Balancing model complexity and inferential capability for disease modelling

by

Alex Bishop

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Centre for Complexity Science

April 2020

Contents

List of Tables

List of Figures

Acknowledgments

Firstly I would like to thank my supervisors, Deirdre and Thomas, for the advice and support they provided.

Huge thanks go to the staff and students of the Centre for Complexity Science for providing a real sense of community. Particular thanks go to Janis, Michael, and the rest of 'Bike gang' for helping me find my love of cycling.

Thanks to various long-term housemates: Jo, for being my best friend for over a decade; John, for many years of stimulating conversations heavily fuelled by beer and whisky; Janis, for offering sage wisdom solely through the medium of peep show quotes, and cleaning up after John.

Eternal thanks go to my family for their unwavering support and love over the years even though I may have been distant.

Finally, I'd like to thank Libby for bringing the light back into my life, believing in me when I didn't believe in myself, and all the things you've done to support me over an excessively long writing-up period.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Parts of the work in Chapter 2 follow on from work that formed part of the author's MSc in Complexity Science at the University of Warwick in 2014.
Chapter 3 has been published as:

- Bishop, A., Kiss, I.Z. and House, T., 2018, December. Consistent approximation of epidemic dynamics on degree-heterogeneous clustered networks. In International Conference on Complex Networks and their Applications (pp. 376-391). Springer, Cham.

My co-authors contributions were as supervisors, helping identify the right question and contributing to discussions of methods.
The methods developed in Chapter 5 have been published as:

- Dyson, L., Marks, M., Crook, O.M., Sokana, O., Solomon, A.W., Bishop, A., Mabey, D.C. and Hollingsworth, T.D., 2017. Targeted Treatment of Yaws With Household Contact Tracing: How Much Do We Miss?. American journal of epidemiology, 187(4), pp.837-844.

The details of this analysis are not detailed in the thesis as I mainly acted in a consultative role on the model structure and inference methodology.
The material in Chapter 7 was shared with stakeholders and the World Health Organisation Guidelines Development Group for Lymphatic Filariasis and is referenced in the WHO Guidelines "Guideline – Alternative mass drug administration regimens to eliminate lymphatic filariasis" [151] as 'the Markov model' (there are no details of

the model provided in this document). The analysis was wholly performed by myself with Dr Jonathan King, WHO, providing data and making decisions on which policy scenarios to simulate and which results were most policy relevant to present.

Abstract

Mathematical models for study of infectious diseases have a rich history but it is only in recent years that directly fitting highly complex models to data has become possible. This has lead to a quantum leap in the capabilities of mathematical modelling for contributing to an evidence base for policy decisions. There still remains a large gap between the most complex models we can simulate and the most complex models we can perform inference on resulting in a trade-off between model complexity and inferential capability. This thesis tackles three separate problems with this trade-off in mind.

First, we study the dynamics of epidemics on degree heterogeneous clustered networks. Network models have many attractions but possess drawbacks such as one must generally resort to stochastic simulation for clustered networks, which represent realistic societal structure, as closed-form approximations of the dynamics do not hold in the highly clustered regime. Furthermore, data for these systems are hard to collect as they must measure the pairwise interactions of each individual - this lack of data limits the possibilities for applying inference. We develop a new model not requiring extensive simulations that approximates these dynamics more accurately than previous approaches, thus improving on the first problem mentioned.

Second, across several chapters we analyse the role of household structure in the transmission and control of soil-transmitted helminths (STH). Starting with a hierarchical negative binomial regression for which inference can easily be performed but which neglects the non-independence of observations, we move on to develop a general methodology for constructing and performing Bayesian inference on stochastic household models that consider different transmission dynamics within and between the households in a population. This permits us to estimate the extent to which transmission occurs within, compared to between, households and simulate the effectiveness of various control strategies – with some exploiting the household structure. The limits to which this general methodology may be extended to arbitrary demographic classes and infection levels before inference with exact-likelihood methods no longer become computationally feasible is explored.

Finally, we build a model of the global control programme for lymphatic filariasis at a regional level, forecasting the number of treatments required each year and their costs in order to reach elimination. A scenario where existing guidelines remained in place and a scenario where proposed guidelines incorporating a new treatment were considered. Our analysis was used by WHO as part of the evidence base for adopting precisely these new guidelines.

Abbreviations

ER: Erdős-Rényi

MCMC: Markov Chain Monte Carlo

CTMC: Continuous Time Markov Chain

ME: Master Equation

STH: Soil-Transmitted Helminths

ZINB: Zero Inflated Negative Binomial

LF: Lymphatic Filariasis

NTD: Neglected Tropical diseases

NUTS: No U-Turn Sampler

MH: Metropolis-Hastings

HMC: Hamiltonian Monte-Carlo

IU: Intervention Unit

WHO: World Health Organisation

GPELF: Global programme to eliminiate lymphatic filariasis

MF: Microfilarae

IVM: Ivermectin

ALB: Albendazole

DEC: Diethylcarbamazine

TAS: Transmission Assessment Survey

pre-TAS: pre-Transmission Assessment Survey

IDA: Treatment with Ivermectin, Diethylcarbamazine, and Albendazole

JRF: Joint Reporting Form

JRSM: Joint Request for Select preventative chemotherapy Medicines

Notation	**Meaning**
$\mathcal{G} = (\mathcal{N}, \mathcal{L})$	A network with node set $\mathcal{N}$ and link set $\mathcal{L}$
$a, b, c, \ldots$	Elements of $\mathcal{N}$ (i.e. nodes of the network)
N	Size of the network (equal to $\lvert\mathcal{N}\rvert$)
G_{ab}	the (a, b)-th elements of the adjacency matrix of $\mathcal{G}$
k_a	Degree of node a
$\psi(x)$	PGF of degree distribution
θ	The Volz variable
n	Degree of a regular network
M	Maximum degree in the network
$\langle f(k) \rangle$	Mean of a function of degree across the network
ϕ	Clustering coefficient of the network
S, I, R	Disease states: Susceptible; Infective; Removed
X_a	Disease state of node a
γ	Recovery rate
τ	Rate of disease transmission across a network link
$A, B, C, \ldots$	Possible disease states (any of S, I and R)
$[A]$	Expected number of nodes in state A
$i, j, l, s, \ldots$	Counts of susceptible or infective contacts
$[A_{s,i}]$	Expected number of nodes in state A with s susceptible contacts and i infective contacts
$[AB]$	Expected number of A–B pairs
$[ABC], [ABC]_{\triangle}, [ABC]_{\wedge}$	Expected number of A–B–C triples (all, closed, unclosed)
$\mathcal{C}_{AB}$	Correlation between states A and B on the network

Table 1: Mathematical notation for epidemic models on networks used in chapter 3.

CHAPTER 1

Introduction

Disease has preyed upon humankind for as long as we have existed, causing unforetold misery and the majority of global deaths [1]. Between 1990 and 2013 worldwide deaths from communicable diseases decreased; however the health gap between rich countries and poor countries remains vast, and the gap between rich and poor in poor countries possibly even more so with over a billion people in low-income populations of developing tropical countries suffering from a diverse group of tropical infections called Neglected Tropical Diseases (NTDs) - so called because they affect the very poorest in society. Because of this, despite the progress being made against disease worldwide the study of infectious diseases remains just as relevant as ever. Furthermore, whilst serious communicable diseases are mostly a thing of the past in developed countries there remain 'heavy tails' with the potential for emerging diseases such as SARS or zoonotic spillover of diseases such as Ebola and influenza having the potential to cause pandemics that would lead to untold human and economic losses without timely, effective action informed by mathematical modelling.

The continued realisation of the increasing computational capabilities promised by Moore's law have seen an increased intensity and prevalence of fitting mathematical models directly to data in order to provide an evidence base for policies aiming to understand and control the spread of infectious diseases. Doing this often requires a pragmatic compromise between a model's realism (i.e. complexity), the data available, and the inference procedures the model can be fitted with. Within this thesis we tackle three broadly separate problems requiring this compromise to take several different forms. Chapters 4 and 5 consider this compromise more directly as they investigate the computational limits of inference for a certain class of models on a sliding scale of complexity.

Chapter 2 introduces the minimum background information on mathematical models for infectious diseases, stochastic processes, and practical Bayesian inference so that the later chapters may be understood. The history and use of the first

compartmental epidemiological models is introduced before moving on to consider how various shortcomings of this original approach may be overcome. Basic definitions, properties, and methods for the solution of Markov processes are then briefly covered before a longer section introducing Bayesian inference with MCMC paying particular attention to performing this in an effective manner. Additional tailored literature reviews are included in the introductory sections of each chapter.

The dynamics of epidemics on degree heterogeneous networks with large clustering coefficients are studied in **Chapter 3**. We develop a new model that makes use of a moment-closure approximation in order to give a closed set of ODEs that more closely approximate the 'exact' dynamics (which we obtain via. exhaustive simulation) than previous approaches in the literature. In addition to this we study the error properties of this type of closure finding that our results agreed with the work of Pellis et al. [157]. Whilst providing a hyper-granular representation of the interactions that facilitate the spread of infectious diseases, network models for clustering are highly complex with the focal point of ongoing research being improving the more efficient solution of these models which needs to be over come before inference can be easily applied to this type of model.

Not all diseases require such a granular representation of dynamics. Key dynamics can still be captured by keeping only the most important unit of structure, such as a household unit. **Chapter 4** analyses the role of household structure in the transmission and control of soil-transmitted helminths (STH) where there is evidence for household-clustering of infections [200]. The data are explored and a hierarchical negative binomial Bayesian regression fitted to the data which displays evidence of household clustering. The empirical model approach of this chapter allows for a lot of the heterogeneity of the system to be captured and is highly amenable to inference; however as it assumes independence of observations it neglects the key feature of a spreading process. **Chapter 5** corrects this by developing a general methodology for constructing and performing Bayesian inference on stochastic household models of arbitrary demographic classes and infection levels. This allows for heterogeneity in age and infectiousness as well as household clustering to be modelled in a more realistic manner than an empirical model. Furthermore, with this approach we forward simulate the epidemic dynamics for a model parameterised by the posterior simulation and perform *in silico* experiments to estimate the distribution of the time to elimination under various hypothetical control strategies. Developing this general methodology leads to multiple challenges when scaling the complexity of the household model considered (in terms of the demographic classes and infection levels). **Chapter 6** focuses on circumventing these challenges in order to find

the reasonable computational limit of this methodology - at least within an exact likelihood framework.

