with transition

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}-\mathbf{e}_6}^{(i,j)}.$$

iii. Waning immunity

Often in epidemic models, an individual has life-long immunity to the infectious disease. However, we define the event of waning immunity to allow for the possibility that immunity fades over time because for endemic diseases, the time scale is much longer. This means we can investigate diseases that continue to infect people even if they have been sick with the infectious disease previously.
A recovered adult will lose immunity after they recover, on average, in $\frac{1}{\omega}$ years. Then the event where an adult loses their immunity, in a household type $H_{\mathbf{v}}^{(i,j)}$, occurs at rate $\omega R_a m_{\mathbf{v}}^{(i,j)}$, with transition

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_1}^{(i,j)}.$$

Similarly, we see a recovered child lose immunity in a household type $H_{\mathbf{v}}^{(i,j)}$ at rate $\omega R_c m_{\mathbf{v}}^{(i,j)}$, with transition

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_4}^{(i,j)}.$$

iv. Within-household transmission

In a household of type $H_{\mathbf{v}}^{(i,j)}$, where $i + j \geq 2$, an infectious individual can transmit the disease to a susceptible individual in the same household. We call this a *within-household*, or *local, transmission*. Let $\boldsymbol{\beta}$ be a matrix of transmission parameters such that,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{aa} & \beta_{ac} \\ \beta_{ca} & \beta_{cc} \end{pmatrix},$$

where β_{aa} is the local transmission parameter corresponding to an adult transmitting the disease to another adult, and similarly β_{ac} is the parameter corresponding to an adult to child transmission, β_{ca} corresponds to a child transmitting the disease to an adult, and lastly β_{cc} is the parameter corresponding to a child-to-child transmission.

An adult will become infected by other members of their household at rate

$$\frac{S_a\left(\beta_{aa}I_a + \beta_{ca}I_c\right)}{i+j-1}\, m_{\mathbf{v}}^{(i,j)}$$

where the associated transition is given by

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_2-\mathbf{e}_1}^{(i,j)}.$$

Similarly, we define a child being infected from within their own household at rate

$$\frac{S_c\left(\beta_{ac}I_a + \beta_{cc}I_c\right)}{i+j-1}\, m_{\mathbf{v}}^{(i,j)}$$

where the associated transition is given by

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_5-\mathbf{e}_4}^{(i,j)}.$$

v. Between-household transmission

In the population of M households, an infectious individual can also transmit the disease to any susceptible individual in the population, which we call *between-household*, or *external, transmission*. Let I_a^T be the total number of infectious adults in the population, and I_c^T be the total number of infectious children in the population. As we did for local transmission, we can create a matrix of

between-household transmission parameters and label it $\boldsymbol{\alpha}$, where,

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{aa} & \alpha_{ac} \\ \alpha_{ca} & \alpha_{cc} \end{pmatrix},$$

where α_{aa} is the between-household transmission parameter corresponding to an adult transmitting the disease to another adult, and similarly α_{ac} is the parameter corresponding to an adult to child transmission, α_{ca} is the parameter corresponding to a child transmitting the disease to an adult, and lastly α_{cc} is the parameter corresponding to a child-to-child transmission.

Then a susceptible adult in a household of type $H_{\mathbf{v}}^{(i,j)}$ will contract the disease and hence become exposed to the pathogen at rate,

$$\frac{S_a\left(\alpha_{aa}I_a^T + \alpha_{ca}I_c^T + M\epsilon\right)}{N} m_{\mathbf{v}}^{(i,j)},$$

where N is the total number of individuals in the population, M is the total number of households in the population, and ϵ is a small importation term per household that allows a disease to migrate into the population. The transition is given by

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_2-\mathbf{e}_1}^{(i,j)}.$$

Similarly, a susceptible child in a household of type $H_{\mathbf{v}}^{(i,j)}$ will contract the disease and hence become exposed to the pathogen at rate,

$$\frac{S_c\left(\alpha_{ac}I_a^T + \alpha_{cc}I_c^T + M\epsilon\right)}{N} m_{\mathbf{v}}^{(i,j)}$$

with transition given by

$$H_{\mathbf{v}}^{(i,j)} \to H_{\mathbf{v}+\mathbf{e}_5-\mathbf{e}_4}^{(i,j)}.$$

4.1.2 Demography events

While the demography events have already been outlined in Section 3.1, the rates and transitions must be reformulated to align with the new definition of the state space introduced earlier in Section 4. Here we redefine the population dynamics events for a household configuration of $H_{\mathbf{v}}^{(i,j)}$.

i. Births

In this event, a household with two adults and no more than three children can experience a birth event. We assume that a child is born susceptible. That is, the household will gain one susceptible child. Let $i = 2$ and $j \leq 3$. Then we can define a birth event by the transition,

$$H_{\mathbf{v}}^{(2,j)} \to H_{\mathbf{v}+\mathbf{e}_4}^{(2,j+1)}$$

at rate $b_j m_{\mathbf{v}}^{(2,j)}$.

ii. Deaths

There are three different types of deaths: adult (1), adult (2), and children. We look at each of them separately.

Adult Death (1): This event occurs in households with only one adult. We remove the adult from the household and assume that any children left parentless in the household are removed from the population. Hence this is equivalent to just removing the household as an emigration event. Here, because we only have a single adult, we let $i = 1$. Then a single adult death event is characterised by the transition

$$H_{\mathbf{v}}^{(1,j)} \to H^{(0,0)}$$

with rate $d_A m_{\mathbf{v}}^{(1,j)}$.

Adult Death (2): Alternatively, if there are two adults in a household, one is picked randomly and is removed from the household. Now, let $i = 2$. Then a couple household (with or without children) death event can be defined as

$$H_{\mathbf{v}}^{(2,j)} \rightarrow \begin{cases} H_{\mathbf{v}-\mathbf{e}_1}^{(1,j)} & \text{with rate } S_a d_A m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}-\mathbf{e}_2}^{(1,j)} & \text{with rate } E_a d_A m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}-\mathbf{e}_3}^{(1,j)} & \text{with rate } I_a d_A m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}}^{(1,j)} & \text{with rate } R_a d_A m_{\mathbf{v}}^{(i,j)} \end{cases}.$$

Children Death: Lastly, for deaths, in a household with at least one child, we randomly pick one child and remove them from the system. For $j \geq 1$, we can define children death events as the transition

$$H_{\mathbf{v}}^{(i,j)} \rightarrow \begin{cases} H_{\mathbf{v}-\mathbf{e}_4}^{(i,j-1)} & \text{with rate } S_c d_C m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}-\mathbf{e}_5}^{(i,j-1)} & \text{with rate } E_c d_C m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}-\mathbf{e}_6}^{(i,j-1)} & \text{with rate } I_c d_C m_{\mathbf{v}}^{(i,j)} \\ H_{\mathbf{v}}^{(i,j-1)} & \text{with rate } R_c d_C m_{\mathbf{v}}^{(i,j)} \end{cases}.$$

iii. Children moving out

If $j \geq 1$, then an event where a child moves out of home can occur. Just as was defined in Chapter 3, there are two possibilities. A child can move out of home and either leave the population, effectively emigrating, or move into an empty household and hence become an adult. If we define the function that determines the probability of an individual deciding to move into an empty household to be the preference function, Equation (3.3), then we can write down the transitions and rates at which they occur.

Move into empty household: A child will leave home and move into an empty household in the following way:

$$\left(H_{\mathbf{v}}^{(i,j)}, H^{(0,0)}\right) \to \begin{cases} \left(H_{\mathbf{v}-\mathbf{e}_4}^{(i,j-1)}, H_{\mathbf{e}_1}^{(1,0)}\right) & \text{with rate } S_c m_{out} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}-\mathbf{e}_5}^{(i,j-1)}, H_{\mathbf{e}_2}^{(1,0)}\right) & \text{with rate } E_c m_{out} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}-\mathbf{e}_6}^{(i,j-1)}, H_{\mathbf{e}_3}^{(1,0)}\right) & \text{with rate } I_c m_{out} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}}^{(i,j-1)}, H_{\mathbf{0}}^{(1,0)}\right) & \text{with rate } R_c m_{out} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \end{cases}.$$

Emigrating: In the case where a child decides to move out of home and leave the population, the transitions and rates are:

$$H_{\mathbf{v}}^{(i,j)} \to \begin{cases} H_{\mathbf{v}-\mathbf{e}_4}^{(i,j-1)} & \text{with rate } S_c m_{out} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}-\mathbf{e}_5}^{(i,j-1)} & \text{with rate } E_c m_{out} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}-\mathbf{e}_6}^{(i,j-1)} & \text{with rate } I_c m_{out} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}}^{(i,j-1)} & \text{with rate } R_c m_{out} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \end{cases}.$$

iv. Couple separation

When $i = 2$, couples can separate. Each adult in a household has equal probability of being removed from the household, given that a couple separation event is occurring. As the couple separation event in a household with two adults occurs with rate c_s, each adult is removed from the household due to couple separation at a rate of $\frac{1}{2}c_s$. As with children moving out, the adult leaving the household can either leave the population or move into an empty household. Again, we use the preference function, Equation (3.3), for determining the probability of an adult staying in the population and hence moving into an empty household.

Move into empty household: When an adult separates from their partner and chooses to move into an empty household, the transitions and rates can be written down in the following way:

$$\left(H_{\mathbf{v}}^{(2,j)}, H^{(0,0)}\right) \to \begin{cases} \left(H_{\mathbf{v}-\mathbf{e}_1}^{(1,j)}, H_{\mathbf{e}_1}^{(1,0)}\right) & \text{with rate } c_s \frac{S_a}{2} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}-\mathbf{e}_2}^{(1,j)}, H_{\mathbf{e}_2}^{(1,0)}\right) & \text{with rate } c_s \frac{E_a}{2} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}-\mathbf{e}_3}^{(1,j)}, H_{\mathbf{e}_3}^{(1,0)}\right) & \text{with rate } c_s \frac{I_a}{2} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \\ \left(H_{\mathbf{v}}^{(1,j)}, H_{\mathbf{0}}^{(1,0)}\right) & \text{with rate } c_s \frac{R_a}{2} m_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) \end{cases}.$$

Emigrating: In the case where the adult leaves the population instead of moving into an empty household, the transitions and rates are:

$$H_{\mathbf{v}}^{(2,j)} \to \begin{cases} H_{\mathbf{v}-\mathbf{e}_1}^{(1,j)} & \text{with rate } c_s \frac{S_a}{2} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}-\mathbf{e}_1}^{(1,j)} & \text{with rate } c_s \frac{E_a}{2} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}-\mathbf{e}_1}^{(1,j)} & \text{with rate } c_s \frac{I_a}{2} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \\ H_{\mathbf{v}}^{(1,j)} & \text{with rate } c_s \frac{R_a}{2} m_{\mathbf{v}}^{(i,j)} \left(1 - f\left(\frac{m^{(0,0)}}{M}\right)\right) \end{cases}.$$

v. Couple formation

If $i = 1$, then a household can undergo the event of couple formation, where a lone adult can merge with another household with one adult. The transitions for

these events can be denoted by the following:

$$\left(H_{\mathbf{v}}^{(1,j)}, \left\{ \begin{array}{l} H_{\mathbf{e}_1}^{(1,0)} \\ H_{\mathbf{e}_2}^{(1,0)} \\ H_{\mathbf{e}_3}^{(1,0)} \\ H_{\mathbf{0}}^{(1,0)} \end{array} \right\} \right) \rightarrow \left\{ \begin{array}{ll} \left(H^{(0,0)}, H_{\mathbf{v}+\mathbf{e}_1}^{(2,j)} \right) & \text{with rate } \frac{c_f}{M} m_{\mathbf{e}_1}^{(1,0)} m_{\mathbf{v}}^{(i,j)} \\ \left(H^{(0,0)}, H_{\mathbf{v}+\mathbf{e}_2}^{(2,j)} \right) & \text{with rate } \frac{c_f}{M} m_{\mathbf{e}_2}^{(1,0)} m_{\mathbf{v}}^{(i,j)} \\ \left(H^{(0,0)}, H_{\mathbf{v}+\mathbf{e}_3}^{(2,j)} \right) & \text{with rate } \frac{c_f}{M} m_{\mathbf{e}_3}^{(1,0)} m_{\mathbf{v}}^{(i,j)} \\ \left(H^{(0,0)}, H_{\mathbf{v}}^{(2,j)} \right) & \text{with rate } \frac{c_f}{M} m_{\mathbf{0}}^{(1,0)} m_{\mathbf{v}}^{(i,j)} \end{array} \right. .$$

vi. Emigration and immigration

Households of all types can emigrate (leave the population), and any type of household can immigrate (enter the population) through an empty household. Define $p_{\mathbf{v}}^{(i,j)}$ to be the probability that a household of type (i,j) and configuration $\mathbf{v}$ will experience an emigration or immigration event, conditioned on an emigration or immigration event occurring in the population.

Emigration: A household will leave the population at instantaneous transition rate $\phi p_{\mathbf{v}}^{(i,j)} m_{\mathbf{v}}^{(i,j)}$. Then the transition is

$$H_{\mathbf{v}}^{(i,j)} \rightarrow H^{(0,0)}.$$

Immigration: An empty household will become inhabited by an immigrating household at instantaneous transition rate $\xi p_{\mathbf{v}}^{(i,j)} f\left(\frac{m^{(0,0)}}{M}\right) m^{(0,0)}$. The preference function is used to measure how likely a family is to choose a population with $m^{(0,0)}$ empty households. The transitions are

$$H^{(0,0)} \rightarrow H_{\mathbf{v}}^{(i,j)}.$$

vii. New empty household

New empty households, $H^{(0,0)}$, are built at rate $Mg > 0$.

4.2 Seasonal forcing

There is strong evidence that the number of infectious individuals can peak seasonally for certain diseases [31, 32] due to the increased transmission rates in children during the school term when children are in close proximity to one another and there are many contacts between individuals. As we are primarily motivated by measles, a disease that predominantly impacts children, we introduce seasonal forcing to our model, which essentially increases transmission of the disease during particular seasons which is outlined in Table 4.2. This table presents an example school year based on Adelaide, South Australia school terms in 2016. During the school terms, the rates of transmission are changed.

Time of Year	**Dates**
Summer holidays	1st Jan – 31st Jan
Term 1	1st Feb – 15th April
Term 1 holidays	16th April – 1st May
Term 2	2nd May – 8th July
Term 2 holidays	9th July – 24th July
Term 3	25th July – 30th Sept
Term 3 holidays	1st Oct – 16th Oct
Term 4	17th Oct – 16 Dec
Summer holidays	17th Dec – 31st Dec

Table 4.2: The dates of school terms in 2016.

Let $\boldsymbol{\beta}(t)$ and $\boldsymbol{\alpha}(t)$ be the time dependent within and between household transmission matrices, respectively. While it is common to use sinusoidal functions as a means to seasonal forcing [16], the step function more accurately captures the

dynamics [5]. We can write,

$$\boldsymbol{\beta}(t) = \boldsymbol{\beta}^{(h)} + \mathbb{I}_{\{t\in T_s\}}\boldsymbol{\beta}^{(s)},$$
$$\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}^{(h)} + \mathbb{I}_{\{t\in T_s\}}\boldsymbol{\alpha}^{(s)},$$

where $\boldsymbol{\beta}^{(h)}$ and $\boldsymbol{\alpha}^{(h)}$ are the transmission matrices parameterised as the holiday transmission rates, and where $\boldsymbol{\beta}^{(s)}$ and $\boldsymbol{\alpha}^{(s)}$ are the changes to the holiday transmission matrices during the school term, and where $\mathbb{I}_{\{\bullet\}}$ is the indicator function and T_s is a set containing all the times at which children are at school as shown in Table 4.2.