Finally, in **chapter 7** we move beyond the local-scale impact of treatment that the previous three chapters focused on in order to focus on the analysis of a global control programme. This leap in scale required a comparatively straightforward modelling approach. A model of the global control programme for lymphatic filariasis at a regional level is built and used to forecast the number of treatments required each year and their costs in order to reach elimination. This is performed for a scenario where existing guidelines remained in place and under a scenario where proposed guidelines incorporating a new treatment with the capability to accelerate control programmes was adopted. Our analysis was used by WHO as part of the evidence base for adopting these new guidelines and provides an example where even simple models that make the right compromises can have real-world impact.

Over the course of this thesis, a range of epidemiological modelling approaches, and inference approaches are investigated and applied to scientific challenges at different scales with appropriate compromises to be found between the realistic representation of population structure in epidemic models, model tractability, and identifiability from data. In addition to the substantive insights gained for the problems in hand, this thesis gives general methodological insights into how to resolve this compromise, which are ubiquitous in the study of complex systems.

CHAPTER 2

Background

Within this chapter we give a brief exposition of several approaches to the mathematical modelling of infectious diseases (section 2.1), basic stochastic processes (section 2.2), and performing practical Bayesian inference (section 2.3). This background material provides context for later chapters.

2.1 Mathematical modelling of infectious diseases

Mathematical models may be used to understand the spread of infectious diseases by encoding underlying assumptions into a set of equations that describe the dynamics of the spreading process. Simpler models that permit some analytical solution can be used to provide a caricature of the dynamics and derive simple relationships between parameters, such as the rough proportion of a homogeneously mixing population that must be immunised in order to contain the spread of a disease via. herd immunity [49]. Reality is typically much messier with populations that mix heterogeneously meaning that more complex models that do not permit analytical solutions are often required to answer real-world policy problems. The advent of modern computing has enabled the utilisation of more complex models and their parameterisation by directly fitting models to data and has thus led to new questions being able to be answered. We now give a brief overview of the building blocks of mathematical epidemiology and show some of the more complex features that may be added to them in order to better model different systems.

2.1.1 Compartmental models

The seminal work of Kermack and McKendrick [105] introduced the notion of compartmental models for epidemics using ordinary differential equations (ODEs). This prototypical SIR model partitions the population within which a disease is spreading into one of three compartments:

- Susceptible (S): not infected with the disease and has no immunity.
- Infected (I): infected with the disease and currently infectious.
- Recovered (R): immune.

The population typically starts off in the susceptible compartment apart from a small number of infectious individuals who start the epidemic. These infected individuals spread the infection and susceptible individuals move into the infected compartment. As infected individuals recover from the disease they move into the recovered compartment and remain there (acquire life-long immunity).

Considering an infinite, large, closed population (no births, deaths or migration) we can write down differential equations for this deterministic limit under the assumption of homogeneous (mass-action) mixing,

$$\frac{dS(t)}{dt} = -\beta/NS(t)I(t) \tag{2.1}$$
$$\frac{dI(t)}{dt} = \beta/NS(t)I(t) - \gamma I(t) \tag{2.2}$$
$$\frac{dR(t)}{dt} = \gamma I(t) \tag{2.3}$$

$S(t)$, $I(t)$ & $R(t)$ denote the number of people in the population of $N = S(t) + I(t) + R(t)$ in each compartment respectively (generally the time dependence of the states is dropped from notation for simplicity); β denotes the transmission rate and represents the encounter rate between the susceptible and infected classes along with the probability of transmission; γ is the recovery rate ($1/\gamma$ denotes the average infectious period). This scenario is obviously a huge simplification of the real system but by making some more realistic assumptions, these compartmental models are very effective modelling tools. For example, (2.1-2.3) allows us to derive the condition under which an epidemic will grow (2.4) - this is known as the basic reproductive ratio R_0 [121]. R_0 gives the average number of secondary cases produced by an infected individual in a wholly susceptible population. Given we know R_0 we can then use (2.4) to derive the threshold proportion of the population, $V^* = 1 - \frac{1}{R_0}$, whom if vaccinated (unable to be infected) provide herd immunity to the population.

$$\frac{dI}{dt} > 0 \implies \frac{S(0)}{N} > \frac{\gamma}{\beta} = \frac{1}{R_0} \tag{2.4}$$

Whilst this classic model highlights how vaccination programmes can in theory stop the spread of a disease if some level of vaccination is reached in the pop-

ulation, there are many shortcomings. Which shortcoming to address is typically a disease and research question specific scenario as not every shortcoming can be addressed at once by some 'best model'. This can be because we do not possess the mathematical/computational apparatus to solve or perform inference on such a model or because the data to inform that model simply do not or cannot exist. Three problems in particular exist with the classic SIR model which are most important to the diseases modelled in the later chapters of this thesis: populations do not mix homogeneously in reality; spreading processes are not deterministic; and not all diseases have dynamics whereby you are infected, have some constant level of infectiousness, and then recover with permanent immunity. The next three subsections briefly address these main shortcomings as they relate to the thesis.

2.1.2 Stochastic models

The ODE representation of the classic SIR model (2.1-2.3) provides a low-dimensional and easy to solve representation of disease dynamics albeit a deterministic one - the same trajectory is always observed (given the same initial conditions). This approximation is reasonable if the problem being considered is for an epidemic process that has already taken off in a large population, as stochastic fluctuations will be small. In smaller populations or for low $I(t)$ the randomness of the dynamics becomes important. Another simple epidemic model is the SIS model - after recovering from infection individuals become susceptible again instead of acquiring life-long immunity. The SIS model yields deterministic ODEs similar to its SIR counterpart,

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \gamma I \\ \frac{dI}{dt} &= \beta SI - \gamma I \end{aligned}$$

Stochastic dynamics may be added by writing down the Master/Kolmogorov equations [112] of the SIS epidemic. Once formulated, the Master equations of a stochastic process may be: solved directly (for a small subset of models such as the simple SIS model); solved to numeric precision using numerical linear algebra; or realisations of the stochastic process may be sampled using stochastic simulation algorithms such as the Gillespie algorithm (Algorithm 1). Here we show how to formulate the Master equations so that we may solve them to machine precision in Chapter 5. The Master equations are given by a set of ODEs, for the probability of finding the population in every possible state, e.g. $P_{S,I}(t)$ is the probability of having S susceptible and I infected ($S, I \in \mathbb{Z}$) at time t. As $\frac{dS}{dt} = -\frac{dI}{dt}$ and $S + I = const.$

we do not need to explicitly keep track of the susceptibles and we are left with a 1 dimensional system.

Conceptually, P_I may be considered as the proportion of an infinite number of simulations that have I infected. Four events occur that modify the proportion of simulations in state I:

1. $I \to I-1$: a simulation in state I may have an infected individual recover at rate γI, leaving $I-1$ infected.

2. $I \to I+1$: a simulation in state I may have a susceptible become infected at rate $\beta SI = \beta(N-I)I$, leaving $I+1$ infected.

3. $I+1 \to I$: a simulation in state $I+1$ may have an infected individual recover at rate $\gamma(I+1)$, leaving I infected.

4. $I-1 \to I$: a simulation in state $I-1$ may have a susceptible become infected at rate $\beta(S+1)(I-1) = \beta(N-I+1)(I-1)$, leaving I infected.

These 4 processes yield a set of $N+1$ ODEs ($I = 0 \dots N$) each representing the flow of probability mass in/out of the $N+1$ states of our system.

$$\frac{dP_I}{dt} = -P_I\gamma I - P_I\beta(N-I)I + P_{I+1}\gamma(I+1) + P_{I-1}\beta(N-I+1)(I-1) \quad (2.5)$$

When solved these equations allow a rich understanding of the stochastic dynamics of the system.

One may make a small modification to the model by adding an external force of infection, ϵ, representing an import of infection into the population from an external source. By following the process above we may arrive at a slightly modified set of Kolmogorov equations,

$$\frac{dP_I}{dt} = -P_I\gamma I - P_I(\beta I+\epsilon)(N-I) + P_{I+1}\gamma(I+1) + P_{I-1}(\beta(I-1)+\epsilon)(N-I+1) \quad (2.6)$$

These equations may then be reformulated as a matrix equation $\frac{d\mathbf{p}}{dt} = \mathbf{Q}\mathbf{p}$, where $\mathbf{p}$ is a column vector of the $N+1$ probabilities, P_I and $\mathbf{Q}$ is the $(N+1)\times(N+1)$ tri-diagonal transition matrix of a Markov Chain (section 2.2.1),

$$\mathbf{Q} = \begin{bmatrix} -N\epsilon & \gamma & 0 & \dots & 0 \\ N\epsilon & -((\beta+\epsilon)(N-1)+\gamma) & 2\gamma & \dots & \vdots \\ 0 & (\beta+\epsilon)(N-1) & -((2\beta+\epsilon)(N-2)+2\gamma) & \ddots & 0 \\ \vdots & \vdots & (2\beta+\epsilon)(N-2) & \ddots & N\gamma \\ 0 & \dots & \vdots & \ddots & -N\gamma \end{bmatrix} \tag{2.7}$$

with the solution given by $\mathbf{p}(t) = \exp(\mathbf{Q}t)\mathbf{p}(0)$ where $\exp(\boldsymbol{Q}t)$ denotes the matrix exponential (section 2.2.2) of $\boldsymbol{Q}t$ [103].

2.1.3 Structured models

The level and type of structure added to a model depends not only on the particular disease being considered but on both the data available and any other complexities (such as stochastic dynamics) added to the model. Transmission could be modelled on a network [141] such that transmission only occurs between nodes (individuals) in the network with an edge between them, or based on the strength of the link between them. A link between two individuals could represent the time spent in close contact if modelling a disease such as SARS [127], or it could represent the existence of a sexual relationship if we were interested in modelling the spread of HIV [115]. Network models are the subject of chapter 3 and are discussed in depth there. Alternatively, transmission could be modelled with a spatial transmission kernel for diseases such as Foot and Mouth disease [100] - though we do not consider any spatial transmission in this thesis and therefore refer the interested reader to Keeling and Rohani [99]. Networks provide an incredibly flexible framework for modelling local & global spatial structure and even temporal structure [192] but they make inference and the incorporation of other modelling considerations harder therefore an alternative approach is to consider a lower dimensional representation that captures the key heterogeneity. One example are meta-population-type models [123] which consider a set of homogeneously mixing local populations that either mix with other populations at another rate or migrate between populations. A subset of meta-population type models are household models [17, 18, 120] that represent the population as household units with two levels of mixing/transmission occurring: transmission from another of the N household members at rate $\beta/(N-1)$ and transmission from households at rate $\lambda_g J(t)$ where $J(t)$ is the total number of people infected at time t. Household models are the key modelling component of chapter 5

and are discussed in depth there.

2.1.4 Macroparasite modelling

Microparasites such as viral or bacterial infections typically rapidly multiply and saturate within a susceptible host, making the discrete transition from 'Susceptible' to 'Infected' a reasonable approximation. Not all diseases can be effectively modelled with this single infectious stage. For example, Macroparasitic diseases such as Helminth infections have multiple infection levels corresponding to the macroparasitic load within the host, and diseases such as Malaria have multiple hosts which must also be modelled.

Anderson and May [5] introduced various models of host-macroparasite interactions for parasites that do not reproduce directly within their host, but which produce transmission stages such as eggs that develop outside of the host. The basic model of the number of hosts, $H(t)$, and parasites, $P(t)$, at time t is given by the following ODEs,

$$\frac{dH}{dt} = (b-d)H - \alpha P \tag{2.8}$$

$$\frac{dP}{dt} = \left(\lambda \frac{PH}{H_0 + H}\right) - (d+\mu)P - \alpha H E_t(i^2) \tag{2.9}$$

This model omits density dependent constraints on the host population as a simplification which clarifies insights into how parasites regulate host populations - the host population reproduces at per-capita rate b with natural mortality occurring at per-capita rate d. Furthermore, each parasite alters the host death rate at a linear rate α resulting in the rate of loss of hosts in a population of size $H(t)$ of $\alpha H(t) \sum_{i=0}^{\infty} ip(i) = \alpha P(t)$ where $p(i)$ is the probability a given host harbours i parasites.

Meanwhile, the rate of production of environmental stages (eggs etc.) per parasite is λ, which yields a net rate for the total parasite population of $\lambda P(t)$. Survival of external stages will depend on the density of the host population - a lower density decreases the chances of gaining entry to a host before natural mortalities. The proportion of parasites gaining entry to a host is determined by the transmission factor $H(t)/(H_0 + H(t))$ - small H_0 corresponds to a smaller proportion gaining entry to hosts. This factor is incorporated into the rate at which new parasites are acquired within the host population which is $\lambda P(t)H(t)/(H_0 + H(t))$ - i.e. the rate of production of eggs multiplied by the transmission factor.

Parasite death has three components: losses due to host mortalities at rate d; within-host parasite mortality at rate μ; and losses from parasite induced deaths. The first two components are straightforwardly $(d + \mu)P(t)$. The last component is less straightforward as the per-capita host loss-rate is αi for a host with i worms. This induces a dependence on the mean-square number of parasites per host: $\alpha H(t) \sum_{i=0}^{\infty} i^2 p(i) = \alpha H(t) E_t(i^2)$. If the parasites are distributed independently randomly among hosts then (2.9) may be written in closed form by applying the following substitution,

$$E_t(i^2) = E_t(i) + E_t(i)^2 = P(H + P)/H^2. \tag{2.10}$$

Beyond this simple model, density dependencies and non-random parasite distributions among hosts can be incorporated. For instance, Anderson and May [5] express $E_t(i^2)$ such that the hosts are distributed according to a negative binomial distribution which provides a good empirical model for a large number of observed parasite distributions. Much of the macroparasite literature studying the role of parasite aggregation and parasite-induced mortality make extensive use of moment closure techniques and negative binomial approximations [75, 83, 128]. Trying to capture these host-parasite dynamics accurately entails significant added complexity, which as a result means that other factors such as the contact structure of hosts and fully stochastic dynamics cannot be incorporated in one model.

2.2 Stochastic processes

A stochastic process is a collection of random variables taking a value from the 'state space' indexed by some 'index set' such that each random variable of the process is uniquely associated with an element in the set. Many varieties of stochastic processes exist but for this thesis we stick to one particular subset - Markov processes.

2.2.1 Markov Processes

Markov processes are stochastic processes that satisfy the Markov, or memoryless, property - the conditional probability distribution of future states depends only on the present state [194]. A Markov chain is a Markov process with either a discrete state space or a discrete index set (typically representing time). A discrete-time Markov Chain is a sequence of random variables X_i with the index set $i \in \mathbb{N}$ where the Markov property holds (2.11) and $X_i \in S \ \forall i$ where the state space S is a

countable set.

$$\mathbb{P}(X_{t+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t) = \mathbb{P}(X_{t+1} = x | X_t = x_t) \tag{2.11}$$

A continuous time Markov chain (CTMC) is defined similarly but has a continuous time parameter, $t \in [0, \infty)$. As there are no smallest time steps a way of talking about individual jumps is needed. For any times $t_0, \ldots, t_n$ indexed up to $n \in \mathbb{N}$ with states $s_0, \ldots, s_n$ recorded at these times it can be shown that,

$$\mathbb{P}(X_{t_{n+1}} = s_{n+1} | X_{t_0} = s_0, \ldots, X_{t_n} = s_n) = p_{s_n s_{n+1}}(t_{n+1} - t_n) = p_{ij}(\delta t) \tag{2.12}$$

where $p_{ij}(\delta t)$ (the probability of being in state j at time t_{n+1} given the chain was in state i at time t_n) is the solution of the forward equation,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{Q}\boldsymbol{x}(t) \tag{2.13}$$

$$||x||^1 = 1 \tag{2.14}$$

where $\boldsymbol{x}(0)$ is the identity matrix. The left-stochastic matrix $\boldsymbol{Q} \in \mathbb{R}^{\|S\| \times \|S\|}$ has elements q_{ij} equal to the rate of the process transitions from i to j for $i \neq j$ and elements q_{ii} equal to the negative sum of the rates of each column (2.15-2.16).

$$Tr(\boldsymbol{Q}) = \sum_{\{i,j:i \neq j\}} \boldsymbol{Q} \tag{2.15}$$

$$\sum_{j=1}^{n} \boldsymbol{Q}_{ij} = 1, \ \forall i \in 1, \ldots, n \tag{2.16}$$

$$\tag{2.17}$$

Perhaps the most famous example of a Markov process is that of Google's PageRank algorithm [154] used by Google to rank web pages in their original search engine results. PageRank is a Markov chain in which the states are pages and the transitions are the links between pages and are equiprobable. The PageRank ranking of a page is the probability of arriving at that page after a large number of clicks which is the inverse of the expected number of random clicks required to get from the page back to itself.

2.2.2 Matrix exponential

We have seen that a CTMC leaves us with a set of linear, constant coefficient ordinary differential equations to solve, given some initial condition $\boldsymbol{x}(0)$. This has a solution of the form $\boldsymbol{x}(t) = e^{\boldsymbol{Q}t}\boldsymbol{x}(0)$ where the exponential here is the matrix exponential. The matrix exponential can be defined as the convergent power series,

$$e^{\boldsymbol{X}} = \sum_{k=0}^{\infty} \frac{1}{k!}\boldsymbol{X}^k; \boldsymbol{X}_0 = I \tag{2.18}$$

and the solution can be expressed in terms of eigenvalues, λ_i, and eigenvectors, $\boldsymbol{v}_i$ of $\boldsymbol{Q}$,

$$\boldsymbol{x}(t) = \sum_{i=1}^{n} \alpha_i e^{\lambda_i t}\boldsymbol{v}_i \tag{2.19}$$

This expression is rarely analytically tractable therefore a numerical solution has to be sought with the particular choice of numerical method needing careful consideration [134], with scaling and squaring variants [56] considered the most effective in terms of compute time and error properties.

2.2.3 Stationary distribution of Markov Chains

A time-homogeneous discrete-time Markov chain described by left-stochastic (columns sum to one) matrix, $\boldsymbol{Q}$, has a solution $\boldsymbol{\pi}$ which is called a stationary distribution if $\forall i \in S$,

$$0 \leq \pi_i \leq 1 \tag{2.20}$$

$$\sum_{i \in S} \pi_i = 1 \tag{2.21}$$

$$\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\pi} \tag{2.22}$$

Likewise, a time-homogeneous continuous-time Markov Chain (CTMC), described by transition rate matrix $\boldsymbol{Q}$, has stationary distribution $\boldsymbol{\pi}$ if $\forall i \in S$ (2.20-2.21) hold and,

$$\boldsymbol{Q}\boldsymbol{\pi} = 0. \tag{2.23}$$

2.2.3.1 Existence

The existence of a stationary distribution is dependent on several properties of the Markov chain. Here we give definitions for a continuous time Markov chain, though the discrete time analogues and their subsequent results may be defined in an equivalent manner.

A Markov chain is irreducible if any state is reachable from any other state. More formally, A Markov chain is irreducible if :

$$\exists t > 0 \text{s.t.} \mathbb{P}(X(t) = j | X(0) = i) > 0 \forall (i, j) \in S \times S \tag{2.24}$$

A state, i, of a Markov chain is transient if there is a non-zero probability of never returning to i given we start at i. Defining the hitting time as T_i,

$$T_i = \inf\{t > 0 : X(t) = i | X(0) = i\} \tag{2.25}$$

allows us to write the probability that we return to state i for the first time at or before t as,

$$f_{ii}^{(t)} = \mathbb{P}(T_i \leq t | X(0) = i) \tag{2.26}$$

State i is transient if,

$$\mathbb{P}(T_i < \infty) < 1 \tag{2.27}$$

State i is recurrent if it is not transient. Even if a hitting time is finite with probability one, it need not have a finite expectation. If the mean recurrence time is finite then a chain is positive recurrent, otherwise it is null recurrent (null recurrent chains must have an infinite state space).