4.3 Analysis

Similarly to Section 3.3, we discuss the theory and methods used for analysing our stochastic model. The two methods that we implement are, again, a stochastic simulation method and a deterministic approximation for the expected behaviour of our stochastic system.

4.3.1 Stochastic simulation

We implement the so-called Gillespie algorithm [22] as outlined in Section 3.3.1. Here, however, the generic variables used will be defined. The state space of the process we are simulating is $\mathcal{S}_{dis}$ (Equation 4.4), and we define the initial condition (for the results in Chapter 5) to be an entirely susceptible population except for two households of type $(2, 2)$ and $(2, 3)$, each with one child exposed to the disease. We choose these household types to initialise the disease because they have a large number of individuals in the household and so we increase the probability that the disease will be transmitted to others at the beginning of the outbreak.

Accounting for each household configuration and each household type, there are a total of 30209 different events, many of which will not be possible at each jump time. Let $\mathbf{r}$ be a vector of length 30209 with components corresponding to the rate of each event. Then Λ is the corresponding stoichiometric matrix of size 981×30209, where column c is the net change in the number of households of each type corresponding to the event with rate in the cth component of $\mathbf{r}$. With this information, the system can be simulated using the so-called Gillespie algorithm.

Figure 4.1 shows one realisation of the number of infectious individuals over five years, as well as the solution to the deterministic formulation that we define shortly. Simulating over this period takes approximately 15 minutes, implemented in Matlab. However, the majority of this time is spent in the initial peak as in a population of approximately one million individuals where there are only a very small number of infectious individuals initially, over a thousand individuals will contract the disease, which means that there has to be at least that many transitions in the stochastic simulation, which is very computationally expensive if there are recurring peaks (which can occur for different initial conditions), especially when many realisations are required.

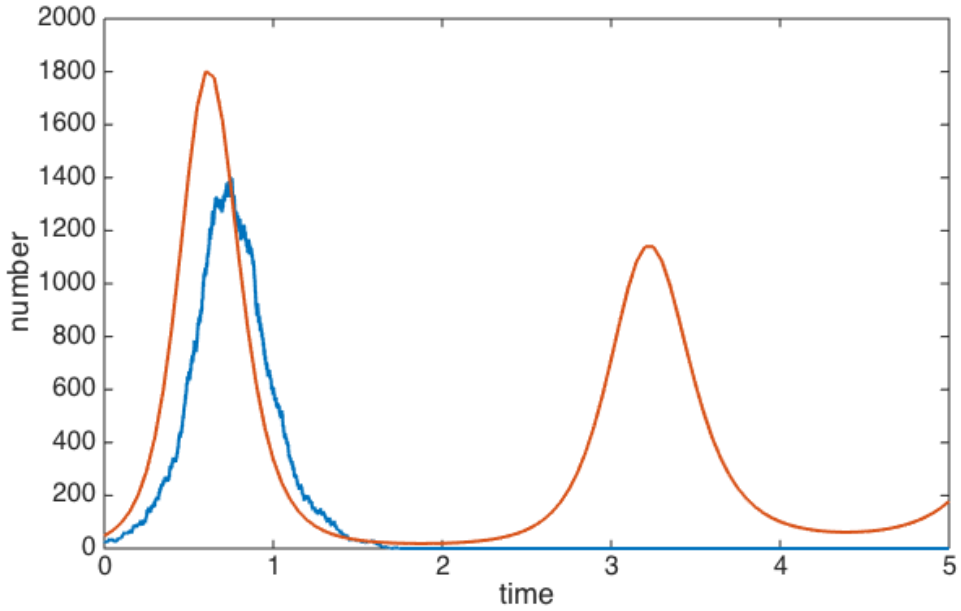


Figure 4.1: A single stochastic realisation (blue) and the deterministic trajectory (orange) of the number of infectious individuals in the population. The initial condition for this is setting the proportion of susceptible individuals in the population to be 8%, and recovered individuals 91%, and 0.5% each of exposed and infectious, and parameters to be defined in Section 4.4.

4.3.2 Deterministic approximation

Stochastic simulation is computationally expensive and so we use the theory proved by Kurtz [23] to develop the differential equations approximation of the expected trajectories of the household configurations. In order to use the approximation as in Section 3.3.2, the CTMC needs to be density dependent. Using the same argument as in Section 3.3.2, it can be shown that our disease model has the required properties in order to use this approximation. The details, however, are omitted due to the large number of events and transition rates, each of which has the same form as that in the demography model; that is, each rate is either linear, quadratic or constant, each of which have been shown to be density dependent previously in Section 3.3.2.

Next, we progress to constructing the system of differential equations. Recall the formulation of the system of differential equations outlined in Section 3.3.2.

The events in Section 4.1 have very similar form and so the system of differential equations will not be linear and hence we construct a number of matrices as we did in Section 3.3.2. Let i_a^T and i_c^T be the total proportion of infectious adults and children in the population. Then we can define the between-households force of transmitting the disease to a susceptible adult and child, respectively, as,

$$\theta_a(\boldsymbol{x}) = \frac{\left(\alpha_{aa} i_a^T + \alpha_{ca} i_c^T + \epsilon\right)}{\left(\sum_{(i,j),\mathbf{v}} x_{\mathbf{v}}^{(i,j)}(i+j)\right)},$$
$$\theta_c(\boldsymbol{x}) = \frac{\left(\alpha_{ac} i_a^T + \alpha_{cc} i_c^T + \epsilon\right)}{\left(\sum_{(i,j),\mathbf{v}} x_{\mathbf{v}}^{(i,j)}(i+j)\right)}.$$

Then the form of the differential equations we are constructing is,

$$\begin{aligned}\frac{d\boldsymbol{x}}{dt} = \Big[& F_1 + f\left(x^{(0,0)}\right)(F_2 + F_4) + \left(1 - f\left(x^{(0,0)}\right)\right) F_3 \\ & + \theta_a(\boldsymbol{x}) F_5 + \theta_c(\boldsymbol{x}) F_6 \\ & + x_{\mathbf{e}_1}^{(1,0)} F_S + x_{\mathbf{e}_2}^{(1,0)} F_E + x_{\mathbf{e}_3}^{(1,0)} F_I + x_{\mathbf{0}}^{(1,0)} F_R \Big] \boldsymbol{x} + G,\end{aligned} \tag{4.5}$$

where $G = (g, 0, ..., 0)^T$, which increases the proportion of empty households at a constant rate g, f is the preference function, and the matrices F_1, F_2, F_3, F_4, F_5, F_6, F_S, F_E, F_I, and F_R are briefly described below, and defined rigorously in Appendix B. They are not detailed in full here because their derivation is very similar to that of the deterministic approximation for the demography model in Section 3.

To describe the matrices $F_1, ..., F_6, F_S, F_E, F_I$, and F_R, the different events are split up into different sets depending on the form of their associated rates. Let $\mathcal{E}_{lin}$ be the set of all the event types with linear rates, except for deaths; we separate deaths from the other events with linear rates because once a household is chosen, an individual must be picked depending on their health status, which complicates the rate formulation. That is, $\mathcal{E}_{lin}$ contains the events: births,

emigration, within-household transmission, infection, recovery, and waning immunity. Let $\mathcal{E}_d$ be the set of child and adult deaths. Let $\mathcal{E}_{dec}$ be the set of events where the individuals have two possible "decisions"; that is, the moving out of home and couple separation events are in $\mathcal{E}_{dec}$. Let $\mathcal{E}_{im}$ be the set containing the immigration event, and let $\mathcal{E}_{cf}$ be the set containing the couple formation event. Lastly, let $\mathcal{E}_{ext}$ be the set containing the external transmission event.

The definitions of the matrices can be found in Appendix B. Here we outline their qualitative interpretation. As in Section 3, these matrices quantify the flux in and out of each household configuration. The matrix F_1 corresponds to events in $\mathcal{E}_{lin}$. F_2 corresponds to the rates and transitions of children moving out of home and couple separation where the individual leaving their previous household *remains* in the population, whereas F_3 is for the children moving out of home and couple separation events where the individual *leaves* the population. The matrix F_4 corresponds to immigration events, and F_5 and F_6 correspond to between-household transmission events for adults and children, respectively. F_S, F_E, F_I, and F_R correspond to couple formation events where the single adult in the household $(1, 0)$ is susceptible, exposed, infectious, or recovered respectively.

Recall Figure 4.1. It shows a single realisation, which fades out stochastically after the first peak. This realisation took just under 15 minutes to run in Matlab, whereas the deterministic trajectory took less than a minute to compute. Hence, where stochastic effects are not important, the deterministic model is utilised due to the overwhelming increase in computational efficiency.

4.4 Parametrisation

To parametrise this model, we need parameter values for both the population and disease dynamics. The parameters associated with the demographic events

are the same as defined in Section 3.4. The only parameters that need to be assigned values are the transmission parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and the associated seasonal forcing transmission parameters, the parameter associated with the latency period, σ, and the infection period, γ, and waning immunity, ω. There is also a parameter, ϵ, which is the immigration of the disease into the population per household, that needs to be determined. In this section we present two sets of baseline disease parameters, for measles- and influenza-like illnesses, and discuss some of the assumptions of the parametrisation. However, we make clear that while these baseline parameters are used extensively throughout the thesis, they may be changed/varied due to the nature of the problems we consider. For example, when we investigate the behaviour of the seasonally-forced model for different transmission rates, we will vary the parameters found in this section. However, we will always make the reader aware of these variations.

Recall the form of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, both of which are matrices encoding transmission rates between adults and children. For diseases that are primarily prevalent in children, it is typical to see the child-to-child transmission element different compared to the others due to the contacts children make at school [1]. To simplify the parametrisation, we have symmetric transmission between adults and children. That is,

$$\beta_{ca} = \beta_{ac}, \quad \alpha_{ca} = \alpha_{ac}.$$

It is also assumed that each element in $\boldsymbol{\alpha}$ is smaller than the corresponding element of $\boldsymbol{\beta}$ as the interaction between individuals within a household is stronger than that between any two individuals in the population.

The parameters σ and γ are chosen such that the mean length of the latency period is $\frac{1}{\sigma}$ years and the mean length of time that an individual is infectious is $\frac{1}{\gamma}$ years. Similarly, the parameter associated with waning immunity, ω is chosen such that individuals lose immunity to the disease after, on average, $\frac{1}{\omega}$ years.

The parameter pertaining to the immigration of the disease, ϵ, is a term that can be used to prevent a disease from dying out stochastically by reintroducing it from an external source. To avoid the effect of imported infection dominating the dynamics, the value of ϵ is relatively small compared to the other disease event parameters which means that the population will get an imported case of the disease infrequently.

All disease parameters, except for those associated with transmission, in Table 4.3 are taken from [1, 33]. The transmission parameters, however, were much harder to determine, especially since there is not an easy way to define R_0 for our model, where R_0 is defined to be the average number of secondary cases arising from an average primary case in an entirely susceptible population [1]. Hence, we make some approximations for our transmission rates for measles and influenza. Measles is often documented to have an R_0 value between 13-18 [1], whereas for influenza R_0 is approximately between 2-4 [1, 33, 34], depending on the strain.

In the SEIR model with births and deaths occuring at rate μ, $R_0 = \frac{\beta\sigma}{(\gamma+\mu)(\sigma+\mu)}$ [1]. However, as the lifespan of an individual is much longer than the length of time associated with the disease dynamics, $R_0 \approx \frac{\beta}{\gamma}$. Hence we pick transmission parameters that will approximately produce γR_0 transmission events over the course of an average infectious period.

We have two levels of mixing, and adults and children each with different transmission rates, so there are many ways to pick transmission parameters that produce approximately γR_0 transmission events. For within-household transmission, we let adult-to-adult and children-to-children rates be larger than the between-household transmission rates so that the model produces approximately γR_0 transmission events, and then adult-to-children, and vice versa, are picked at a much lower rate to reflect the lack of interaction between adults and children outside of a household. All disease parameters are summarised in Table

4.3.

We also parameterise the seasonal forcing model for measles in Table 4.4. That is, we define $\boldsymbol{\alpha}^{(h)}, \boldsymbol{\alpha}^{(s)}, \boldsymbol{\beta}^{(h)}$, and $\boldsymbol{\beta}^{(s)}$. The transmission parameters for the non-forced model (in Table 4.3) are calculated by taking the average of the seasonally-forced transmission matrices. That is, $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(h)} + t_s\boldsymbol{\alpha}^{(h)}$, where t_s is the proportion of the year that children spend in school. We similarly define $\boldsymbol{\beta} = \boldsymbol{\beta}^{(h)} + t_s\boldsymbol{\beta}^{(s)}$.

Parameter	Measles	Influenza
$\boldsymbol{\beta}$	$\begin{pmatrix} 2.4286 & 0.5714 \\ 0.5714 & 2.5177 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.6 \\ 0.6 & 2 \end{pmatrix}$
$\boldsymbol{\alpha}$	$\begin{pmatrix} 1.8571 & 0.2920 \\ 0.2920 & 2.2203 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 1.2 \end{pmatrix}$
σ^{-1}	5 days	1 day
γ^{-1}	7 days	3.8 days
ω	0	0
ϵ	0	0

Table 4.3: Values of the disease dynamics parameters that are used throughout this thesis. The $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ values in the table can be interpreted as contacts per person per day, and are calculated by averaging the seasonally-forced transmission matrices in Table 4.4. It is assumed that neither disease above allows immunity to wane over time, and there is no external infection unless stated otherwise.

Year Round	Forced School Term	Average
$\boldsymbol{\beta}^{(h)} = \begin{pmatrix} 2.4286 & 0.5714 \\ 0.5714 & 2.4286 \end{pmatrix}$	$\boldsymbol{\beta}^{(s)} = \begin{pmatrix} 0 & 0 \\ 0 & 0.1214 \end{pmatrix}$	$\begin{pmatrix} 2.4286 & 0.5714 \\ 0.5714 & 2.5177 \end{pmatrix}$
$\boldsymbol{\alpha}^{(h)} = \begin{pmatrix} 1.8571 & 0.2857 \\ 0.2857 & 2.0000 \end{pmatrix}$	$\boldsymbol{\alpha}^{(s)} = \begin{pmatrix} 0 & 0.0086 \\ 0.0086 & 0.3000 \end{pmatrix}$	$\begin{pmatrix} 1.8571 & 0.2920 \\ 0.2920 & 2.2203 \end{pmatrix}$

Table 4.4: Values of the seasonally forced parameters and their average value used for the non-forced model, as seen in Table 4.3. All other parameters, $\sigma, \gamma, \omega, \epsilon$ are as in Table 4.3.

Chapter 5

Results

A CTMC modelling disease with demography dynamics has been developed in Chapter 4. We consider the problem of *epidemic fadeout* [35, 6] and discuss the *critical community* size [12], and use our model to gain some more insight on the topic. We also consider the periodicity of our model, both with and without seasonal forcing, and investigate the effect of a disease that poses serious health risks for people that are infectious; that is, we increase the death rate for infectious individuals and determine how the population dynamics change.

5.1 Periodicity

Endemic diseases persist in populations for long periods of time. It is, then, intuitive to investigate the long-term behaviour of the household model. The demography dynamics with a fixed number of households will reach an equilibrium whereas the model with the parameter set which increases the number of households at a constant rate, will never reach equilibrium. In this section we consider the parameter set with a fixed number of households.

To find the equilibrium, we approximate the expected behaviour of the CTMC by solving the differential equations derived in Section 4.3.2, which is valid in

the limit as the number of households approaches infinity. We then consider the equilibrium of this deterministic approximation, and determine the proportion of individuals that are infectious and the period of oscillation if the system is perturbed away from the fixed point. To find the period of oscillations, the Jacobian of the system of differential equations is evaluated numerically at the fixed point using D'Errico's *Adaptive Robust Numerical Differentiation* Matlab package [26] as introduced in Section 2.3. Then the eigenvalues of it are calculated numerically using Matlab's inbuilt `eig` function. The eigenvalues of the model are displayed in Figure 5.1. We note that they are all negative, which means the system is stable.