A finite state-space Markov-Chain has a stationary distribution if it is both irreducible and recurrent [194].

2.2.3.2 Finding the stationary distribution

The simplest way of calculating the stationary distribution computationally is to calculate the matrix exponential for large enough t that the stationary distribution is reached, though more efficient methods of computation exist. If the CTMC specified by $\boldsymbol{Q}$ has a stationary distribution then computation of the null eigenvector of $\boldsymbol{Q}$ - the eigenvector, $\boldsymbol{v}_1$, corresponding to the zero eigenvalue, λ_1 (which must exist if the

stationary distribution does) - gives the stationary distribution,

$$\dot{\boldsymbol{x}}(t) = \mathbf{0} \implies \boldsymbol{Q}\boldsymbol{x}(0) = \lambda_1 \boldsymbol{v}_1;\ \lambda_1 = 0 \tag{2.28}$$

$$\lambda_i < 0\ \forall i > 1 \implies \lim_{t\to\infty} e^{\lambda_i t} = 0 \tag{2.29}$$

$$\implies \lim_{t\to\infty} \boldsymbol{x}(t) = \lim_{t\to\infty} \sum_{i=1}^{n} \alpha_i e^{\lambda_i t} \boldsymbol{v}_i = \alpha_1 \boldsymbol{v_1} \tag{2.30}$$

with α_1 specified by the constraint that $\boldsymbol{x}$ must be a probability distribution ($||\boldsymbol{x}||^1 = 1$). Therefore any algorithm capable of solving this linear system will give the stationary distribution, with many algorithms existing that are much faster than the matrix exponential.

2.2.4 Simulation of stochastic processes

For some models the state space may be too complicated or large to compute the solution of a Markov process (or the process may not be Markovian) therefore a large number of individual simulations of the stochastic process are often performed in order to approximate the solution of the Master equations.

The Gillespie algorithm [69] generates a statistically correct trajectory with regard to the Master equations of the underlying stochastic process (algorithm 1). In brief, each iteration of the Gillespie algorithm chooses an event, E_i, to occur in proportion to the rate of each event happening along with a time step that is exponentially distributed with respect to the total sum of the rates and updates the state space time respectively. In the limit of infinite Gillespie simulations, the distribution of states over time will be equal to that of the solution of the Master equations. Referring back to the stochastic SIS model of section 2.1.2 for context, an event E_1 may be - depending on the chosen labelling - the recovery of an infectious individual. Event E_1 occurs at a rate R_1 with the system (which measures the number of infected individuals) updating such that $x(t+\delta_t) = x(t)-1$. The Gillespie algorithm is an example of a Monte-Carlo technique (section 2.3.2).

2.3 Bayesian Inference

Bayesian inference is a method of statistical inference in which Bayes' theorem is used to calculate the probability for a hypothesis. Bayesian probability/inference is conceptually different to the more traditional approach of frequentist probability/inference in that it interprets probability as representing a quantification of per-

Algorithm 1 Gillespie algorithm

Label all possible events $E_1, \ldots, E_n$
Choose stopping time T
Initialise model with initial conditions, $x_0 = x(t = 0)$
while t **do** $<$ T
    Compute rate $R_i(x(t))$ corresponding to event E_i for $i = 1, \ldots, n$
    $R_{\text{cum}}[i](t) = \sum_{j=1}^{i} R_j(x(t))$ ▷ Cumulative sum of rates
    $R_{\text{total}} = R_{\text{cum}}[n]$ ▷ Rate of any event
    $\delta_t \sim \exp(R_{\text{total}}(t))$ ▷ Exponentially distributed time to next event
    $r \sim \text{Uniform}(0, 1)$ ▷ Sample event in proportion to rate distribution
    Find smallest event index i s.t. $R_{\text{cum}}[i] \geq r\delta_t$
    Update state vector $x(t)$ according to E_i.
    $t+ = \delta_t$ ▷ Step forward in time
end while

sonal belief. Under the Bayesian methodology the probability assigned to a hypothesis is in the range from 0 to 1; however, a frequentist hypothesis must either be true or false.

2.3.1 Bayes' theorem

Bayes' theorem is derived from the laws of conditional probability and states that,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.31}$$

where A and B are events and $P(B) \neq 0$. $P(A)$ and $P(B)$ are the probabilities of observing A and B; $P(A|B)$ is the probability of observing event A given that B is true. Under the Bayesian interpretation of probability we may use Bayes' theorem to construct an inferential framework. For a proposition $\boldsymbol{\theta}$ and data $\mathcal{D}$,

- $P(\boldsymbol{\theta})$, the *prior*, denotes the initial degree of belief in $\boldsymbol{\theta}$.
- $P(\boldsymbol{\theta}|\mathcal{D})$, the *posterior*, is the degree of belief in $\boldsymbol{\theta}$ given $\mathcal{D}$.
- $P(\mathcal{D}|\boldsymbol{\theta})$, the *likelihood*, is probability of observing $\mathcal{D}$ given $\boldsymbol{\theta}$ is true.
- $P(\mathcal{D})$ is a normalisation constant.

For the purpose of this thesis, $\boldsymbol{\theta}$ represents a vector of parameters which will parameterise some model and $\mathcal{D}$ represents the data we have collected about our system of interest. Under the Bayesian inferential framework we wish to estimate the posterior, $P(\boldsymbol{\theta}|\mathcal{D})$; however this calculation is seldom analytically tractable as we have

no way to calculate the integral $P(\mathcal{D}) = \int P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})\,d\boldsymbol{\theta}$. As we often have no expression it is common to write,

$$P(\boldsymbol{\theta}|\mathcal{D}) \propto P(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta}) \tag{2.32}$$

With no analytical solution one resorts to numerical integration using Monte Carlo methods to obtain the posterior distribution and to evaluate expectations of quantities under the posterior distribution.

$$\mathbb{E}(g(\boldsymbol{\theta})) = \int_{\boldsymbol{\theta}} g(\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathcal{D})\,d\boldsymbol{\theta} \tag{2.33}$$

2.3.2 Simple Monte Carlo methods

Essentially, a Monte Carlo method [125] is an algorithm to approximate an expectation value using repeated random samples using numerical simulations. They are typically used for optimisation, numerical integration, and sampling from a complex probability distribution.

The power of Monte Carlo methods derives from the law of large numbers; the expected value (of some random variable) we wish to compute can be approximated by taking the sample mean of independent samples which under the law of large numbers will become asymptotically exact. For probability mass function (p.m.f) or probability density function (p.d.f) $f_X(x)$, random variable $X \in \Omega$,

$$X \text{ discrete:} \quad \mathbb{E}(g(X)) = \textstyle\sum_{x\in\Omega} g(x)f_X(x) \tag{2.34}$$

$$X \text{ continuous:} \quad \mathbb{E}(g(X)) = \textstyle\int_{x\in\Omega} g(x)f_X(x)\,dx \tag{2.35}$$

$$\mathbb{E}(g(X)) \approx \tfrac{1}{n}\textstyle\sum_{i=1}^{n} g(x_i) \tag{2.36}$$

But what use is estimating expectations? Well we may recast probabilities and integrals/summations as expectations. For example, probabilities may be expressed as $P(Y \in A) = \mathbb{E}(\mathbb{1}_{\{A\}}(Y))$, and below we show how integrals may be used.

Importance sampling is a Monte Carlo method for numerical integration where one chooses a "good" distribution, $h(\cdot)$, from which to simulate one's random variables in order to decrease the variance of the Monte Carlo estimate - by sampling random variables that have a larger impact on the parameter being estimated.

Say we wish to numerically integrate some quantity $r(x)$ for $x \in \Omega$, we can use importance sampling to reformulate the integral as the expectation of some quantity under our chosen distribution,

$$\int_{x\in\Omega} r(x)\,dx = \int_{x\in\chi} \frac{r(x)}{h(x)} h(x)\,dx = \mathbb{E}_h\left(\frac{r(X)}{h(X)}\right) \tag{2.37}$$

$h(\cdot)$ is introduced into the integrand of interest by sampling X from $h(x)$ and computing the expectation of the likelihood ratio $\frac{r(x)}{h(x)}$ rather than solely $r(x)$ in order to obtain an unbiased sampling estimator.

$$\mathbb{E}_h\left(\frac{r(X)}{h(X)}\right) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{r(X_i)}{h(X_i)}, \text{ where } X_i \sim h(x) \tag{2.38}$$

$r(x)$ is an arbitrary function that often takes the form of a standard expectation. For example, $r(x)$ may correspond to the quantity $\mathbb{E}(g(x))$ (with $X \sim f_X$) which could represent something such as the expected number of infections given some transmission parameter(s) x. Then our integral would correspond to,

$$\mathbb{E}(g(X)) = \int g(x) f_X(x)\,dx = \mathbb{E}_h\left(\frac{g(x) f_X(x)}{h(x)}\right) \tag{2.39}$$

A good importance sampling function, $h(x)$, should be: positive whenever $r(x)$ is non-zero; close to being proportional to $|r(x)|$; easy to compute the density of $\forall x \in \Omega$. In one or two dimensions importance sampling provides an elegantly efficient way of evaluating the numerical integrals for Bayesian inference; however choosing distributions that satisfy these properties in higher dimensions is highly non-trivial. Defining a function that is both non-zero everywhere it needs to be and close to being proportional to the posterior across more than a handful of dimensions simultaneously becomes an exponentially harder constraint to satisfy. For Bayesian inference being able to specify such a function amounts to having a very highly informed prior distribution - which one rarely has. The quest for methods that perform better in higher dimensions leads us to Markov Chain Monte Carlo methods.

2.3.3 Markov Chain Monte Carlo (MCMC)

Consider the same integral as before,

$$\int g(x) f_X(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i); x_i \sim f_X \tag{2.40}$$

Instead of sampling x_i from some distribution $h(x)$, with MCMC we wish to define a stochastic process $\{X_t : t \in T\}$ that is a Markov Chain ($P(X_t|X_1,\ldots,X_{t-1}) = P(X_t|X_{t-1})$) with stationary distribution $\pi(\cdot) = f_X$. This means that by running this Markov chain we draw samples from f_X in order to compute our expecta-

tion/integral. In the case of Bayesian inference we are defining a Markov Chain with a stationary distribution that is equal to our posterior.

A Markov Chain is uniquely defined by its transition probabilities, $P(x'|x)\forall(x', x) \in \Omega \times \Omega$. It has a stationary distribution (section 2.2.3) when the detailed balance condition, $\pi(x)P(x'|x) = \pi(x')P(x|x')$, is met (sufficient but not necessary) which requires that the probability of being in state x and transitioning to state x' is equal to the probability of being in state x' and transitioning to state x. Furthermore, the stationary distribution of a Markov Chain (if it exists) is unique if the Markov Chain is ergodic which requires that the expected number of steps for returning to the same state is finite (positive recurrence) and that the system does not return to the same state at fixed intervals (aperiodicity).

2.3.3.1 Metropolis-Hastings algorithm

The first MCMC algorithm is the Metropolis-Hastings algorithm [126], and is still one of the most commonly used algorithms today. It designs a Markov chain which is ergodic and reversible by rewriting the condition of detailed balance as,

$$\frac{P(x'|x)}{P(x|x')} = \frac{\pi(x')}{\pi(x)} \tag{2.41}$$

$P(x'|x)$ is separated into a proposal distribution, $q(x'|x)$, the conditional probability of proposing a state x' given state x, and the acceptance distribution, $\alpha(x'|x)$, the conditional probability to accept the proposition x'.

$$P(x'|x) = q(x'|x)\alpha(x'|x) \tag{2.42}$$

$$\frac{\alpha(x'|x)}{\alpha(x|x')} = \frac{\pi(x')}{\pi(x)}\frac{q(x|x')}{q(x'|x)} \tag{2.43}$$

If $\alpha(\cdot)$ is chosen such that the condition above is met then we will sample from $\pi(\cdot)$, with the random-walk Metropolis choice being,

$$\alpha(x'|x) = \min\left(1, \frac{P(x')}{P(x)}\frac{q(x|x')}{q(x'|x)}\right) \tag{2.44}$$

$$q(x'|x) \sim \mathcal{N}(x, \sigma^2\Lambda) \tag{2.45}$$

for some choice of scale factor σ^2 and covariance matrix Λ.

We are now equipped with an algorithm that allows us to simulate a Markov Chain with a stationary distribution equal to our posterior distribution. It is impor-

Algorithm 2 Metropolis Hasting's algorithm

$X_0 \sim$ Prior distribution ▷ Initialise
for t=1,2, ... **do**
  $X' \sim q(X_t|X_{t-1})$ ▷ Proposal
  $\alpha(X'|X_{t-1}) = \min\left(1, \frac{q(X_{t-1}|X')\pi(X')}{q(X'|X_{t-1})\pi(X_{t-1})}\right)$ ▷ Acceptance probability
  $u \sim \text{Uniform}(0, 1)$
  if $u < \alpha$ **then**
    $X_t \leftarrow X'$ ▷ Accept
  else
    $X_t \leftarrow X_{t-1}$ ▷ Reject
  end if
end for

tant to note that this is only an asymptotic guarantee. How many samples from our stochastic process do we need to accurately approximate the posterior? We briefly outline how to apply MCMC in practice and refer the reader to [36, 65] for a full treatment of the complexities of applying MCMC in practice.

Each choice of hyper-parameter (σ, Λ) - the particular hyper-parameters will vary depending on which specific MCMC algorithm is used but we will assume random-walk MH for this discussion - specifies a different stochastic process all of which have stationary distribution $\pi(\cdot)$ but some of which will converge faster than others (in the sense that the estimates of the posterior means and standard deviations will achieve a given error threshold faster). If the random walk step sizes are too large then few moves will be accepted resulting in a small acceptance rate. A small acceptance rate is undesirable as this means that many of our samples will be identical resulting in very slow convergence. If the opposite is true and the step sizes are too small then most moves will be accepted; however the samples will be very highly auto-correlated. High auto-correlation results in poor convergence as the number of samples it takes to draw a sample 'far away' from the current position, e.g. one posterior standard deviation, becomes incredibly large. Choosing hyper-parameters that specify an efficient Markov chain is no easy feat but despite many referring to this act as a 'black art' principled ways of making this choice do exist (section 2.3.3.3) but first we discuss how to measure convergence given this choice has already been made.

2.3.3.2 MCMC convergence

Given some choice of hyper-parameters, we need a principled way of knowing how long to run a chain for. There is no fixed answer as it depends on the attributes

of the particular posterior being inferred with a large number of parameters and correlations in the posterior joint distributions making longer chains necessary and the reality is that 'You can't prove convergence, at best you fail to prove a failure to converge'. Put simply, we can never be sure that convergence has been achieved, the only time we can be certain of the state of convergence is when a chain has failed to converge *and* we detect this through a failed convergence test.

Firstly, before considering when to stop we need to decide when to start. After initialisation the chain must be run long enough until it has converged to an area of high posterior probability before we begin counting samples. If we theoretically ran our chain for an infinite number of iterations then this would not be necessary. With only have a finite number of samples then starting from an arbitrary position biases the chain towards unlikely values. To avoid this bias we need to run the chain for long enough until it has 'forgotten' where it started so that the first sample we count is a true posterior sample. This throwing away of samples is referred to as the 'burn-in' period and can be informed by the Geweke diagnostic [67].

Secondly, there is typically a very high correlation between subsequent iterations such that the number of independent draws or effective sample size (ESS) can be far less than the number of iterations of the simulation. It is therefore necessary to quantify this auto-correlation in order to calculate the ESS so that accurate estimates of the Monte-Carlo error can be calculated. Often practitioners will also use it to calculate a relevant thinning parameter, k, taking one in every k sample iterations for any posterior. In reality thinning is not necessary unless storage becomes a problem [119].

When using iterative simulation algorithms such as MCMC it is therefore of utmost importance that care is taken to avoid the pitfalls mentioned. One way to overcome this is to employ more efficient algorithms; however to measure the performance properties of an algorithm we require diagnostics to analyse multiple simulation runs with dispersed starting points in parameter space by comparing the variation within and between each chain with the desirable property that within variance is roughly equal to between variance.

It is important to look at between chain information as we may have two chains which alone look stationary, but together it is obvious they have not converged to the same distribution, perhaps due to either a step size that is too small or a bimodal target distribution with each chain sampling from a different mode. Likewise it is necessary to look at the within chain information as we may have two chains that appear to sample from a common distribution, however looking within each chain neither is stationary.

One can simultaneously check mixing and stationarity by splitting each chain in half and check all the half-chains have mixed. For m half-chains each of length n, define the samples of a given parameter to be tested as $\theta_{ij}(i = 1, \ldots, n; j = 1, \ldots, m)$ and compute the between and within chain variances, B and W:

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\theta}_{.j} - \bar{\theta}_{..})^2 \tag{2.46}$$

$$W = \frac{1}{m} s_j^2 \tag{2.47}$$

where,

$$\bar{\theta}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \theta_{ij} \tag{2.48}$$

$$\bar{\theta}_{..} = \frac{1}{m} \sum_{i=1}^{m} \bar{\theta}_{.j} \tag{2.49}$$

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\theta_{ij} - \bar{\theta}_{.j})^2 \tag{2.50}$$

The marginal posterior variance of the estimand (2.51) may be expressed as a weighted average of W and B. This overestimates the true variance (assuming an appropriately dispersed starting distribution) but is unbiased under stationarity. W should be an underestimate as the individual chains are less likely to have sampled the full range of the target distribution. 'Convergence' of the chains is monitored by estimating the factor, $\hat{R}$, (2.52) [63] by which the scale of the current distribution might be reduced if simulations were continued for $n \to \infty$ which naturally declines to 1 in the limit - $\hat{R}$ is also known as the Gelman-Rubin convergence diagnostic. The prevailing wisdom in the literature is that $\hat{R}$ should be less than 1.1 [63, 65] for all estimands but the recent work of Vehtari et al. [197] analysing the conditions under which $\hat{R}$ breaks down suggest the stricter threshold of 1.01 (calculated with at least 4 chains).

$$\hat{V} = \hat{\text{var}}^{+}(\theta|\mathcal{D}) = \frac{n-1}{n} W + \frac{1}{n} B \tag{2.51}$$

$$\hat{R} = \sqrt{\hat{V}/W} \tag{2.52}$$

This can be extended to the multivariate case where W, B, and $\hat{V}$ now

represent covariance matrices and the potential scale reduction factor of the variance is shown in Brooks and Gelman [35] to be:

$$\hat{R} = \sqrt{\max_a \frac{a^T \hat{V} a}{a^T W a}} = \sqrt{\frac{n-1}{n} + \frac{m+1}{m}\lambda_1} \tag{2.53}$$

Where λ_1 is the largest eigenvalue of $W^{-1}B/n$. This method has the advantage of ensuring the multivariate distribution has achieved approximate convergence, a stricter measure than whether each individual parameter has achieved approximate convergence.