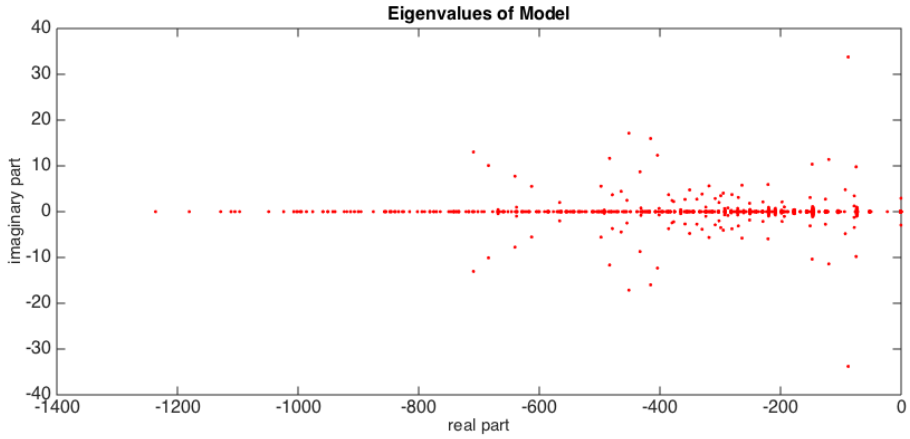


Figure 5.1: Plot of the eigenvalues of the Jacobian matrix associated with the deterministic household model, evaluated at the stable fixed point. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.3.

The imaginary parts of the eigenvalues are associated with the periodicity of the model, whereas the real parts of the eigenvalues are associated with the dampening of the deterministic trajectory [28]. That is, the eigenvalues with the smallest magnitude real part will dominate the dynamics of the system in the long term. Then the periodicity that we are interested in can be found by considering the eigenvalues with smallest magnitude real part that also have a non-zero imaginary part. Figure 5.2 displays the eigenvalues with the smallest

magnitude real parts and the associated imaginary parts.

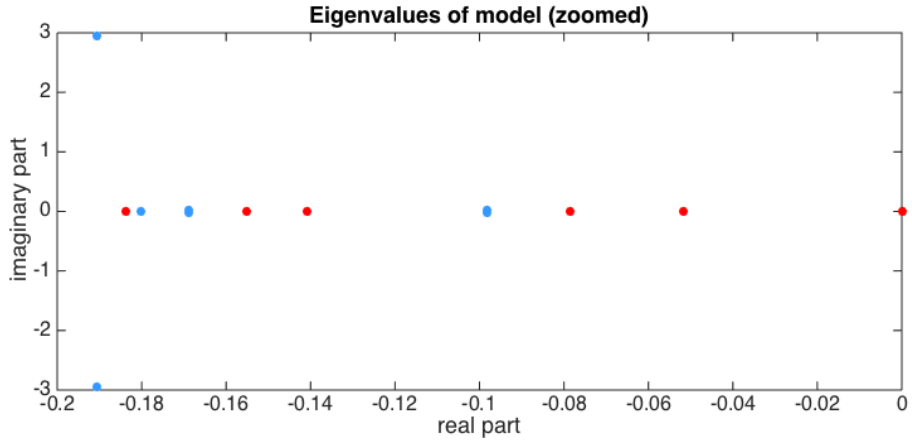


Figure 5.2: The eigenvalues of the Jacobian matrix associated with the deterministic household model, evaluated at the fixed point, with real parts greater than -0.2. The blue points are eigenvalues with non-zero imaginary parts. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.3.

It can be seen that there are four complex conjugate pairs (highlighted in blue) in Figure 5.2 that have a real part with magnitude less than 0.2. Taking the positive conjugate of each of them, the eigenvalues are,

$$\lambda_1 = -0.1907 + 2.9423i, \quad \lambda_2 = -0.0981 + 0.0250i,$$
$$\lambda_3 = -0.1688 + 0.0219i, \quad \lambda_4 = -0.1801 + 0.0092i.$$

The period of oscillation, Θ_i, associated with λ_i is given in [28] by,

$$\Theta_i = \frac{2\pi}{\mathrm{Im}(\lambda_i)}.$$

Hence the periods associated with λ_1, λ_2, λ_3 and λ_4 are 2.14, 251.80, 286.69 and 679.38 years respectively. The periods 251.80, 286.69 and 679.38 are very large and are associated with the demographics, which occur on a much slower time scale than the disease dynamics. We effectively disregard these periods. The other period, 2.14, is associated with the disease dynamics. We expect, for

measles, to see a period of approximately 2 years as it is well established that the number of infectious individuals in a population peaks approximately every two years [2].

Further, we can determine the periodicity of the household model with seasonal forcing as well. The deterministic model is run for approximately 150 years before the trajectories transients die down to the stable attractor, and oscillatory motion repeats. In Figure 5.3 we see the trajectories of the proportions, and the slight changes for different values of $\alpha_{cc}^{(h)}$. We record the proportion of the population that is infectious at the same time each year, and then the system is perturbed by varying $\alpha_{cc}^{(h)}$, which is an easy and effective way to change the rate of transmission between children. This is repeated for a number of different values, and the results are shown in Figure 5.4. This graph shows that the proportion of infectious individuals is approximately the same at the same time each year, implying that the disease has an annual period.

We follow a very similar procedure again but instead of perturbing the external transmission parameter, we perturb the birth rate of a household of type $(2, 0)$, because the replenishment of susceptible individuals is a driving factor in endemic disease dynamics, and births into households of type $(2, 0)$ also increases the number of births into $(2, 1), (2, 2)$ and $(2, 3)$ households. Again, the period of the disease dynamics is annual, as seen in Figures 5.5 and 5.6. In fact, as b_0 increases, there is no significant change in the proportion of infectious individuals that we record each year.

We also investigate the periodicity of the stochastic dynamics. For the measles parameters in Table 4.4, the deterministic model is evaluated over 150 years. The state of the system at the end of 150 years is taken as the initial state for the stochastic model and three hundred and seventy-five simulations are run over 30 years. Of the three hundred and seventy-five simulations, three hundred and thirty-six fade out. We disregard these and consider only those that do not fade

out in order to simplify the analysis. We plot eight of the realisations that did not fadeout in Figure 5.7, and note the period of each realisation is approximately between 2-3 years. The average of the thirty-nine realisations that did not fadeout is plotted in Figure 5.8 and we see that the realisations become out of phase after 2-3 peaks. Between 15-20 years into the trajectory, we see that the peaks begin to appear approximately yearly, which agrees with the deterministic approximation. This has shown that although each realisation has approximate period of 2-3 years, averaging them produces a different periodicity.

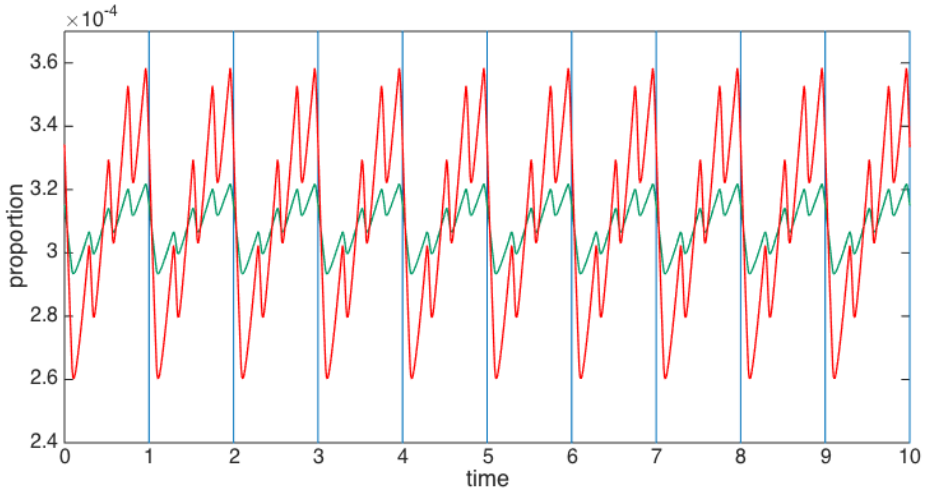


Figure 5.3: The trajectories of the proportion of infectious individuals in the population for $\alpha_{cc}^{(h)} = 0.05$ (green) and $\alpha_{cc}^{(h)} = 0.2$ (red). The light blue reference lines are times that we record the proportion. The demography parameters are those summarised in Table 3.3, and the measles parameters that are not varied in this plot as in Table 4.4.

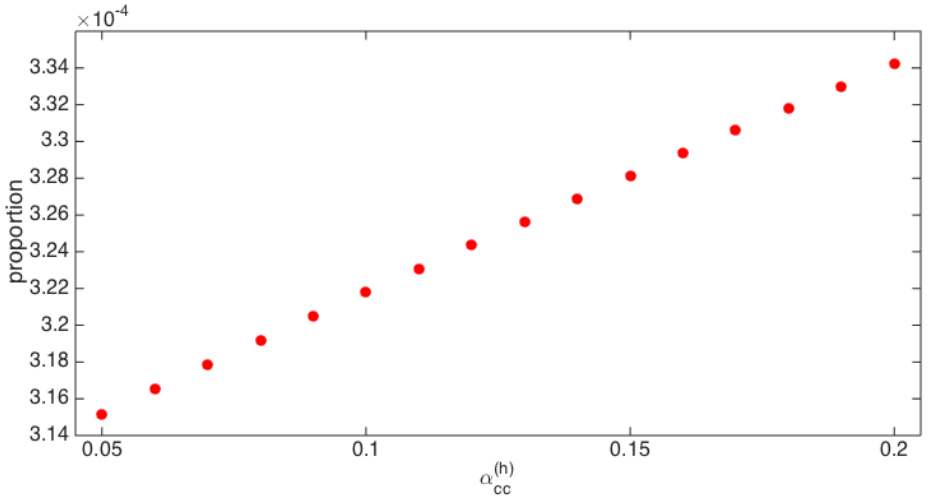


Figure 5.4: The proportion of infectious individuals in the population is recorded at the same time each year. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.4, with $\alpha_{cc}^{(h)}$ being varied.

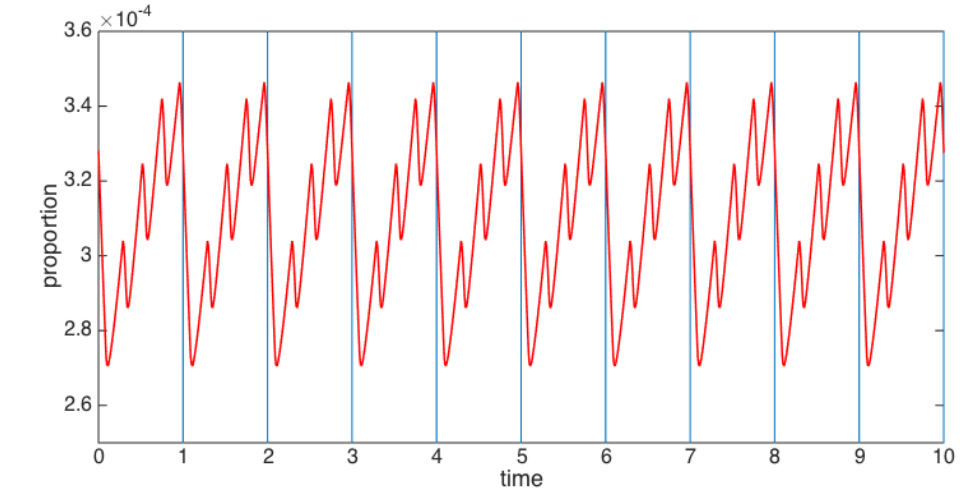


Figure 5.5: The trajectories of the proportion of infectious individuals in the population. The light blue reference lines are times that we record the proportion. The demography parameters are those summarised in Table 3.3 except b_0 is being varied, and the measles parameters as in Table 4.4.

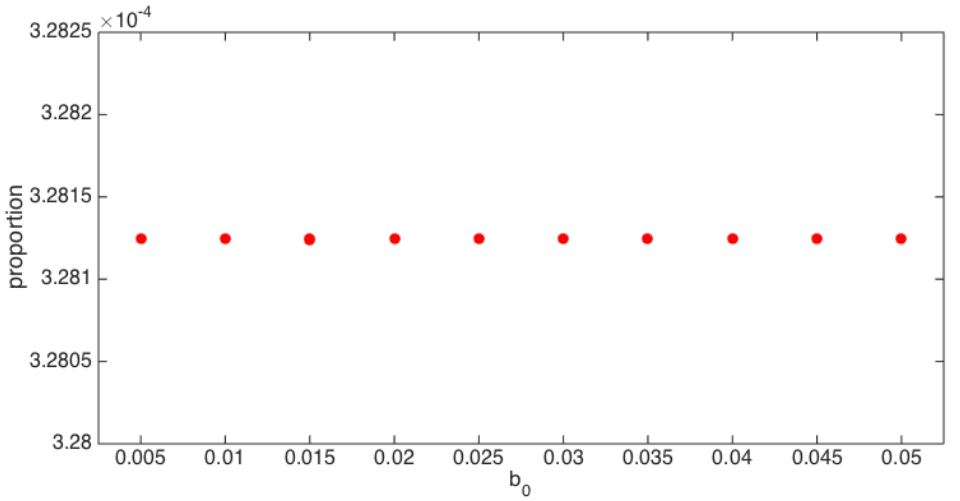


Figure 5.6: The proportion of infectious individuals in the population is recorded at the same time each year. The demography parameters are those summarised in Table 3.3 except b_0 is being varied, and the measles parameters as in Table 4.4.

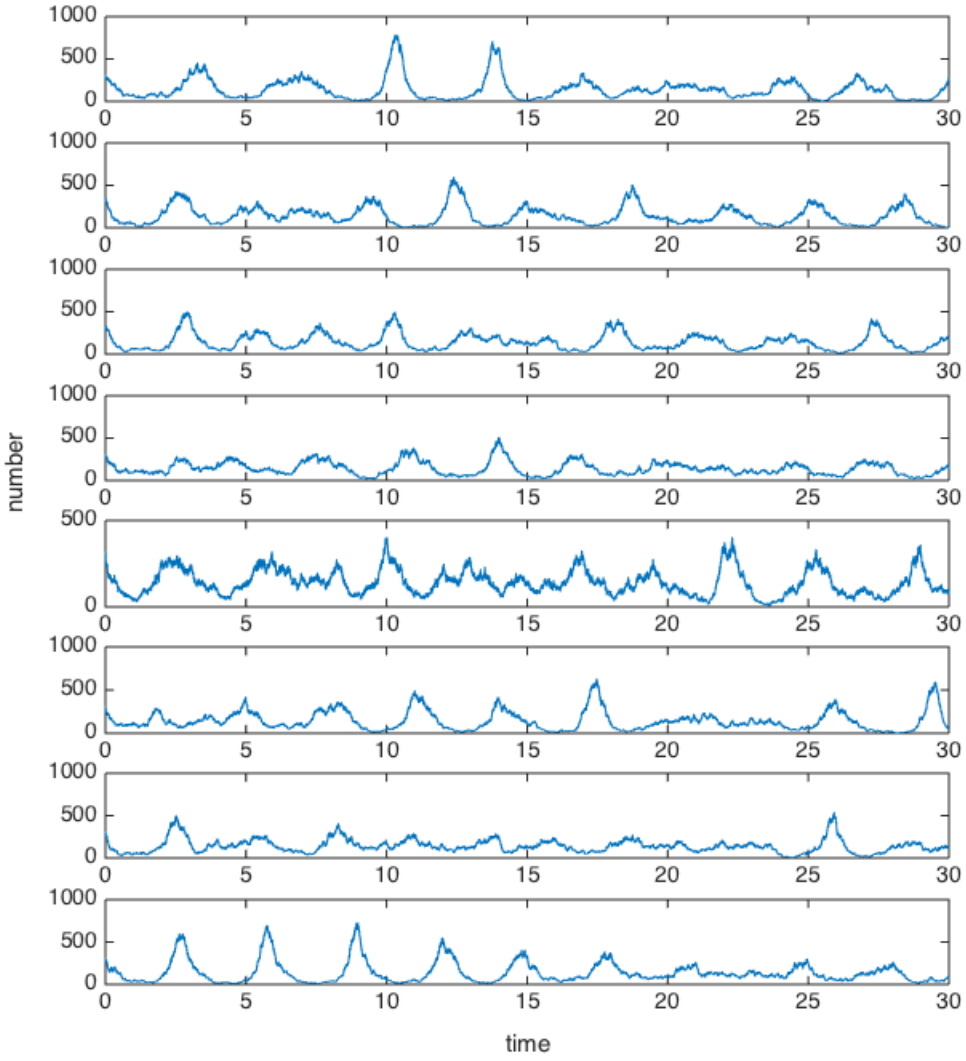


Figure 5.7: Eight realisations that persist in the population over 30 years plotted on separate axes, each with peaks approximately 2-3 years apart. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.4.

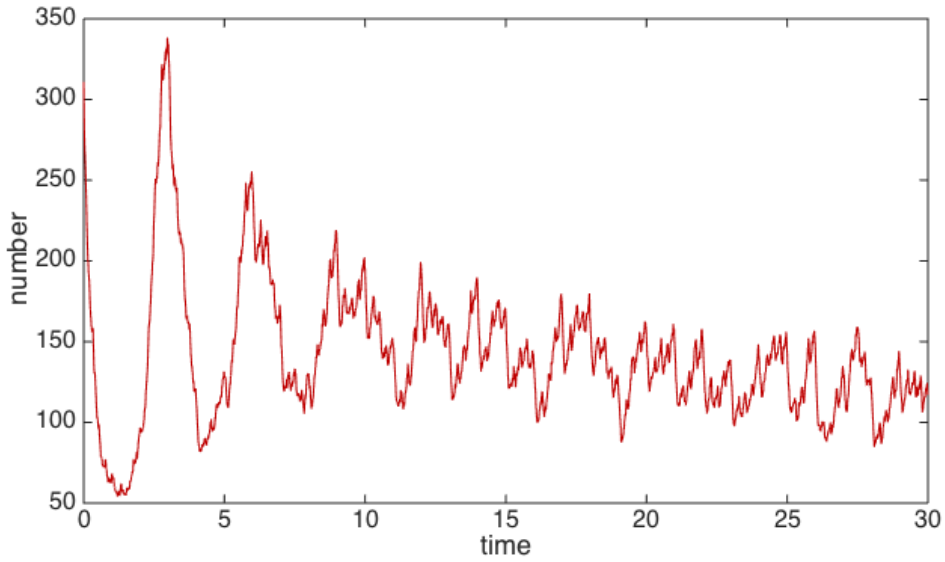


Figure 5.8: The average of thirty-nine realisations. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.4.

5.2 Outbreak of a flu-like disease

In the model developed in this thesis none of the demographic events depend on the status of the disease in the individual. In particular, the rate of death of an individual is independent of their disease status. However, many diseases do actually affect the death rate of an individual that is infected. For example, in communities with a lack of healthcare, measles can kill up to 10% of the individuals who contract the disease [36].

Of particular interest is the scenario where there is an outbreak of a flu-like disease in a city, where this disease kills some proportion of individuals while they are infectious. Let m_r be the proportion of individuals that die while infectious, due to the disease, and call this the mortality rate[1]. Then let d^* be the death (due to infection) rate in infectious individuals. As the time until the next death event is exponentially distributed,

$$\int_0^{\frac{1}{\gamma}} d^* e^{-d^* x} \mathrm{d}x = m_r.$$

This equation says that over the mean infectious period, the probability of dying due to the disease is m_r for a given rate d^*. Hence, the death rate due to the disease is given by,

$$d^* = -\gamma \ln (1 - m_r).$$

Hence an infectious adult dies at rate $d_A + d^*$, and an infectious child dies at a rate of $d_C + d^*$. In a city, it is not realistic to introduce a disease into an entirely susceptible population. Hence, in this scenario, the disease is run until the proportion of infectious individuals has reached equilibrium. Once in equilibrium some percentage of each household configuration with recovered individuals are turned entirely susceptible. This attempts to replicate the proportion of individ-

[1]Note that this is not a 'per time' rate, but a percentage.

uals susceptible to a new strain of flu when it breaks out in the population. From then, we evaluate the model over 20 years to learn how the population structure is impacted and how long until the household structure resettles. We investigate three percentage of loss of immunity: 10%, 30% and 100%.

We consider three possible scenarios: where only children can die due to the disease (pictured in Figure 5.9), only adults can die due to the disease (pictured in Figure 5.10), and the scenario where both adults and children can die due to the disease (pictured in Figure 5.11). To save computational time when evaluating the process, we utilise the deterministic model. This is because a large number of transmission events occur and hence the stochastic simulation method is very computationally expensive. This, however, means we cannot feasibly investigate the stochastic properties of the process.

The increased mortality rates that we consider are $m_r \in \{0, 0.1, 0.2, 0.3, 0.4\}$. Although the largest values here are somewhat extreme, we investigate the effect that these increased mortality rates have on the population.

5.2.1 Increased mortality in children while infectious

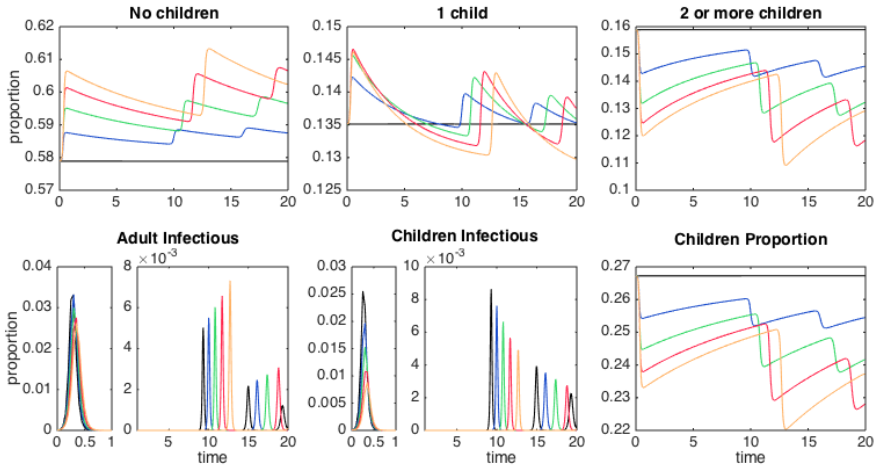
In the case of increased mortality rate during infection in children, as seen in Figure 5.9, we see persisting peaks of infection when only 10% of recovered individuals are returning to the susceptible pool, whereas when the percentage is higher, we see no new peaks of infection within 20 years. This is due to the population being depleted of susceptible individuals in the latter cases, whereas, in the 10% case, there is a smaller proportion of susceptibles, which is akin to when the model is approaching equilibrium, meaning that there is a higher proportion of infectious individuals between peaks and we are less likely to see fade out. These renewed spikes in infection in the 10% case cause fluctuations in the proportions of each household type, which is not seen when 30% or 100%

of recovered individuals are returned to the susceptible class.

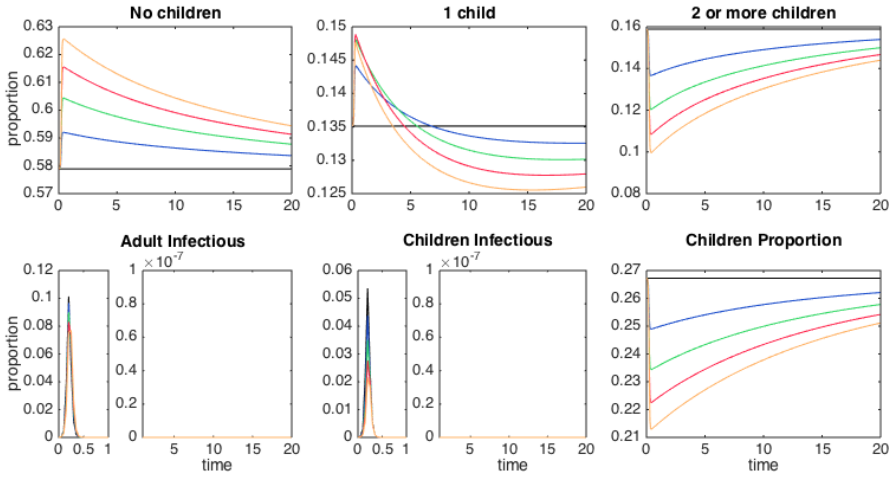
If utilising the stochastic model, the disease would almost definitely fade out during the approximately 10 years following the initial peak because the proportion of infectious individuals is very low. This would mean that the fluctuations that occur for every new peak of infection in the 10% case would actually smooth out, much as it does in the other two cases.

The proportion of children in the population decreases as the mortality rate during infection increases which is due to the increasing number of children that are dying throughout the process. Because, in the 10% case, there are recurring peaks of infection, the proportion of children in the population decreases rapidly during each new outbreak of the disease as there are more individuals with the disease and hence the amount of death increases as well.

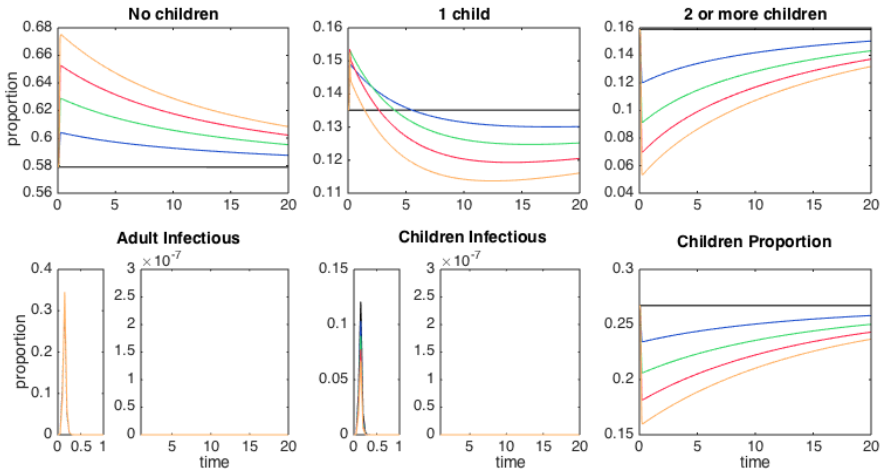
We also see that the proportion of households with 2 or more children decrease in proportion, whereas the proportion of households with less children actually increase initially. However, in the long term, even the proportion of households with 1 child eventually falls below that in the case where there is no increased mortality during the infectious period. This is due to the children in these households becoming adults and moving out of home, and also these families having more children, as evidenced by the households with 2 or more children increasing in proportion over time, which are getting closer to the proportion in that of the case where there is no increased mortality rate in infectious children.



(a) 10% of households with recovered individuals returned to susceptible class



(b) 30% of households with recovered individuals returned to susceptible class



(c) 100% of households with recovered individuals returned to susceptible class

Figure 5.9: The trajectories of proportions of infectious children and adults, and the proportion of children in the population, when only **children** have an increased mortality rate during infectious periods. Also pictured is the proportion of households with no children, one child, and 2 or more children. The mortality rates plotted are $m_r = 0$ (black), $m_r = 0.1$ (blue), $m_r = 0.2$ (green), $m_r = 0.3$ (red) and $m_r = 0.4$ (gold). The demography parameters are those summarised in Table 3.3, and the influenza parameters as in Table 4.3.

5.2.2 Increased mortality in adults while infectious

When it is adults only that have an increased probability of dying during an infectious period, persistent infection is seen when only 10% of recovered individuals are returned to the susceptible class, and also in the 100% case too for mortality rates $m_r = 0.3, 0.4$, but there is no persistence over 20 years in the 30% scenario as seen in Figure 5.10.

In this scenario, where adults are the only individuals that are at risk of a higher rate of death during an infectious period, the proportion of children increases initially to coincide with the initial peak in infectious adults and children. As adults are the only ones that are affected by the increased mortality, the proportion of adults decreases as they are removed from the system.

In the case of 10% of the recovered households becoming susceptible again, the proportion of children is less predictable. Considering just the two largest mortality rates plotted, $m_r = 0.3, 0.4$, we see that the trajectory corresponding to $m_r = 0.3$ gives a larger proportion of children compared to $m_r = 0.4$; or contrapositive: $m_r = 0.3$ yields a lower proportion of adults compared to $m_r = 0.4$. Further, we see that the trajectories for $m_r = 0.2, 0.3$ are very similar, meaning that perhaps there is some maximum impact on adults for $m_r \in [0.2, 0.3]$.

Initially, we see a decrease in all household types plotted, meaning that the proportion of empty households is increasing, which is due to the large number of single adult households being removed from the system. Over time, we see that all the household types tend to approach the trajectory of $m_r = 0$, except in the 10% case, which is likely due to the persistence of the disease during the 20 years.

Because the model was evaluated deterministically, the disease will not fade out. By determining the lowest proportion of infectious individuals in the population—denoted, say, L_p—we can give an approximate population size for the disease to

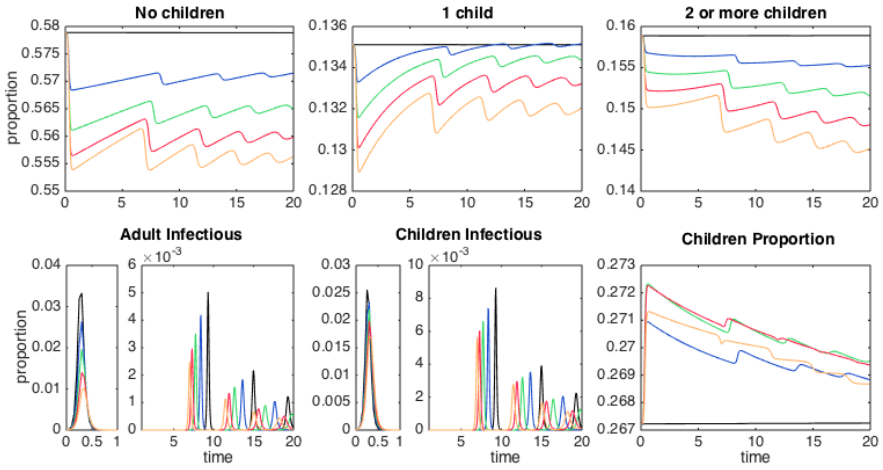
persist if we assume,

$$L_p N \geq 1.$$

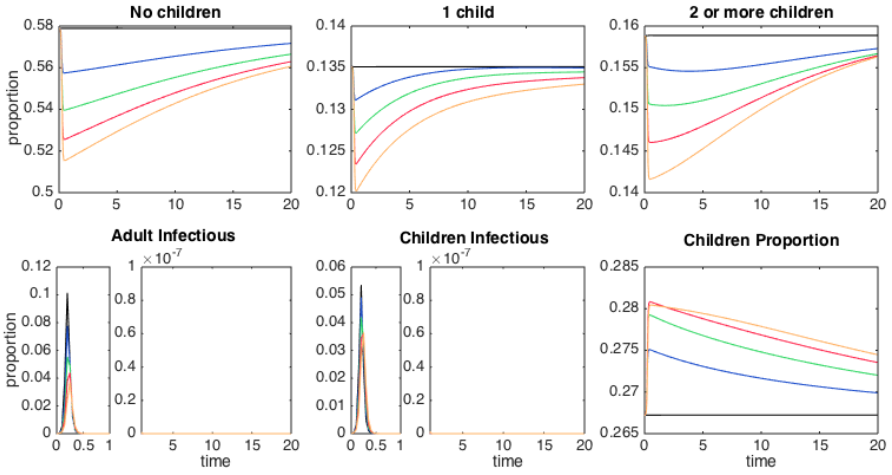
This says that if the population size, N, is large enough such that at least one individual has the disease, then the disease may persist. This is, of course, simply a 'ball-park figure' because the stochastic model has not been used here.