The $\hat{R}$ scale reduction factor answers half the question about when we can stop sampling but as eluded to earlier we also need to know the ESS in order to estimate the Monte-Carlo errors ($\propto \frac{1}{\sqrt{N}}$). ESS calculation for correlated samples is done by considering the efficiency of the average of the simulations $\bar{\theta}_{..}$ as an estimate of the posterior mean $\mathbb{P}(\theta|y)$. Starting with the asymptotic formula for the variance of the average of a correlated sequence:

$$\lim_{n\to\infty} mn\text{var}(\bar{\theta}_{..}) = (1 + 2\sum_{t=1}^{\infty} \rho_t)\text{var}(\theta|y) \tag{2.54}$$

ρ_t is the auto-correlation of θ at lag t. Independent draws would results in $\text{var}(\bar{\theta}_{..}) = \frac{1}{mn}\text{var}(\theta|y)$ and the (effective) sample size would be mn but in the presence of correlation is,

$$n_{\text{eff}} = \frac{mn}{(1 + 2\sum_{t=1}^{\infty} \rho_t)} \tag{2.55}$$

In reality we cannot exactly estimate the ESS as we must estimate the sum of the correlations. Use between and within chain samples, start by computing var^+ and then compute the variogram V_t for each lag t.

$$\hat{\rho}_t = 1 - \frac{V_t}{2\text{var}^+} \tag{2.56}$$

For large t the correlation is too noisy therefore a partial sum is computed until T, where T is the first odd lag for which two successive lags are negative (though other choices exist).

$$\hat{n}_{\text{eff}} = \frac{mn}{(1 + 2\sum_{t=1}^{T} \hat{\rho}_t)} \tag{2.57}$$

2.3.3.3 Hyper-parameter choices

Adaptive versions of MCMC algorithms exist which algorithmically tune the hyper-parameters to improve sampling efficiency by using the history of the Markov chain to tune the hyper parameters of the proposal distribution [168]. These adaptive proposals introduce an element of danger into the MCMC process if not used correctly as adaptive proposals destroy the Markov property and thus ergodicity of our algorithm which is required in order for a unique stationary distribution to exist. Fortunately several methods of maintaining ergodicity exist with the most common two being simply stopping adaptation after some time (though this requires user intervention) and 'controlled MCMC' via. diminishing adaptation [167].

The first such attempt at defining an adaptive MCMC algorithm that we are aware of is the work of Haario et al. [77]. Adaptation occurs by continuously adapting the target distribution to estimate the covariance of the Gaussian proposal distribution centred on the current state. At time $t+1$ we have already sampled $X_0, X_1, \ldots, X_t$. The new proposal distribution for the next step is a Gaussian distribution with mean X_t and covariance given by $C_t = \sigma^2 \Lambda$, where Λ is the covariance matrix determined by the spatial distribution of $X_0, X_1, \ldots, X_t \in \mathbb{R}^d$. The scaling parameter σ depends only on the dimension d of the posterior in order to give an appropriately scaled Gaussian approximation of the target distribution.

$$\alpha(X_t, X') = \min\left(1, \frac{\pi(X')}{\pi(X_t)}\right) \tag{2.58}$$

The acceptance criteria (2.58) resembles that of the normal Metropolis-Hastings algorithm; however here the choice for acceptance probability is not based on symmetry (reversibility) conditions as the exactness of simulation must be studied separately. Indeed, theorem 1 of Haario et al. [77] verifies that the algorithm possesses the correct ergodicity properties required to sample from our posterior despite the chain not being Markovian. Haario et al. [77] exploit the fact that the asymptotic dependence between the elements of the chain is weak enough to apply known theorems of large numbers for Mixingales [124].

The tuning and updating of the covariance matrix occurs according to,

$$C_{t+1} = \begin{cases} C_0 & : t \leq t_0 \\ \sigma^2 \mathrm{cov}(X_0, X_1, \ldots, X_t) + \sigma^2 \epsilon I_d & : t > t_0 \end{cases} \tag{2.59}$$

where $\epsilon > 0$ and small, ensures C_t will not become singular; $t_0 > 0$ is a free choice but adaptation will be slower the larger it is (reflects trust in initial covariance C_0); and $\sigma^2 = (2.4^2)/d$ [64]. For $t \geq t_0 + 1$, C_t may be efficiently calculated with the

recursion formula,

$$C_{t+1} = \frac{t-1}{t}C_t + \frac{s_d}{t}\left(t\bar{X}_{t-1}\bar{X}_{t-1}^T - (t+1)\bar{X}_t\bar{X}_t^T + X_tX_t^T\right) + \epsilon I_d \tag{2.60}$$

Whilst Haario et al. [77] use a mixingale approach, the later work of Roberts and Rosenthal [167] presented simpler conditions which ensure ergodicity and stationarity - that an adaptive MCMC algorithm satisfies the conditions of *diminishing adaptation* and *bounded convergence*. Diminishing adaptation requires that the amount of adaptation decreases as $t \to \infty$. In practice, diminishing adaptation is implemented by either modifying the adaptation parameters less overtime or by performing adaptations with smaller probabilities over time. Bounded convergence requires that the convergence time of a transition kernel for a given set of adaptive parameters is bounded in probability [168] - this is most easily satisfied whenever the joint state space of the chain and stochastic process defining the adaptation are finite.

2.3.4 Bayesian Model selection

When fitting a Bayesian model we need a way to evaluate and compare the predictive accuracy of models. Several approaches exist of which we will briefly discuss the most popular.

2.3.4.1 Bayes' factors

The Bayes' factor [96], K, assesses the plausibility of two models, M_1 and M_2, parameterised by θ_1, θ_2, and is given by,

$$K = \frac{\mathbb{P}(D|M_1)}{\mathbb{P}(D|M_2)} = \frac{\mathbb{P}(M_1|D)\mathbb{P}(M_2)}{\mathbb{P}(M_2|D)\mathbb{P}(M_1)} \tag{2.61}$$

The Bayes' factor simplifies to the ratio of posterior probabilities of M_1 and M_2 given that the two models are equally likely a priori. Bayes' factors are not hugely popular due to several shortcomings. They are not well defined for improper priors; they make an implicit assumption that one model is correct (which is almost never the case); and they are heavily dependent on the prior.

2.3.4.2 Information criteria

An alternative set of model selection techniques are information criteria, which have their roots in information theory. The Akaike information criterion (AIC) [3] is the first and best known example, it estimates the information lost when a given model is

used to represent the process that generated the data. AIC, and information criteria in general reward goodness of fit (using the likelihood function) and penalise model complexity (most commonly using the effective number of parameters). Information criteria are asymptotic approximations as the sample size becomes large. Given a statistical model with k estimated parameters, and maximum likelihood value $\hat{L}$ the AIC is given by,

$$\text{AIC} = 2k - 2\ln(\hat{L}) \tag{2.62}$$

Being based on the maximum likelihood value AIC is not a Bayesian measure. Therefore several efforts to come up with a Bayesian information criterion were made with perhaps the most popular being the Deviance Information Criterion (DIC) [182]. Its popularity may be attributed to the fact that it is extremely convenient to compute for Bayesian model selection problems where the posterior has been obtained using MCMC because it is calculated based on the likelihoods of the sampled posterior. Define the deviance as $D(\theta) = -2\log(\mathbb{P}(y|\theta)) + C$ for data y, parameters θ, and constant C (cancels out). p_D then gives the effective number of parameters which may be chosen to be one of $p_D = \bar{D} - D(\bar{\theta})$ or $p_D = 0.5\hat{\text{var}}(D(\theta))$.

$$\text{DIC} = p_D + \bar{D}(= D(\theta) + 2p_D) \tag{2.63}$$

A major shortcoming of DIC is that it tends to select over-fitted models as the observed data are used to both construct the posterior distribution and evaluate models. Ando [7] introduced the Bayesian predictive information criterion (BPIC) to overcome the issues with DIC; however in practice BPIC is impractical to calculate. To this end Ando [8] developed a new information criterion which can be easily calculated which is equivalent to DIC with double the model penalty term. This information criterion seems to have not been named in the literature or is often confused with BPIC - we shall refer to it as DICc.

One final information criterion, the Watanabe-Akaike information criterion [201] (WAIC - colloquially referred to as the Widely-Applicable-Information-Criterion), is the generalisation of the AIC onto singular statistical models (singular Fisher information matrix/posterior distribution is not normal/log-likelihood not approximated by quadratic form - e.g. hidden Markov model). More importantly, unlike AIC and DIC the WAIC averages over the posterior distribution rather than conditioning on a point estimate meaning it can be said to be fully Bayesian [66].

$$\text{WAIC} = -2(\text{LPPD} - p_D) \tag{2.64}$$

$$\text{LPPD} = \sum_{i=1}^{N_{obs}} \log\left(\sum_{j=1}^{n} \mathbb{P}(y_i|\theta_j)/(n)\right) \tag{2.65}$$

$$p_D = \sum_{i=1}^{N_{obs}} \hat{var}(\{\log(\mathbb{P}(y_i|\theta_j))/(n) : j = 1, \ldots, n\}) \tag{2.66}$$

Where LPPD denotes the log posterior predictive density and n denotes the chain length.

Whilst WAIC could be said to be the best practically computable information criterion, it is still an asymptotic approximation. Without losing too much of the convenience of WAIC/DIC we may turn to leave-one-out Cross-validation (LOO) which becomes asymptotically equivalent to WAIC, and therefore will provide us with a better model selection criterion [196].

Exact cross-validation requires re-fitting the model with different training sets which would be impractical; however approximate leave-one-out cross-validation (LOO) may be performed using importance sampling. LOO results in a noisy estimate as the variance of the importance weights may be large. Therefore Pareto Smoothed Importance Sampling [195] (PSIS) is used to overcome these shortcomings by fitting a pareto distribution to the upper tail of the importance weight distribution. WAIC and PSIS-LOO can be seen as improvements over DIC - an already popular and pragmatic model selection choice. On the other hand, in the context of modelling the spread of infectious diseases PSIS-LOO may not offer an improvement over WAIC as the independence assumptions inherent in the use of leave one out cross-validation are not valid in the context of epidemics - leaving out a data-point 'censors' the chain of transmission. PSIS-LOO is only used in the context of chapter 4 where this assumption is already violated by the chosen models, and should in general be avoided in epidemic modelling.

CHAPTER 3

Epidemics on degree heterogeneous clustered networks

3.1 Introduction

Networks offer an unprecedented opportunity to represent and model contacts between interacting units at all scales ranging from proteins and individuals to countries via air transportation. This additional degree of freedom has led to extensive modelling and analysis in mathematical epidemiology and has allowed the development of a number of network-based models, which are either parameterised by real network data, if available, or by using synthetic networks which reflect and reproduce some observed local or global properties of real-world networks [45, 98, 110, 156]. Idealised networks often offer a greater degree of analytical tractability, which in turn offers clearer insight into the impact of network properties on epidemic outbreak threshold, final epidemic size, prevalence, and on effectiveness or choice of control measures.