In the 10% case, for mortality rate $m_r = 0.4$, the lowest proportion of infectious adults is approximately 3.03×10^{-8} and the lowest proportion of infectious children is approximately 6.38×10^{-8}. Hence, if the population of adults in a city exceeded 33 million people, and the population of children exceeded 16 million, persistence of the disease may be observed. As of the end of 2016, no city has this many inhabitants but perhaps in the future this result may be more meaningful. However, for $m_r < 0.4$ (and for all mortality rates in the 30% and 100% cases) the lowest proportions of infectious children and adults are even smaller and as such require a significantly larger population again to see disease persist, which is very unlikely in any city around the world, let alone an Australian one.

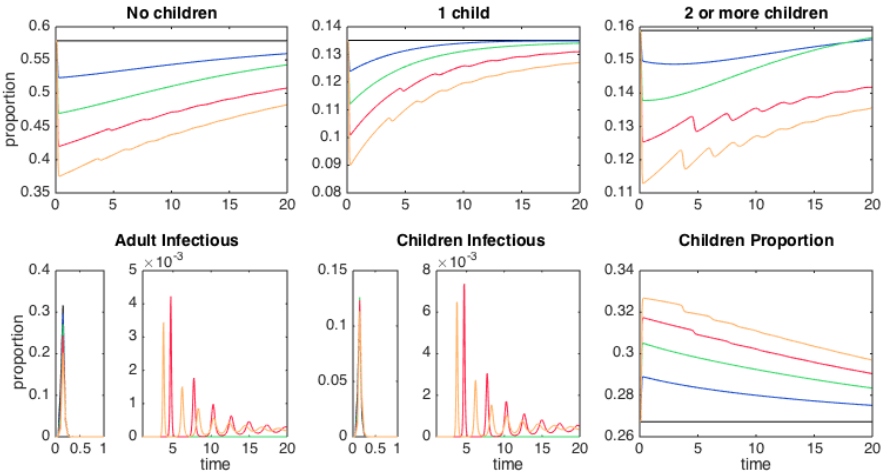
Lastly, in the 100% case, for the mortality rate $m_r = 0.4$, the lowest proportion of infectious adults and children is 6.14×10^{-7} and 1.38×10^{-6}, respectively. This means that in a population with approximately 1.63×10^6 adults and 7.25×10^5 children, persistence of the disease may be observed. There are many cities in the world where such a population size is seen. In Australia cities such as Sydney and Melbourne have the largest population size, in excess of 4 million people, meaning that persistence may be observed in them if a flu-like disease that kills 40% of adults that contract the disease. For all other mortality rates, we see a much lower proportion of infectious individuals such that persistence is not likely to be seen in any current city in the world.



(a) 10% of households with recovered individuals returned to susceptible class



(b) 30% of households with recovered individuals returned to susceptible class



(c) 100% of households with recovered individuals returned to susceptible class

Figure 5.10: The trajectories of proportions of infectious children and adults, and the proportion of children in the population, when only **adults** have an increased mortality rate during infectious periods. Also pictured is the proportion of households with no children, one child, and 2 or more children. The mortality rates plotted are $m_r = 0$ (black), $m_r = 0.1$ (blue), $m_r = 0.2$ (green), $m_r = 0.3$ (red) and $m_r = 0.4$ (gold). The demography parameters are those summarised in Table 3.3, and the influenza parameters as in Table 4.3.

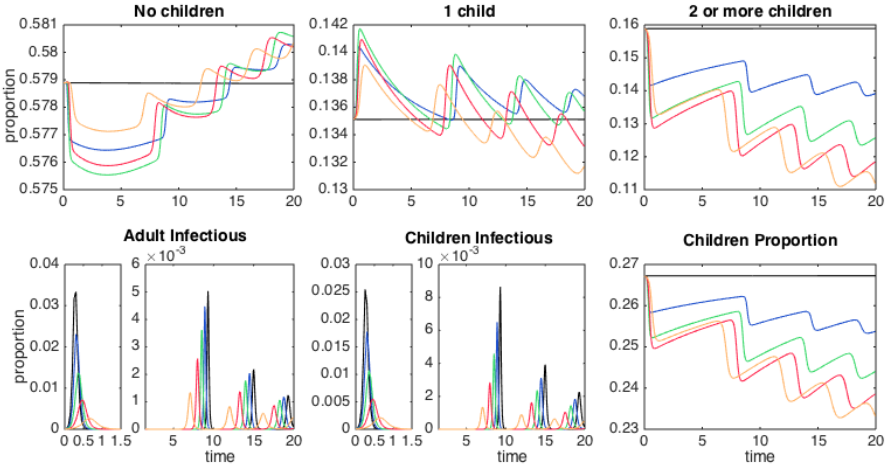
5.2.3 Increased mortality in both children and adults while infectious

In the scenario where both children and adults have increased mortality during the period of infectiousness, the trajectories of the proportions of household types appear to be a combination of the 'children only' and 'adults only' dynamics, as seen in Figure 5.11. Only when 10% of the recovered individuals are returned to the susceptible class do we see any subsequent outbreaks of the disease after the initial peak within 20 years. Further, it is very possible that these subsequent outbreaks could appear if a stochastic model was used to evaluate this scenario too, because for $m_r = 0.3$, if the population exceeds 1.2 million people, or for $m_r = 0.4$ if the population exceeds even as little as a few hundred people, then stochastic realisations may persist.

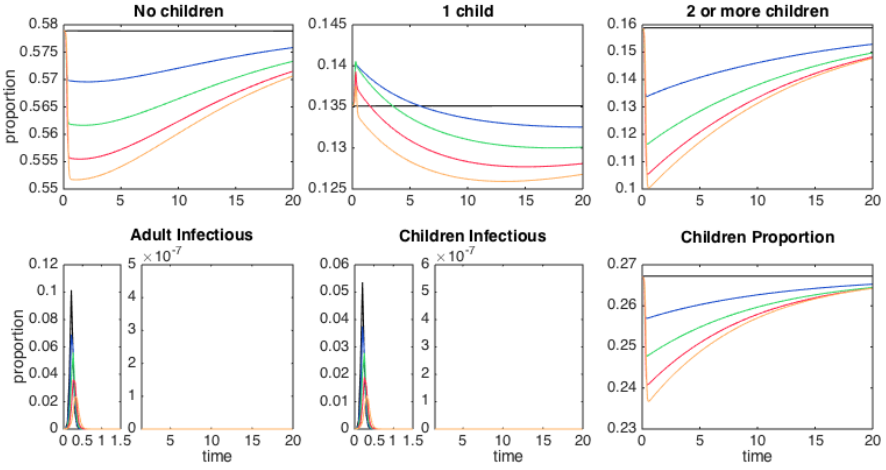
In this scenario, where both adults and children are at risk of an increased mortality rate during their infectious period, the 10% case suggests that the proportion of households with no children will increase beyond the associated proportion where there is no increased mortality rate, albeit after a small initial drop. This is due to the death of adults and children, reducing the size of each household. We also see oscillations about the base mortality rate $m_r = 0$ for households with 1 child, similar to that seen in the scenario where only children are affected by increased mortality rate.

The increased death rate in children means less adults are being created and hence less children are being born, reducing the susceptible pool which results in less infection. Over 20 years, we see a decrease in the proportion of children (except in the case of all recovered individuals returning to the susceptible class) which could be due to the removal of children as well as adults when a single adult dies (recall that the entire household is removed). Hence, while it appears that children are dying more than adults because the proportion of children is

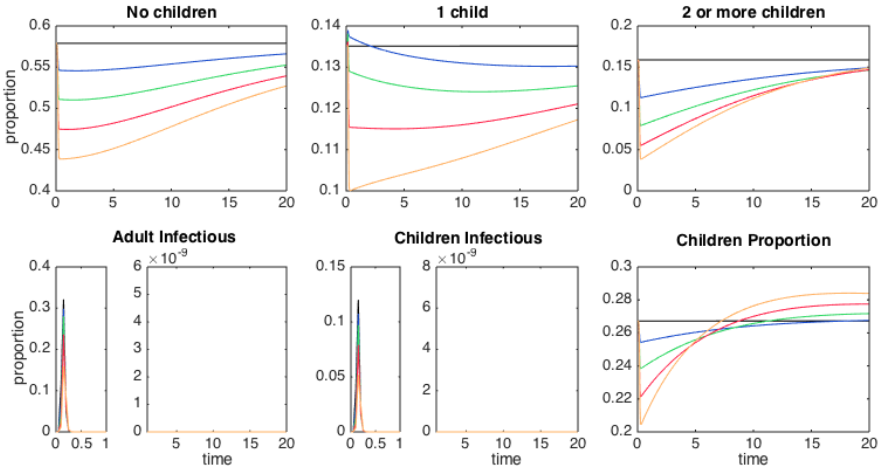
decreasing, some proportion of these are in fact just being removed from the system. This is a feature of our model and a more realistic formulation may show contrary results in this regard.



(a) 10% of households with recovered individuals returned to susceptible class



(b) 30% of households with recovered individuals returned to susceptible class



(c) 100% of households with recovered individuals returned to susceptible class

Figure 5.11: The trajectories of proportions of infectious children and adults, and the proportion of children in the population, when both **adults and children** have an increased mortality rate during infectious periods. Also pictured is the proportion of households with no children, one child, and 2 or more children. The mortality rates plotted are $m_r = 0$ (black), $m_r = 0.1$ (blue), $m_r = 0.2$ (green), $m_r = 0.3$ (red) and $m_r = 0.4$ (gold). The demography parameters are those summarised in Table 3.3, and the influenza parameters as in Table 4.3.

5.3 Epidemic fadeout

An integral part of investigating endemic diseases is determining why and how they persist in communities. Two large factors in the persistence of diseases, particularly for disease parameters akin to measles, is the birth rate and how large the population is. The latter has been explored in detail in the form of the critical community size [12, 14]. When a population is sufficiently large, the disease will persist without external introductions more often than not. On the other hand, populations of size smaller than the critical community size will have the disease fade out much more frequently. Here, we contribute to the discussion by investigating the critical community size for measles in our household model with both disease and demography.

In order to investigate this phenomenon, we define epidemic fadeout to occur when the number of infectious individuals is zero. The critical community size is largely accepted to be between 250-400 thousand individuals [14].We run simulations for a number of population sizes and count the number of realisations that fadeout and hence make deductions on the critical community size.

5.3.1 Non-forced model

The question of how to set up the simulations sensibly arises. The naive approach to intialising the system is by introducing one (or at least a very small number of) infectious individual(s) into the population, but if we were to measure how many realisations fade out after the initial peak, we would find that most realisations die out as the initial peak depletes so many of the susceptible individuals and births have not had the time to replenish them. Hence, it makes sense to let the disease run its course for longer in order to 'mimic' the disease persisting in the population for some time. To do this, the deterministic model is run over a long period until the number of infectious individuals is in equilibrium. The final state

is a fixed point and this becomes the initial condition for each simulation.

We use a homogeneous model with demography (where births and deaths are not linked, but have the same birth and death rate) to compare with the household model. Recall that b_j is the birth rate into a household of type $(2, j)$, d_A is the rate of death for an adult and d_C is the rate of death for a child. Then we define the total births and deaths, respectively, for the household model in equilibrium,

$$b_{HH} = b_0 m_{(2,0)} + b_1 m_{(2,1)} + b_2 m_{(2,2)} + b_3 m_{(2,3)},$$
$$d_{HH} = d_A \sum_{i=1}^{2} \sum_{j=0}^{4} i m_{(i,j)} + d_C \sum_{i=1}^{2} \sum_{j=1}^{4} j m_{(i,j)}.$$

The homogeneous model is calibrated by minimising the difference between the total births, the total deaths, and the total amount of infection in equilibrium, in the two models. That is, the homogeneous model was fitted to the household model by minimising,

$$\min_{b,\beta} \frac{1}{2} \left(|b_{HH} - b\left(N - S(b,\beta)\right)| + |d_{HH} - bN| \right) + |I_{HH} - I_{homo}(b,\beta)|,$$

where I_{HH} is the total number of infectious individuals in the household model at equilibrium and I_{homo} is the total number of infectious individuals in the homogeneous model at equilibrium. Recall that in this homogeneous model, the birth and death rates are assumed to be equal, but are not coupled.

In Figure 5.12, we see that in the household model over 15 years, infection tends to die out, even in large sized population that correspond to where the homogeneous model sees persistence of the disease. This suggests that the household model is more unstable and has higher variability. Also note that the proportions pictured fluctuate. This is because we have simply plotted the proportion of realisations that have 0 infectious individuals at a given time; however, because

we utilise the SEIR structure, it possible for there to be 0 infectious individuals but more than 0 exposed individuals which consequently increases the number of infectious individuals to be greater than 0 once more, which accounts for the very slight fluctuations.

Of course, we expect to see the proportion of realisations that fade out to be above 0.5 to concretely talk about the critical community size but due to the household model seemingly having a higher variability than the homogeneous, such proportions are not found after 15 years. Within the first 5 years in the household model, though, there is an interval in time where the proportion of realisations that have faded out flattens; this can be seen somewhat prominently for households between 250 and 400 thousand, and during this period the proportions are well above 0.5. This can be interpreted as persistence of the disease over 5 years is likely more often than not. As our model has such high variability, the first five years of the realisations may hold the most relevant information about the critical community size. This would then suggests that the household model without seasonal forcing has a critical community size at about 250 thousand households, which is approximately between 450 - 500 thousand individuals, which is larger than estimated in Bartlett [12], yet similar. Our results assume no external importation of measles over the course of the 15 years studied, whereas for the data in [12] it cannot be assumed that there was no importation of measles from surrounding cities and towns.

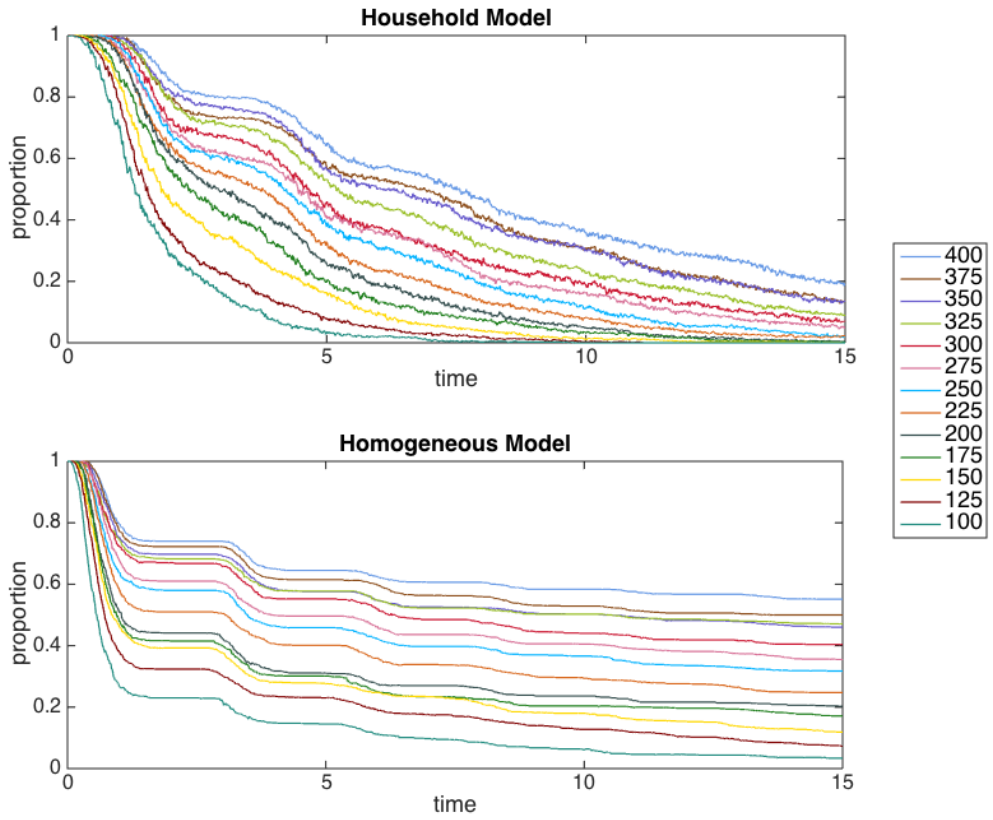


Figure 5.12: The proportion of 800 realisations that fade out in the household and fitted homogeneous model, for different population sizes, where the legend denotes the number of households in thousands. The number of individuals in the population for a given household size is approximately just below double the number of households. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.3.