The many degrees of freedom offered by networks come at the cost of computational and mathematical complexity. In order to handle this, and for a systematic investigation and understanding of network processes, it is desirable not to rely solely on stochastic simulations. As a result, a number of differential equation-based models have been developed including pairwise [89, 97, 161, 185], PGF or edge-based compartmental models [129, 132] and effective degree models [73, 117], to name just a few. The goal of all these models is the same: to derive a low-dimensional system of ordinary differential equations (ODEs) where variables correspond to some average quantity from the stochastic process—e.g. expected prevalence— over time. All such mean-field models require choice of a 'state space'. For example, pairwise models initially concentrate on the expected number of nodes in different states, while effective degree models concentrate on the expected number of star-like structures, i.e. a central node with all its neighbours, and the possible state that these can be in.

Once chosen, evolution equations for these variables are derived. This, often heuristic, step still involves a precise book-keeping which in general yields a dependency on higher order states or moments, e.g. for pairwise models the expected number of infected nodes depends on the expected number of edges where one node is infected and the other is susceptible. These newly introduced higher-order structures or moments require further equations and hence to curtail the dependency on higher-order moments and the fast growth in the number of equations, 'closures' are needed. This amounts to approximating higher-order moments in terms of lower order ones. The performance of mean-field models are often tested by direct comparison to results based on explicit stochastic network simulations. If the process is successful and the mean-field model works well, the analysis of the stochastic epidemic on networks is mapped into analysing a system of ODEs. This can be done using dynamical systems tools and such analysis often leads to analytical results which explicitly reveal the interaction between network and disease characteristics. More importantly, it shows how the fundamental properties of the network impact on growth rate, epidemic threshold, final epidemic size and so on.

Degree or contact heterogeneity and degree-based mixing is well accounted for in existing differential equation models, but epidemics on networks which are clustered (i.e. where two nodes with a common neighbour are highly likely to be neighbours of each other [202]) pose more of a challenge. A major factor of this difficulty is the non-unique way in which global or network-level clustering can be achieved, and the same level of clustering can be achieved while keeping the degree distribution the same but using different distributions or different sets of motifs [74, 109, 142, 163].

In general, clustered bond percolation-type [72, 95, 130, 144] or PGF-based models [89, 164, 198] consider specific forms of clustering (e.g. non-overlapping triangles) while pairwise models [87, 89, 91, 97, 101] usually work best in the case where clustering is 'randomly' distributed, as is the case when using rewiring algorithms such as the big-V [19, 74, 90], which precludes certain kinds of interaction between clustering and degree heterogeneity. The question of how to create better models of epidemic dynamics on clustered networks remains open, and is the topic of this chapter.

For clustered networks it is difficult to derive accurate differential equation models for epidemics. In this chapter, we present an approach based on generalisation of the effective degree model [117] to clustered networks, and we show that this newly-derived model outperforms the state-of-the-art models and displays excellent agreement with results based on stochastic simulations for a range of degree

distributions and clustering values.

The chapter is structured as follows: In section 3.2 an overview of relevant complex network concepts and terminology are presented; section 3.3 introduces methods of generating clustered networks with a specific degree distribution; in section 3.4 existing ODE models are introduced and discussed; section 3.5 presents the extension of ODE models to dynamics on clustered networks including our novel ODE model; section 3.6 details the simulation study performed to compare between models and tests the closure performance against error properties conjectured by [157].

This chapter is quite notationally dense, table 1 provides an overview of the conventions used.

3.2 Model Definitions

3.2.1 Networks

A *network* (or *graph*) is a pair $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a size-N set of *nodes* and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ is a set of *links*. The information about a network can be usefully encoded in terms of an *adjacency matrix*, which has elements $G_{ab} = 1$ if $(a, b) \in \mathcal{L}$, and zero otherwise.

We will consider simple undirected networks without self edges meaning that $G_{aa} = 0$ and $G_{ab} = G_{ba}$, $\forall a, b \in \mathcal{N}$. The *degree* of node a is the number of links that it participates in, i.e.

$$k_a := \sum_{b \in \mathcal{N}} G_{ab} \ . \tag{3.1}$$

We assume that all degrees are integers between 1 and the *maximum degree* M. We will use angled brackets to refer to mean values of functions of degree across the network, i.e. for an arbitrary function f,

$$\langle f(k) \rangle := \frac{1}{N} \sum_{a \in \mathcal{N}} f(k_a) \ . \tag{3.2}$$

One particularly important such expectation is the *probability generating function* (PGF) which is $\psi(x) := \langle x^k \rangle$.

A final network property of interest in this work is the *clustering coefficient*,

$$\phi := \frac{\sum_{a,b,c \in \mathcal{N}} G_{ab} G_{bc} G_{ca}}{\sum_{a,b,c \neq a \in \mathcal{N}} G_{ab} G_{bc}} \in [0, 1] \ . \tag{3.3}$$

Our interest is in networks with large size ($N \gg 1$), degree distributions with a non-infinite variance ($\langle k \rangle^2 \leq \langle k^2 \rangle < \infty$), and clustering coefficients that can be non-zero but are not particularly large ($\phi \in [0, 0.3]$).

3.2.2 Epidemic Dynamics

We will consider SIR (Susceptible–Infected–Removed) dynamics. At the individual level, an individual a has a random state $X_a \in \{S, I, R\}$. Individuals in state I move to state R at rate γ, and individuals in state S move to state I at a rate τ multiplied by the number of I individuals they are linked to on the network.

At the population level, we will consider expected total node, pair and triple counts in given disease state configurations,

$$[A] = \mathbb{E} \sum_{a \in \mathcal{N}} \mathbf{1}_{\{X_a = A\}} \ , \tag{3.4}$$

$$[AB] = \mathbb{E} \sum_{a,b \in \mathcal{N}} \mathbf{1}_{\{X_a = A \& X_b = B\}} G_{ab} \ , \tag{3.5}$$

$$[ABC] = \mathbb{E} \sum_{a,b,c \neq a \in \mathcal{N}} \mathbf{1}_{\{X_a = A \& X_b = B \& X_c = C\}} G_{ab} G_{bc} \ , \tag{3.6}$$

where $A, B, C \in \{S, I, R\}$ and $\mathbf{1}$ is the indicator function. We will distinguish notationally between closed and open triples,

$$[ABC]_\triangle = \mathbb{E} \sum_{a,b,c \in \mathcal{N}} \mathbf{1}_{\{X_a = A \& X_b = B \& X_c = C\}} G_{ab} G_{bc} G_{ca} \ , \tag{3.7}$$

$$[ABC]_\wedge = [ABC] - [ABC]_\triangle \ , \tag{3.8}$$

which will be particularly important later. We will also consider more detailed states—in particular $[A_{s,i}]$ represents the expected number of nodes in state A with s susceptible and i infected contacts. Another important definition is the *correlation* between states

$$\mathcal{C}_{AB} = \frac{N}{\langle k \rangle} \frac{[AB]}{[A][B]} \ , \tag{3.9}$$

which expresses how much more likely an $[AB]$ edge is over the null model. The understanding of this is key to the role of networks in shaping epidemic dynamics [102].

Our aim in this chapter is to find methods for approximation of the expected population-level behaviour of the epidemic dynamics that are, as much as possible, logically consistent, well motivated, and accurate.

3.3 Models of Network Generation

Typically, a full epidemic network is not directly available from data and so a standard method is to work with probabilistic models for the network that respect certain observable summary statistics—in our case, the degree distribution and clustering coefficient. We will now present several such algorithms in outline form—more detail on these can be found in relevant papers and textbooks, e.g. Newman [143].

In n-regular networks, all nodes have the same degree i.e. $k_a = n, \forall a \in \mathcal{N}$. A typical such network can be generated by starting with a network that is atypical but assuredly n-regular, for example a one-dimensional $(n/2)$-nearest neighbour ring. The network is then rewired by removing two edges (A, a) and (B, b) sampled (without replacement) uniformly at random, and then adding new edges (A, B) and (a, b) (figure 3.1 ④ → ⑤). Performing a large number of such 'edge swaps' will result in a n-regular network that is representative of this class of graphs.

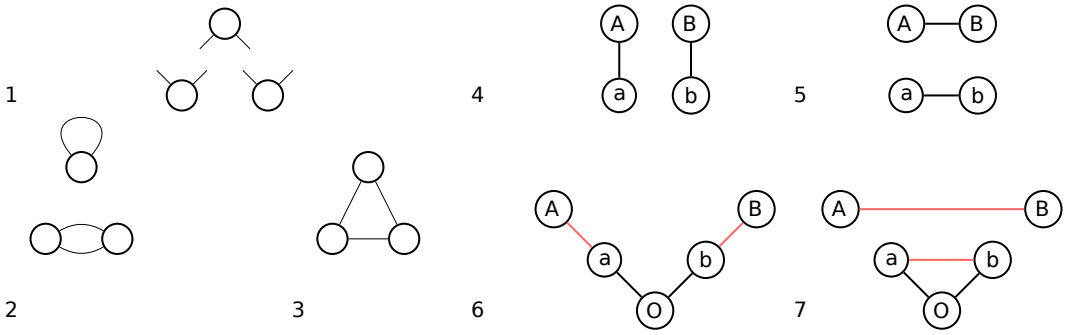

Figure 3.1: Illustration of network rewiring techniques. The configuration model starts as a set of nodes with half-edges (①) and constructs a network by pairing these up to form a full-edge at random, resulting in configurations like ② & ③. ② shows an example where the configuration model can lead to self and duplicate edges, whereas ③ gives a valid configuration. ④ → ⑤ will rewire a network such that local structure is lost - performing many of these edge swaps yields a network with the same degree distribution, negligible clustering, negligible degree assortativity, and introduces the small world property to a lattice. The clustering of a network may be increased by looking for V-shaped configurations (⑥) and performing the rewiring ⑥ → ⑦ if it increases the overall clustering coefficient of the network.

Erdős-Rényi networks [68], often referred to simply as 'random graphs' due to their importance, are defined such that each possible link is present with independent probability q. This leads to a binomial degree distribution although practically this will always be very well approximated by a Poisson distribution with mean $\langle k \rangle = (N-1)q$.

In the configuration model (CM), nodes are given a number of 'stubs', which are then paired uniformly at random [135]. This construction is useful for the development of asymptotic results, but can cause problems for a given finite value of N due to the presence of 'defects'—self-edges, multiple links between nodes, and stubs that are not eventually paired with others (figure 3.1 ②-③).