5.3.2 Seasonally-forced model

In the seasonally-forced model, the deterministic model does not reach equilibrium, instead oscillating periodically. Hence it is more difficult to determine a reasonable initial condition. We have chosen to simply pick a time corresponding to the period after the number of infectious individuals settles into the repeating oscillatory motion.

Calibrating the household model to the homogeneous model here is also more difficult to do because of the oscillations in the level of infection. To get parameters

calibrated to the household model, we minimise the following,

$$\min_{b,\beta_s,\beta_h} |b_{HH} - b(N - S(b,\beta_s,\beta_h))| + |d_{HH} - bN| \\ + |p^{(h)}_{HH} - p^{(h)}_{homo}(b,\beta_s,\beta_h)| + |p^{(l)}_{HH} - p^{(l)}_{homo}(b,\beta_s,\beta_h)|,$$

where β_s and β_h are the transmission parameters during the school term and holidays, respectively, b_{HH} and d_{HH} are defined as in Section 5.3.1, $p^{(h)}_{HH}$ and $p^{(h)}_{homo}$ are the highest points in the oscillations that are seen in the long term in the household and homogeneous models, respectively, and $p^{(l)}_{HH}$ and $p^{(l)}_{homo}$ are the lowest points in the oscillations that are seen in the long term in the household and homogeneous models, respectively. This means that we fit the homogeneous by attempting to get a similar amount of births and deaths in each model and also have the level of infection fluctuating between the same two values.

As seen in Figure 5.13, the household model again fades out more often than in the homogeneous case. Further, in the homogeneous model we do not see the flattening out of the proportions as we did in the non-forced model. What we do see however, is that in comparison to the non-forced model, the seasonal forcing appears to increase the number of realisations that fade out.

Although there are no significant jumps in the homogeneous model, there appears to be a large jump between the 225 and 250 thousand households in the household model. This suggests that populations of larger than 225 thousand households could see a large difference in fadeout, especially within the first five years. Over the course of 15 years, however, the difference between the proportions decreases. However, we see an interval within the first 5 years where the proportion flattens out somewhat at about 0.5 for populations with 250 thousand households. In fact, we see this flattening out of the proportion in most of the populations with household sizes greater or equal to 250 thousand, and to a lesser extent in populations with 200 and 225 thousand households too.

Just as in the non-forced model, this interval where the proportion of fadeout temporarily stays constant could suggest that the critical community size is in fact between 225 and 250 thousand households, or approximately 375 to 500 thousand individuals, which is closer to the critical community size presented by Bartlett [12], but still higher. Our estimates, in both the seasonally forced model and the non-forced, can be explained, at least in part, by our parameterisation. Bartlett was interested in learning about the critical community size in England in the mid 1900s when the birth rate was considerably higher than in modern Australia. This means that our model produces less susceptible children and hence there is a greater chance that the disease will die out, and that the critical community size is larger than in [12, 14].

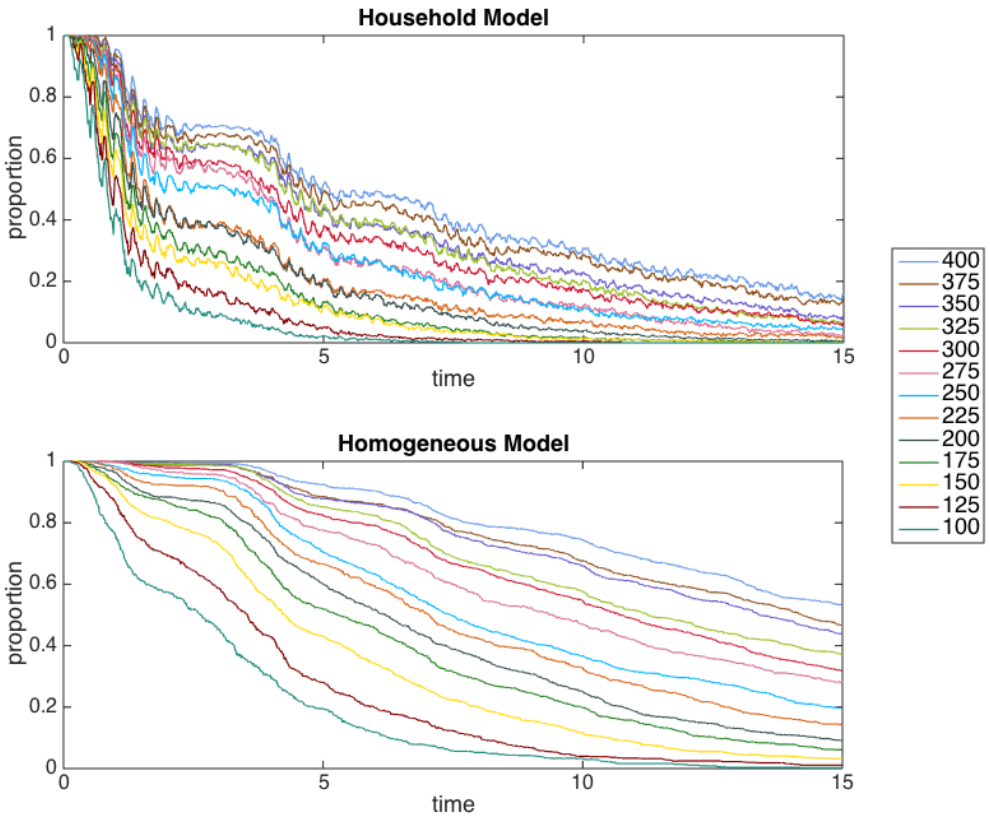


Figure 5.13: The proportion of 900 realisations that fade out in the household and fitted homogeneous model with **seasonal forcing**, for different population sizes, where the legend denotes the number of households in thousands. The number of individuals in the population for a given household size is approximately just below double the number of households. The demography parameters are those summarised in Table 3.3, and the measles parameters as in Table 4.4.

Chapter 6

Discussion

In this thesis, a household model for endemic diseases was developed as a CTMC. A changing population structure is assumed to contribute significantly to the disease dynamics, and so demography dynamics are modelled as a CTMC before extending the model to include disease events too. We also define a household model with seasonal forcing.

Two methods are used to evaluate the dynamics of the model: simulations using the so called Gillespie's algorithm [22] and a deterministic approximation. The deterministic model is derived in Section 3.3.2 and Appendix B, and numerical differential equation solvers in Matlab, such as `ode45` and `ode23s`, are utilised to evaluate the dynamics of the model.

In addition to the development of the household model, this thesis has investigated the relationship between population size and the probability of the disease fading out by running simulations and recording the proportion of realisations that fade out over 15 years. This study was complemented by an investigation into how an outbreak of a flu-like disease, where infectious individuals have an increased mortality rate, may impact the population structure.

In this section, we discuss the properties that the household model possesses in comparison to the homogeneous model and make further comparisons between

the seasonally-forced and non-forced model. We also detail the important model design choices, the limitations, the strengths and weaknesses of the model, and the improvements that could be made.

6.1 Results and model discussion

Like any model, the household models in this thesis have strengths and weaknesses. The models also have a number of features that are not found in other similar models, which we detail below. We discuss some of the properties of the household model and how it compares to the homogeneous model, and discuss some of the most important design choices, assumptions and limitations.

Both the household model with and without seasonal forcing appear to be less stable in comparison to their homogeneous counterparts. The disease tends to fade out more often in the household model which indicates more variability. This is due to how far the stochastic realisations can move away from the mean trajectory, which is the solution to the deterministic approximation. This is further evidenced in the seasonally-forced model where the periodicity of a single realisation is approximately between 2-3 years, but when averaged the realisations become out of sync, and hence the mean of the realisations eventually settles into an annual period, at least by the end of the 30 years studied, which is what the deterministic model yields. A more robust analysis of the stochastic period could be carried out by undertaking a spectral analysis [37].

When finding the periodicity of the stochastic seasonally forced household model, over 350 simulations were computed. However, despite the population dynamics representing South Australian census data (South Australia has a much higher population compared to what the literature says the critical community size should be), a large majority of the realisations faded out over 30 years. This behaviour was alluded to earlier when looking at the epidemic fadeout problem,

where we saw that the proportion of realisations that faded out steadily declined over 15 years, despite the homogeneous model flattening out in the non-forced model. This, in conjunction with the fact that so many realisations faded out when simulating a population with size approximating Adelaide, implies that the household model does not have a population size (at least up to 1.2 million individuals) that sees the disease persistent more than 50% of the time, over time periods longer than 10 years. We have assumed, when simulating these realisations, however, that no external disease is imported into the population, when a realistic feature would include the possibility of measles being reintroduced into the population over the 15-30 years studied. Future work could be investigating the effect of importations of measles at different time intervals and the effect this has on the persistence of measles in the population.

What we did see, however, was that within the first 5 years of stochastic simulation, the household models have persistence in the population above 50% of the time, when the number of households was greater than approximately 225-250 thousand, and for some years too, meaning that initially it appears as if the disease will persist as expected but eventually the variability takes over and reduces the number of infectious individuals until there are none. If we take 225-250 thousand households (or approximately 375-500 thousand individuals) to be the range in which the critical community size for our household models lies, then it is higher than other estimates made by Bartlett [12]. This can be explained by our demographic parameterisation, where there are less births and less children compared to in England in the mid 1900s, which is the time period that Bartlett was concerned with. Less births means less susceptible individuals and hence the disease may struggle to persist as frequently. Future work in this area includes calibrating the demography dynamics in our model to England in the 1950s and studying the effect of imported disease into the population.

Our household model relaxes many assumptions made in the homogeneous model

such as providing a more realistic population structure for endemic diseases with the addition of partitioning households and demographic events. This makes our model one of the best for investigating population dynamics throughout the duration of an endemic disease. In addition to this, because our model is able to differentiate between adults and children, we are able to investigate in detail diseases that affect adults and children separately. We have seen the difference in population structure when a flu-like disease breaks out in a population and affects exclusively adults or children. Although not included in this thesis, we can also explore the dynamics of the population structure in the case where adults and children experience different mortality rates, rather than the same, as seen in Section 5.2. This flexibility, gained by considering separately adults and children, can lead into much more complex problems, such as how vaccinations can be utilised in such a population to reduce infection.

To study the effect of vaccines in our household model, we can extend the model to include additional events such as vaccinations, where susceptible individuals are transitioned into the recovered class at some rate. By applying different rates and different strategies of distributing the vaccines we could learn even more about how disease spreads through the population. The different rates could take the form of linear or constant rates, where the associated parameters are different for children and adults. Further, how the vaccines are distributed is known to be important in vaccinating populations efficiently [38, 39, 40]. As such, some of the strategies we would consider initially include randomly allocating the vaccines uniformly to individuals across the population, randomly allocating vaccines to entire households across the population, so that an entire household is in the recovered class, and distributing vaccines to children only.

The results focus on the stable population size parameterisation but the model was also parameterised to account for a growing population. Although this thesis has not explored the effects this has on the disease dynamics, it would be

relatively easy to implement once an appropriate comparison to the equilibrium in the non-growing model has been established. That is, in Section 5.2, where a flu-like disease outbreak occurs, and in Section 5.3, where we investigate the problem of epidemic fadeout, the initial condition is a fixed point. However, in the case of a growing population, the number of households is always increasing, meaning there is no equilibrium for the process. This means that investigating epidemic fadeout in such a model may be difficult to initialise because choosing the population size to be a particular value to assess critical community size, while simultaneously attempting to start the stochastic simulation from a state that is comparable to the fixed population size model is not easy or intuitive. Preliminary investigation shows, though, that the trajectories of the proportions of each of the household types appear to settle into a steady state, so that each household type is increasing at the same rate. That is, the ratios of each of the households are fixed, and this could be utilised in future work.

There were many design choices made in the process of developing the household model. First and foremost was choosing to let the time between events be exponentially distributed. While this is not the most realistic assumption, it is one that simplifies the model and the mathematics considerably. One of the most obvious consequences of the exponentially distributed time between birth events is that as soon as a birth has occurred, another birth can potentially occur almost immediately after. However we know that there must be at least 9 months between births (except in the case of twins, triplets, etc), but our model does not encapsulate that. In order to capture this biological feature, albeit at the cost of making it even more complex, a birth event could transition a household into another household configuration that cannot undergo a birth event, and after an exponentially (or Erlang, in order to reduce the variance) distributed time with mean of 9 months, the household is returned to the configuration where they can, once more, birth another child, provided that the number of children is less than

4. Further, we could use Erlang distributions for any events where the variance of the time between events should be smaller, or more precise, than is the case when exponentially distributed.

One of the many other important design choices in developing the household model was regarding the inclusion of the empty households as opposed to defining the model without them. The inclusion of empty households easily allowed us to keep the number of households constant, or, if desired, increase the number of total households by some controlled amount via the event which we denoted as a new empty household with rate g, per household in the population. While the number of empty households does not directly impact the disease dynamics, their importance comes in another form. They are used to define the rates of immigration, couple separation, and children moving out of home, as each of the rates of these events depend on the preference function. If the empty households were not included in the model, three events would need to be redefined alternatively, as they each rely on the number of empty households in the population. These changes would further complicate the model and necessitate alternative assumptions in order to model them. This is because we have to assume that either these events rely on other household types or that they do not depend on any household type, either of which require an assumption, arguably less feasible than assuming the events do rely on the number of empty households, which is no more beneficial than simply including the household type (0, 0) in the model.

One of the biggest drawbacks to including empty households is the extra household type. One of the goals in developing the household model was to keep it as simple as practical while relaxing as many assumptions about the population as possible. Another design choice that had to be made early was to decide how many different age groups would be modelled and consequently how to define these age groups. Measles primarily affects children, and so it was decided, be-

cause our model is motivated by measles, that there would be at least two age groups: children and adults. An elderly class was considered, as they can be more susceptible to diseases, and they would definitely have different rates for death events (which would increase), births (would would become very small, if not zero), couple formation (which would decrease), couple separation (which would decrease) and migration (which would decrease). However, by adding only one extra age group, the size of the state space would become even bigger than the final choice of 981 different household configurations, which was an infeasible option due to computer memory constraints. Instead, the elderly are included in the adult age group, where children are assumed to be individuals that are dependent on at least one parent; that is, an individual becomes an adult once they move out of home.

One of the main restrictions of the model is the number of children within a household. The model assumes that no more than 4 children can live in one household. This assumption was made as a simplification, because if the maximum number of children increased to 5, the number of household configurations would exceed a total of 1700 different household configurations. This would not add any extra realism to the model as Australian census data shows a very small percentage of the population in households with more than 4 children. It is also assumed that there are no group households in our model; that is, three or more adults cannot live together. This assumption was made also to simplify the model, but was justified in the sense that a very small proportion of Australian homes classify as a group household.

Two of the main assumptions made with regard to the population dynamics is in the birth and couple formation events. We assume that births only occur in households with two adults in them, disregarding the births that occur in households with single parents, and also partners who do not live together. This assumption was to simplify the model, but as births are a real driving factor

in endemic diseases, the model could be extended to include birth rates for households with only one adult in them, although the rates for these birth events would be significantly lower than those in households with two adults.

Couple formation requires that one of the two households that are joining together be a lone adult household, $(1, 0)$. This assumption was made to simplify the event, because by ensuring that one of the households does not have any children, the addition of children in the two households never exceeds 4. If we do not make this assumption, then it is possible to have more children in a household than our model allows. This is a problem which has many solutions, such as simply truncating the number of children in a household at 4; limiting the couple formation of single adult families to the cases where the total number of children is no more than 4 (that is, a household type $(1, 2)$ can combine with $(1, 0)$, $(1, 1)$, or $(1, 2)$ but not $(1, 3)$ or $(1, 4)$). The consequences of the former solution is that the number of children in the population is not being conserved in the event. To counteract this lack of conservation of children, the excess children could be distributed to other households, or they could create their own households, or some combination of each, but this increase the complexity of the model and would require making even more assumptions such as that the excess children are old enough to form their own single household, for example.

The other solution, where we limit couple formation to single adult households where the total number of children is no more than 4, is also not feasible for two particular reasons. The first is that by restricting the types of households that can join in the couple formation event, we are effectively making just as many assumptions as if we simply assume that a lone adult has to be one of the household types involved in a couple formation. The second reason this solution is infeasible is that this is a complication in the model that is much more difficult

to implement. Recall the form of the differential equations,

$$\begin{aligned}\frac{d\boldsymbol{x}}{dt} = \Big[& F_1 + f\left(x^{(0,0)}\right)(F_2 + F_4) + \left(1 - f\left(x^{(0,0)}\right)\right) F_3 \\ & + \theta_a(\boldsymbol{x}) F_5 + \theta_c(\boldsymbol{x}) F_6 \\ & + x^{(1,0)}_{\mathbf{e}_1} F_S + x^{(1,0)}_{\mathbf{e}_2} F_E + x^{(1,0)}_{\mathbf{e}_3} F_I + x^{(1,0)}_{\mathbf{0}} F_R \Big] \boldsymbol{x} + G.\end{aligned} \tag{6.1}$$

The couple formation event is captured within the terms,

$$x^{(1,0)}_{\mathbf{e}_1} F_S + x^{(1,0)}_{\mathbf{e}_2} F_E + x^{(1,0)}_{\mathbf{e}_3} F_I + x^{(1,0)}_{\mathbf{0}} F_R, \tag{6.2}$$

and for each new possible configuration that is allowed to form a couple, there is an extra term, making the system of differential equations needlessly complex. Hence, we assume that a household of type $(1, 0)$ has to be one of the household types in the couple formation event so that there are only four terms corresponding to this type of event, as displayed in Equation (6.2).

Another way that could be used to further simplify the model is to alter the death events. For diseases like measles and influenza, the latent and infectious period is short relative to diseases such as HIV, and so it could be argued that the rate of death occurring in the latent and infectious period is effectively negligible because of the vast difference in the time scales of population dynamics compared to disease dynamics, and so death events are only defined for individuals in the susceptible and recovered classes. However, the problem with such a simplification in the model is that it depends on the disease that is of interest, and also depends on what questions are being investigated. For example, with such a simplification in the household model presented in this thesis, we could not have investigated the effect of an outbreak of a flu-like disease in a population where individuals are an increased risk to death during the infectious period.

6.2 Method discussion

In addition to the design choices made for the household model, many methods of analysis were attempted to make evaluating the model less computationally expensive. Simulating realisations of the trajectories of infectious individuals was very slow and so the questions and topics of interest that required a lot of simulation, such as repeating the increased mortality rate in an outbreak of a flu-like disease except with stochastic effects, were left unexplored. In order to use the household model to provide an answer to these questions and topics of interest, better methods of simulation need to be used.

One particular method that we attempted to implement, to approximate exact realisations, was the *tau leap* method [41]. This is an algorithm where instead of exactly simulating the realisations for the process, a realisation is approximated by stocastically determining how many of each event occurs over a time period of τ, where the number of events is a Poisson distributed random variable. Ideally, τ is relatively small [41]. However, the problem with using this method for our particular household model is that there are so many events and it is required to generate just as many random variables. Also, there are two distinct time scales (demography and disease), and so the value of τ must be very small to capture the disease dynamics. Because the tau leap method approximates the number of events over a time period, τ , the rates of these events are given by the rate at the start of the interval we are approximating over, which means that if τ is too large then the approximation is likely to be inaccurate as the state of the system does not update throughout the interval, but only at the end when the state space may have already reached an unrealistic value; for example, if τ is too large and the population has 20 infectious individuals, the method may approximate the number of recovery events to be, say, 50, leaving the total number of infectious individuals below zero. The required value of τ to get approximations that are

comparable to the exact realisations using Gillespie's algorithm [22] are so small that the computational time to run such approximations is of the same order as the exact simulations. There are methods, however, to overcome this problem, which could be implemented in future, such as using a binomial distribution instead of a Poisson distribution [42].

Although this approximation technique ultimately did not improve the efficiency of our model, other methods of approximation may be of more use. As there are two time scales, it may be of advantage to consider the events in each time scale independently because the value of τ does not need to be very small for a good approximation to the demographic dynamics. Then it may be feasible to exactly simulate only the disease dynamics because there are a lot less events and hence the rate vector is much smaller, because the demographic dynamics are being approximated, and so less needs to be recalculated after each transition.

Another method to improve the efficiency that was not attempted but could be investigated in future work is converting the current model into a hybrid model, where the demographic dynamics are deterministic. We saw in Chapter 3 that each realisation of the demographic dynamics do not venture far from the deterministic trajectory, meaning that stochastic effects in the demographic dynamics are not as significant as the disease dynamics. It was noted when determining the parameters for the demographic dynamics that the aim of including them was to simply provide some movement between households, rather than developing a robust model for the populations changing structure. Hence, this idea of a hybrid model could potentially be a vast improvement in terms of computational efficiency, as there would be a much smaller rate vector that needs to be updated after each transition.

Another technique we tried to apply utilised a dependency graph, which is a function that maps what household configuration each event depends on [43]. That is, it is a data structure that controls which elements of the rate vector

to change when a particular event occurs. Ideally, utilising a dependency graph means that instead of updating the entire rate vector, a much smaller number of components of the rate vector need to be updated at each time step, which would considerably decrease the computational time [43]. To compute this dependency graph, we must find all the household configurations that each event depends on which is quite difficult for a model with so many household configurations and events. In addition to creating this, we need to create another graph that tells us what household configurations are effected by an event, which can be easily computed using the stoichiometric matrix. Creating the dependency graph is difficult but achievable. However, it was much more difficult to implement than initially thought due to the complexity of the model. Further work is required to implement these graphs, in order to simulate more efficiently.

It was not only the simulations that proved difficult to compute. Many results required an equilibrium to be computed for the deterministic model, which can be found by solving the differential equations over very long time periods until the rate of change is 0 in each differential equation. This is, however, a naive approach and a technique that could be used to evaluate the proportion of infectious individuals in equilibrium for a parameter set without solving the differential equations over such long periods of time would be desirable.

Recall the form of the differential equations, Equation (6.1). Initially, we tried a number of search algorithms, such as the bisection method [44], to solve these but due to the non-linear form of the differential equations, regular search methods were unsuccessful. Although not actually of the exact same form, our differential equations shared a passing resemblance to the Riccati Equations, which are a specific form of first-order differential equations that are quadratic in their unknown term [45]. As such, we attempted to utilise, in some way, a number of the methods for solving such equations but they were not easily adaptable to our household model. In order to find new ways to compute the number of

infectious individuals in equilibrium, we should attempt to extend some of the algorithms in [46, 45] to be of use to the household model, or develop an entirely new method.

6.3 Conclusion

Throughout this thesis, we have developed a household model that includes demographic events in addition to SEIR disease dynamics, both with and without seasonal forcing. The parameters associated with the demographic events are fitted to a stable population where there is a fixed number of households, and also to a growing population, each of which are based on census data in South Australia. This model is evaluated stochastically using the so-called Gillespie's algorithm [22] and deterministically utilising theory proved by Kurtz [23]. The deterministic model is derived as a system of differential equations as a matrix equation.

For measles-like disease parameters, we find that the period of the non-seasonal forced model is approximately two years, which matched very well to the established biennial periodicity [2]. In the model with seasonal forcing, the periodicity is more complicated, with an annual period in the deterministic, and approximately 2-3 year periods for individual stochastic realisations, which, when averaged, tend towards the annual periodicity to match the deterministic, due to the realisations going out of sync.

The household model without seasonal forcing was used to investigate the case of a flu-like disease suddenly becoming fatal in children, adults, and both. We found that when the population has a small proportion of susceptible individuals there is a higher chance that the disease persists when compared to the population with a larger proportion of susceptible individuals. Significant change in household proportions are recorded initially during the first outbreak of the disease when

the mortality rate increases during the infectious period of individuals.

In both the models with and without seasonal forcing, a measles-like disease fades out more in the household models compared to the homogeneous models. In each of the household models, within the first five years, we see periods of approximately 2-3 years where the proportion of realisations that fade out stay essentially constant for populations of greater than 225 household households in the seasonally forced model, and above 250 thousand in the non-forced model. During these periods, the proportion of realisations that have not faded out is above 0.5, which is suggestive of a critical community size between 375-500 thousand individuals in both household models, which is higher than in the existing literature [12, 14].

The household models in this thesis are extremely flexible in regards to the parameters and formulation, for example: household types can have unique birth rates; children and adults can be differentiated between; and although many population structure assumptions are made, the partitioning of the population into households alone relaxes some of the biggest population/contact assumptions required in developing a homogeneous model. The drawback of such a flexible model is the complexity and hence evaluation of such a model is time consuming and difficult, requiring either new methods or a renewed in-depth search into existing literature to efficiently simulate realisations and compute the equilibrium of the differential equations.

The household models in this thesis present a more detailed model than presented in [19], and the determinsitic approximation can be evaluated more efficiently than the model in [21], although each of these models have their own unique set of benefits. While the household models in this thesis can indeed be improved upon, especially in regards to the methods used to evaluate them, they can definitely be used to further investigate many problems related to endemic diseases and public health in the future.

Appendix A

Demography deterministic matrices derivation

$\mathcal{E}_{lin}$ events

Define $r^{(k)}_{(i,j)}$ to be the rate of event type $k \in \mathcal{E}_{lin}$ occurring in a household of type (i,j). Next, we define an event type specific transition vector, $\Lambda^{(k)}_{(i,j)}$. This column vector quantifies the state transition of the event type k happening in a household of type (i,j).

Define $F^{(k)}_{(i,j)}$ to be the vector that measures the rate into and out of each household in the case where event type k occurs in a household of type (i,j). That is,

$$F^{(k)}_{(i,j)} = r^{(k)}_{(i,j)} \Lambda^{(k)}_{(i,j)}. \tag{A.1}$$

Then define,

$$F^{(k)} := \left[F^{(k)}_{(0,0)}, F^{(k)}_{(1,0)}, ..., F^{(k)}_{(2,4)} \right]. \tag{A.2}$$

Finally, the total rate matrix for the events $k \in \mathcal{E}_{lin}$ is given by

$$F_{lin} = \sum_{k \in \mathcal{E}_{lin}} F^{(k)}.$$

$\mathcal{E}_{dec}$ and $\mathcal{E}_{im}$ events

For the events types $k \in \mathcal{E}_{dec} \cup \mathcal{E}_{im} \cup \mathcal{E}_{cf}$, we note that the rates of these events are not linear. This will mean that the functional form of these matrices will be different, but the approach we take in defining the matrices is very similar to that of the events with linear rates. The event types $k \in \mathcal{E}_{dec} \cup \mathcal{E}_{im}$ have rates that depend upon a preference function. Recall the event type "child moving out of home". When this event occurs, the child either moves into an empty household or leaves the population. Due to this feature, the rate of this event type is non-linear. Similarly for couple separation and immigration. Our aim is to construct two matrices such that we have the form

$$f(x)F_1 + (1 - f(x))F_2,$$

where $f(x)$ is the preference function, and F_1, F_2 are matrices of a similar form to that of F_{lin}.

Let k_1 denote the moving out event, k_2 be the couple separation event, and k_3 the immigration event. Then,

$$\mathbf{r}^{(k_1)}_{(i,j)} = \begin{cases} m_{out} & \text{for } i = 1, 2 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, let us define,

$$\mathbf{r}^{(k_2)}_{(i,j)} = \begin{cases} c_s & \text{for } i = 2 \text{ and } j = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{r}^{(k_3)}_{(i,j)} = \begin{cases} \xi p_{(i,j)} & \text{for } i = 1, 2 \text{ and } j = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

For each $m = 1, 2$, define k_m^{stay} to be the event type k_m but the individual

stays in the population and k_m^{leave} to be event type k_m but the individual leaves the population. Let $\Lambda_{(i,j)}^{(k_m^{stay})}$ be the state transition corresponding to the event type k_m in a household of type (i,j), where the individual moves *into an empty household*, and let $\Lambda_{(i,j)}^{(k_m^{leave})}$ be the state transition corresponding to the event type k_m in a household of type (i,j), where the individual *emigrates out of the population*.

Set,

$$F_{(i,j)}^{(k_j^{stay})} = r_{(i,j)}^{(k_j)} \Lambda_{(i,j)}^{(k_m^{stay})}, \quad F_{(i,j)}^{(k_j^{leave})} = r_{(i,j)}^{(k_j)} \Lambda_{(i,j)}^{(k_m^{leave})}.$$

Hence write $F^{(k_j^{stay})}$ and $F^{(k_j^{leave})}$ as the matrices that make up the column vectors $F_{(i,j)}^{(k_j^{stay})}$ and $F_{(i,j)}^{(k_j^{leave})}$, respectively, same as in Equation (A.2). That is,

$$F^{(k_m^{stay})} = \left[F_{(0,0)}^{(k_m^{stay})}, F_{(1,0)}^{(k_m^{stay})}, ..., F_{(2,4)}^{(k_m^{stay})} \right], \quad F^{(k_m^{leave})} = \left[F_{(0,0)}^{(k_m^{leave})}, F_{(1,0)}^{(k_m^{leave})}, ..., F_{(2,4)}^{(k_m^{leave})} \right].$$

As the immigration event is only defined for when the family chooses to enter the population, we need to find a different matrix for it. Let $\Lambda_{(i,j)}^{(k_3)}$ be the state transition corresponding to the immigration event when a household of type (i,j) enters the population. Set,

$$F_{(i,j)}^{(k_3)} = r_{(i,j)}^{(k_3)} \Lambda_{(i,j)}^{(k_3)}.$$

Then combining each household type, we can write,

$$F^{(k_3)} = \left[F_{(0,0)}^{(k_3)}, F_{(1,0)}^{(k_3)}, ..., F_{(2,4)}^{(k_3)} \right].$$

Hence we can define two matrices, F_1 and F_2, where F_1 corresponds to the events that depend on the preference function where each individual in the affected household stays in the population (or are entering the population in the case of the immigration event), and F_2 corresponds to the events that depend on the

preference function where an individual leaves the population. That is,

$$F_1 = F^{(k_3)} + \sum_{m=1}^{2} F^{(k_m^{stay})}, \quad F_2 = \sum_{m=1}^{2} F^{(k_m^{leave})}.$$

Couple formation events

We consider the event type $k \in \mathcal{E}_{cf}$. That is, the couple formation event. Write,

$$r_{(i,j)}^{(k)} = \begin{cases} c_f & \text{for } i = 1 \text{ and } j = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise.} \end{cases}.$$

Construct $\Lambda_{(i,j)}^{(k)}$ so as to be the state transitions, according to Table 3.1, and write,

$$F_{(i,j)}^{(k)} = r_{(i,j)}^{(k)} \Lambda_{(i,j)}^{(k)}$$

and hence finally define,

$$F_3 = \left[F_{(0,0)}^{(k)}, F_{(1,0)}^{(k)}, ..., F_{(2,4)}^{(k)} \right].$$

Appendix B

Demography with disease deterministic matrices derivation

Events with linear rates

Recall that $\mathbf{v} = (S_a, E_a, I_a, S_c, E_c, I_c)$. Consider the set of all configurations of the vector, $\mathbf{v}$. Next, recall the ordering of the configurations as in Equation 4.4. Hence denote the first configuration to be $\mathbf{v}_1$ and $\mathbf{v}_{981}$ to be the last; that is, $\mathbf{v}_1$ is the empty configuration and $\mathbf{v}_{981}$ is the household type $(2, 4)$ where each of the adults and children are recovered. Let $\mathcal{V} = \{\mathbf{v}_1, ..., \mathbf{v}_{981}\}$.

Recall Section 3.3.2. Let $r_{\mathbf{v}}^{(k)}$ be the rate of an event $k \in \mathcal{E}_{lin}$ occurring in the household configuration of $\mathbf{v}$ and let $\Lambda_{\mathbf{v}}^{(k)}$ be the vector which quantifies the number of each household configuration that is gained or lost after the event $k \in \mathcal{E}_{lin}$ occurs in household configuration $\mathbf{v}$. If household configuration $\mathbf{v}$ has household type (i, j),

$$r_{\mathbf{v}}^{(k)} = \begin{cases} b_j & \text{for birth events} \\ \phi p_{(i,j)} & \text{for emigration events} \\ \frac{S_a(\beta_{aa}I_a+\beta_{ca}I_c)}{i+j-1} & \text{for adult transmission events} \\ \frac{S_c(\beta_{ac}I_a+\beta_{cc}I_c)}{i+j-1} & \text{for child transmission events} \\ \sigma E_a & \text{for adult infection events} \\ \sigma E_c & \text{for child infection events} \\ \gamma I_a & \text{for adult recovery events} \\ \gamma I_c & \text{for child recovery events} \\ \omega R_a & \text{for adult waning immunity events} \\ \omega R_c & \text{for child waning immunity events} \end{cases}.$$

Then for each $k \in \mathcal{E}_{lin}$ and for each $\mathbf{v}$, let,

$$F_{\mathbf{v}}^{(k)} = r_{\mathbf{v}}^{(k)} \Lambda_{\mathbf{v}}^{(k)}.$$

Then we can write

$$F^{(k)} = \left[F_{\mathbf{v}_1}^{(k)}, ..., F_{\mathbf{v}_{981}}^{(k)}\right].$$

The events $k \in \mathcal{E}_d$ have rates that depend upon the number of individuals with particular disease statuses, which of course depends further on the household configuration. Let ψ be the generic symbol denoting a particular disease status; that is, $\psi \in \{S, E, I, R\}$. Then let,

$$r_{\mathbf{v},\psi}^{(k)} = \begin{cases} d_C \psi_c & \text{for children death events} \\ d_A \psi_a & \text{for both types of adult death events} \end{cases}.$$

Define $\Lambda_{\mathbf{v},\psi}^{(k)}$ be the vector which details the change in household configurations

for the event with rate $r^{(k)}_{\mathbf{v},\psi}$. Hence write

$$F^{(k)}_{\mathbf{v},\psi} = r^{(k)}_{\mathbf{v},\psi}\Lambda^{(k)}_{\mathbf{v},\psi}.$$

Then we can write

$$F^{(k)}_{\psi} = \left[F^{(k)}_{\mathbf{v}_1,\psi}, ..., F^{(k)}_{\mathbf{v}_{981},\psi}\right].$$

Finally, we let,

$$F^{(k)} = \sum_{\psi} F^{(k)}_{\psi}.$$

Since all the events in this section have linear rates, we can add each of the matrices together. That is, set,

$$F_1 = \sum_{k\in\mathcal{E}_{lin}\cup\mathcal{E}_d} F^{(k)}.$$

Events with two outcomes

Now we wish to find the matrices for the events with non-linear rates. The events $k \in \mathcal{E}_{dec}$ have two outcomes, specifically that the individuals can either remain in or emigrate out of the population, and hence there are two associated matrices. Let k_1 be when the event $k \in \mathcal{E}_{dec}$ occurs and the individual remains in the population and enters an empty household, and let k_2 be the case when the individual emigrates out of the population. As with the events in the set $\mathcal{E}_d$, the rates depend on the household configuration. For each $m = 1, 2$ let,

$$r^{(k_m)}_{\mathbf{v},\psi} = \begin{cases} m_{out}\psi_c & \text{for the children leaving home events} \\ \frac{c_s}{2}\psi_a & \text{for the couple separation events} \end{cases}.$$

Define $\Lambda^{(k_m)}_{\mathbf{v},\psi}$ to be the vector which details the change in household configurations

for the event with rate $r^{(k_m)}_{\mathbf{v},\psi}$. Hence write

$$F^{(k_m)}_{\mathbf{v},\psi} = r^{(k_m)}_{\mathbf{v},\psi}\Lambda^{(k_m)}_{\mathbf{v},\psi}.$$

Then we can write

$$F^{(k_m)}_{\psi} = \left[F^{(k_m)}_{\mathbf{v}_1,\psi}, ..., F^{(k_m)}_{\mathbf{v}_{981},\psi}\right].$$

Finally, we let,

$$F_2 = \sum_{k\in\mathcal{E}_{dec}}\sum_{\psi} F^{(k_1)}_{\psi} \quad \text{and} \quad F_3 = \sum_{k\in\mathcal{E}_{dec}}\sum_{\psi} F^{(k_2)}_{\psi}.$$

Immigration events

Although similar to the events with two outcomes, in that the rate of immigration requires the preference function, here immigration only has one option — a household enters the population. For the event $k \in \mathcal{E}_{im}$, in which a new household of configuration $H^{(i,j)}_{\mathbf{v}}$ enters an empty household, let $\Lambda^{(k)}_{\mathbf{v}}$ be the vector that quantifies the removal of an empty household and replaces it by a household of type $H^{(i,j)}_{\mathbf{v}}$. Then set,

$$F^{(k)}_{\mathbf{v}} = \xi p_{(i,j)}\Lambda^{(k)}_{\mathbf{v}}.$$

Hence, the matrix F_4 is given by,

$$F_4 = \left[F^{(k)}_{\mathbf{v}_1}, ..., F^{(k)}_{\mathbf{v}_{981}}\right].$$

Couple formation events

In this section, we create four matrices as the couple formation event has rates that depend on the number of household types $(1, 0)$ as well as the number of household configurations that the $(1, 0)$ household is pairing with. However, the $(1, 0)$ household type has four associated configurations that correspond to the

adult being susceptible, exposed, infectious or recovered, and this non-linearity requires four separate matrices to be constructed. Let k denote the couple formation event. For each disease status, represented generically by ψ, let $\Lambda^{(k)}_{\mathbf{v},\psi}$ be the vector that quantifies the change in the number of households of each configuration after a couple formation event where the household configuration $\mathbf{v}$ combines with the lone adult household where the adult has disease status represented by ψ. Then let,

$$F^{(k)}_{\mathbf{v},\psi} = c_f \Lambda^{(k)}_{\mathbf{v},\psi},$$

and then,

$$F_\psi = \left[F^{(k)}_{\mathbf{v}_1,\psi}, ..., F^{(k)}_{\mathbf{v}_{981},\psi}\right].$$

Between-household transmission events

The transmission events that occur between households also have rates that are non linear. As such two more matrices have to be constructed; one for adults moving from the susceptible to exposed class and one for the children. We let k_a denote the event when the disease is transmitted between households to an adult, whereas k_c is the event when the disease is transmitted between-households to a child. Then $\Lambda^{(k_a)}_{\mathbf{v}}$ is the vector that quantifies the change in the number of household configurations when an adult in the household configuration $\mathbf{v}$ has the disease externally transmitted to them between-households. Similarly if a child has the disease transmitted to them, $\Lambda^{(k_c)}_{\mathbf{v}}$.

Recall that the rate of a susceptible adult in a household of type (i, j) and configuration $\mathbf{v}$ becoming exposed to the disease through external transmission is of the form,

$$S_a \frac{\left(\alpha_{aa} I^T_a + \alpha_{ca} I^T_c + M\epsilon\right)}{\left(\sum_{(i,j),\mathbf{v}} m^{(i,j)}_{\mathbf{v}} (i+j)\right)},$$

where M is the total number of households in the population, I^T_a is the total

number of infectious adults in the population and I_c^T is the total number of infectious children in the population. Note here that the entire fraction term depends on the number of each of the household configurations in the population at the time. Hence, here we construct our matrices F_5, F_6 using the part that does not depend on $\mathbf{m}$; that is, S_a and S_c, respectively. For a household configuration $\mathbf{v}$, we write

$$F_{\mathbf{v}}^{(k_a)} = S_a \Lambda_{\mathbf{v}}^{(k_a)} \text{ and, } F_{\mathbf{v}}^{(k_c)} = S_c \Lambda_{\mathbf{v}}^{(k_c)}.$$

Then let,

$$F_5 = \left[F_{\mathbf{v}_1}^{(k_a)}, ..., F_{\mathbf{v}_{981}}^{(k_a)}\right] \quad \text{and} \quad F_6 = \left[F_{\mathbf{v}_1}^{(k_c)}, ..., F_{\mathbf{v}_{981}}^{(k_c)}\right].$$

Bibliography

[1] M. J. Keeling and P. Rohani. *Modelling Infectious Diseases in Humans and Animals*. Princeton University Press, New Jersey, USA, 2008.

[2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, UK, 1992.

[3] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115:700–721, 1927.

[4] J. Lessler, N. G. Reich, R. Brookmeyer, T. M. Perl, K. E. Nelson, and D. A. T. Cummings. Incubation periods of acute respiratory viral infections: a systematic review. *The Lancet Infectious Diseases*, 9:291 – 300, 2009.

[5] A. J. Black and A. J. McKane. Stochastic amplification in an epidemic model with seasonal forcing. *Journal of Theoretical Biology*, 267:85 – 94, 2010.

[6] P. G. Ballard, N. G. Bean, and J. V. Ross. The probability of epidemic fade-out is non-monotonic in transmission rate for the markovian SIR model with demography. *Journal of Theoretical Biology*, 393:170 – 178, 2016.

[7] T. Britton. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225:24 – 35, 2010.

[8] P. C. Cross, P. L. F. Johnson, J. O. Lloyd-Smith, and W. M. Getz. Utility of R_0 as a predictor of disease invasion in structured populations. *Journal of The Royal Society Interface*, 4:315–324, 2007.

[9] A. J. Black, T. House, M. J. Keeling, and J. V. Ross. The effect of clumped population structure on the variability of spreading dynamics. *Journal of Theoretical Biology*, 359:45 – 53, 2014.

[10] J. M. Heffernan and M. J. Keeling. Implications of vaccination and waning immunity. *Proceedings of the Royal Society of London B: Biological Sciences*, 276:2071–2080, 2009.

[11] S. Krugman, J. P. Giles, H. Friedman, and S. Stone. Studies on immunity to measles. *The Journal of Pediatrics*, 66:471 – 488, 1965.

[12] M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)*, 120:48–70, 1957.

[13] F. L. Black. Measles endemicity in insular populations: Critical community size and its evolutionary implication. *Journal of Theoretical Biology*, 11:207 – 211, 1966.

[14] M. J. Keeling and B. T. Grenfell. Disease extinction and community size: Modeling the persistence of measles. *Science*, 275:65–67, 1997.

[15] N. Becker. An epidemic chain model. *Biometrics*, 36:249–254, 1980.

[16] N. J. T. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, London, 2 edition, 1975.

[17] F. Ball and G. Mollison, D.and Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7:pp. 46–89, 1997.

[18] F. Ball, T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63 – 67, 2015.

[19] K. Glass, J. M. McCaw, and J. McVernon. Incorporating population dynamics into household models of infectious disease transmission. *Epidemics*, 3:152 – 158, 2011.

[20] N. Geard, J. M. McCaw, A. Dorin, K. B. Korb, and J. McVernon. Synthetic population dynamics: A model of household demography. *Journal of Artificial Societies and Social Simulation*, 16:8, 2013.

[21] N. Geard, K. Glass, J. M. McCaw, E. S. McBryde, K. B. Korb, M. J. Keeling, and J. McVernon. The effects of demographic change on disease transmission and vaccine impact in a household structured population. *Epidemics*, 13:56 – 64, 2015.

[22] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403 – 434, 1976.

[23] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7:pp. 49–58, 1970.

[24] S. Ross. *Introduction to Probability Models*. Academic Press, Oxford, UK, 2014.

[25] D. W. Jordan and P. Smith. *Nonlinear Ordinary Differential Equations: An Introduction to Dynamincal Systems*. Oxford University Press, New York, USA, 1999.

[26] J. D'Errico. Adaptive robust numerical differentiation.

`http://au.mathworks.com/matlabcentral/fileexchange/13490-adaptive-robust-numerical-differentiation`, 2006.

[27] J. N. Lyness and C. B. Moler. Generalized romberg methods for integrals of derivatives. *Numerische Mathematik*, 14:1–13, 1969.

[28] E. Kreyszig, H. Kreyszig, and E. J. Norminton. *Advanced Engineering Mathematics*. John Wiley & Sons, Inc, New Jersey, USA, 1999.

[29] Australian Bureau of Statistics. Census of population and housing: Time series profile, 2011.

[30] Australian Bureau of Statistics. Census of population and housing: Basic community profile, 2011.

[31] H. E. Soper. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, 92:34–73, 1929.

[32] M. S. Bartlett. Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Health*, pages 81–109, Berkeley, Calif., 1956. University of California Press.

[33] N. Goeyvaerts, L. Willem, K. Van Kerckhove, Y. Vandendijck, G. Hanquet, P. Beutels, and N. Hens. Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence. *Epidemics*, 13:1 – 9, 2015.

[34] C. E. Mills, J. M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 2004.

[35] O. A. van Herwaarden. Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak. *Journal of Mathematical Biology*, 35:793–813, 1997.

[36] World Health Organisation. *Measles Fact Sheet*, 2016.

[37] W. Arveson. *A Short Course on Spectral Theory*. Springer-Verlag New York Inc., New York, USA, 2001.

[38] N. G. Becker and K. Dietz. The effect of household distribution on transmission and control of highly infectious diseases. *Mathematical Biosciences*, 127:207 – 219, 1995.

[39] N. G Becker and R. Hall. Immunization levels for preventing epidemics in a community of households made up of individuals of various types. *Mathematical Biosciences*, 132:205 – 216, 1996.

[40] N. G. Becker and D. N. Starczak. Optimal vaccination strategies for a community of households. *Mathematical Biosciences*, 139:117 – 132, 1997.

[41] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115:1716–1733, 2001.

[42] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis. Binomial distribution based τ-leap accelerated stochastic simulation. *The Journal of Chemical Physics*, 122:024112, 2005.

[43] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104:1876–1889, 2000.

[44] K. E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley, New York, USA, 1978.

[45] T. Haqiri and F. Poloni. Methods for verified stabilizing solutions to continuous-time algebraic Riccati equations. *ArXiv:1509.02015 e-prints*, September 2015.

[46] D. A. Bini, B. Iannazzo, and F. Poloni. A fast newton's method for a nonsymmetric algebraic riccati equation. *SIAM Journal on Matrix Analysis and Applications*, 30:276–290, 2008.