Networks with a given degree distribution can be generated following the procedure described in Del Genio et al. [48], which generates a sample of a given degree distribution that is graphical (meaning that a network with this degree distribution can be constructed without defects such as self or multiple edges) and then directly samples a statistically independent network. This algorithm is better behaved than the Configuration Model as it always produces a simple graph without defects, backtracking or rejections.

Algorithms that generate clustered heterogeneous networks using one random process typically generate special network topologies. We therefore generate clustered networks via a two-step process in which we first generate a network of the required degree distribution, and then increase the clustering coefficient using a rewiring method known as the 'big-V' [19, 74, 90]. This involves finding a 'V' configuration in the network (figure 3.1 ⑥) and proposing a rewiring which produces a triangle and a separate edge (figure 3.1 ⑦). This rewiring is then performed provided the clustering coefficient is increased. This method preserves the degree distribution whilst increasing the clustering coefficient (empirically) up to between $\phi = 0.3$ and $\phi = 0.4$ before the acceptance ratio of proposed moves critically slows down. The origin node, O, is selected with a weight proportional to $k_O(k_O - 1)$, where k_O denotes its degree, so that the expected proportion of possible triangles present around each node does not depend on its degree [178].

3.4 Epidemic Dynamics on Locally Tree-like Networks

In the limit as $\phi \to 0$ while $N \to \infty$, it is possible to obtain results for the expected population-level epidemic dynamics that are known to be asymptotically exact [20].

The following four sub-sections introduce an approach from the literature, with these approaches varying in complexity, accuracy, and intuitiveness. This allows section 3.5 to review extensions for two of these models, and introduce a novel extension for the effective degree model such that they better model dynamics on clustered networks. As the model we develop in section 3.5.3 relies on the formulation of the effective degree model, we cover the intuition behind the effective degree model to a greater extent than the other models of this section.

3.4.1 Simple SIR Model

It is possible to show [180] that the following equations hold for an arbitrary network:

$$\frac{\mathrm{d}}{\mathrm{d}t}[S] = -\tau[SI] \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}[I] = \tau[SI] - \gamma[I] \ . \tag{3.10}$$

There are two limits in which we can accurately approximate the pair variable $[SI]$ in terms of the node variables: (i) for an n-regular graph, as $n \to N-1$; and (ii) for an ER graph with mean degree $\langle k \rangle$. In each case we take $[SI] \approx \langle k \rangle [S][I]/N$ with $\beta := \tau \langle k \rangle$ to end up with a special case of the classic model in mathematical epidemiology introduced almost a century ago by Kermack and McKendrick [105], often called the simple SIR model,

$$\frac{\mathrm{d}}{\mathrm{d}t}[S] = -\frac{\beta}{N}[S][I] \ , \qquad \frac{\mathrm{d}}{\mathrm{d}t}[I] = \frac{\beta}{N}[S][I] - \gamma[I] \ . \tag{3.11}$$

Homogeneous mixing in a population is a poor assumption and more structural information is required to focus on the underlying network of contacts.

3.4.2 Pairwise Model

The pairwise model was one of the first steps toward more realistic contact structure, and primarily concerns n-regular graphs. In this approach, one continues to write down equations of the form (3.10) for higher order structures—go from expressing the evolution of single nodes in terms of single nodes and pairs of connected nodes to expressing the evolution of pairs of nodes in terms of triplets of connected nodes and pairs of nodes,

$$\frac{\mathrm{d}}{\mathrm{d}t}[SS] = \ -2\tau[ISS], \tag{3.12}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[SI] = \tau\Big([ISS] - [ISI] - [IS]\Big) - \gamma[IS], \tag{3.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[II] = 2\tau\Big([ISI] + [IS]\Big) - 2\gamma[II] \ . \tag{3.14}$$

For instance, (3.12) expresses the fact that the number of Susceptible-Susceptible ($[SS]$) pairs in the network decreases at a rate based on the number of Infected-Susceptible-Susceptible triplets ($[ISS]$) in the network, multiplied by two times the per-link transmission rate (τ). The factor of two arises from the fact that pairs are ordered such that $[IS] = [SI]$, thus the rate of change of symmetric terms involves a double contribution.

The closed pairwise model [102, 160] is gained by taking the unclosed ODEs (3.10 and 3.12–3.14) and approximating the number of triples, $[ABC]$, in the system using a moment closure originally attributed to Kirkwood [108],

$$[ABC] \approx \frac{\langle k \rangle - 1}{\langle k \rangle} \frac{[AB][BC]}{[B]} \ . \tag{3.15}$$

The intuition behind the Kirkwood closure lies in trying to express the proportion of links out of the B node in an $A - B$ pair that will connect to a C node. On average, there are $(\langle k \rangle - 1)[AB]$ free-links out from the B nodes of $A - B$ pairs. The proportion of those free-links connecting to C nodes is the quotient of the number of $B - C$ links, $[BC]$, and the total number of B links, $\langle k \rangle [B]$. Putting these terms together yields the Kirkwood closure.

3.4.3 Effective Degree (ED) Model

Ball and Neal [17] introduced the notion of an effective degree. In the Ball and Neal construction, individuals (nodes) begin with a number of unpaired half-links/stubs—an *effective degree*—and a contact network is constructed along with SIR epidemic dynamics.

The dynamics occur as follows: stubs of infected nodes randomly connect with stubs of susceptible nodes at rate τ, reducing the effective degree of both nodes by one and infecting the susceptible node; infected nodes become recovered at rate γ, at which point they connect their remaining stubs uniformly at random to any remaining unpaired stubs in the network thus reducing their effective degree to zero. This process results in a configuration model network [135].

Following Ball and Neal [17], Lindquist et al. [118] developed an effective degree model categorising each node by its own disease state (S, I or R) along with the number of neighbours in each disease state resulting in the network being separated into classes representing the state of a node and its neighbours. For example, $S_{s,i}$ denotes the star motif corresponding to a susceptible node where s and i are the number of susceptible and infected neighbours of the central susceptible node.

Analogously to the Pairwise model (see section 3.4.2), ODEs describing the time evolution of network motifs may be written down; however for the Effective Degree model there are $M(M + 3)$ equations representing the time evolution of each of the possible star motifs (where M is the maximum degree) compared to the equations of the pairwise model (3.12-3.14) which represent the time evolution of all node and all edge combinations, of which there are 5 ($[S], [I], [SS], [SI], [II]$).

The susceptible node at the centre of a $S_{s,i}$ star experiences a force of infec-

tion, τ, from each of its i infected neighbours with infection resulting in the transition from an $S_{s,i}$ star to an $I_{s,i}$ star. Each of the i infectious neighbours recover at rate γ and thus transition from the $S_{s,i}$ class to the $S_{s,i-1}$ class at rate $\gamma i[S_{s,i}]$. By the same reasoning the rate at which transitions from $S_{s,i+1}$ to $S_{s,i}$ occur is $\gamma(i+1)[S_{s,i+1}]$. Infection of one the s susceptible neighbours results in a transition from $S_{s,i}$ to $S_{s-1,i+1}$ occurring at a rate of

$$\frac{\sum_{j,l}\tau jl[S_{j,l}]}{\sum_{j,l}j[S_{j,l}]}s[S_{s,i}]\ , \tag{3.16}$$

where above and henceforth notation of the form $\sum_{j,l}=\sum_{k=1}^{M}\ \sum_{j+l=k}$ is adopted to aid brevity. The rate derives from the fact that new infections are generated at rate $\sum_{j,l}\tau l[S_{j,l}]$ which in turn causes the effective degree of the susceptible neighbours of the now infected $[S_{j,l}]$ to change at rate $\sum_{j,l}\tau jl[S_{j,l}]$. Put in the notation of section 3.4.2 this can be expressed as the rate at which transitions $[ISS]\rightarrow[IIS]$ occur which is $\tau[ISS]$ where,

$$[ISS]=\sum_{j,l}[IS_{j,l}S]=\sum_{j,l}jl[S_{j,l}]\ . \tag{3.17}$$

As we wish to express the rate at which a susceptible neighbour of a $S_{s,i}$ star becomes infected, i.e. the rate of $[ISS_{s,i}]\rightarrow[IIS_{s-1,i+1}]$, we must account for the probability that the S–S link in the $[ISS]$ triple will connect to a $S_{s,i}$ star which is given by the following identity,

$$[ISA_{s,i}]=[ISA]\frac{s[A_{s,i}]}{\sum_{j,l}jA_{j,l}},\ A\in\{S,I\}. \tag{3.18}$$

Putting this all together then yields the rate given in (3.16). Mutatis mutandis the rate of transition from $S_{s+1,i-1}$ to $S_{s,i}$ is obtained.

Using the above rates the set of ODEs describing the time evolution of star motifs with a central susceptible are obtained as (3.19) below. Following similar reasoning allows (3.20) below to be derived with the extra term $\gamma[I_{s,i}]$ corresponding

to the recovery of the infectious individual at the centre of the star:

$$\frac{\mathrm{d}[S_{s,i}]}{\mathrm{d}t} = \gamma\Big((i+1)[S_{s,i+1}] - i[S_{s,i}]\Big) + \frac{\sum_{j,l}\tau jl[S_{j,l}]}{\sum_{j,l} j[S_{j,l}]}\Big((s+1)[S_{s+1,i-1}] - s[S_{s,i}]\Big) - \tau i[S_{s,i}] \tag{3.19}$$

$$\frac{\mathrm{d}[I_{s,i}]}{\mathrm{d}t} = \gamma\Big((i+1)[I_{s,i+1}] - i[I_{s,i}]\Big) + \frac{\sum_{j,l}\tau l^2[S_{j,l}]}{\sum_{j,l} j[I_{j,l}]}\Big((s+1)[I_{s+1,i-1}] - s[I_{s,i}]\Big) - \gamma[I_{s,i}] + \tau i[S_{s,i}] \tag{3.20}$$

The constraints upon s and i are given by $\{(s,i) : s \geq 0,\ i \geq 0,\ s+i \leq M\}$. Figure 3.2 summarises these two sets of equations governing the system in graphical form.

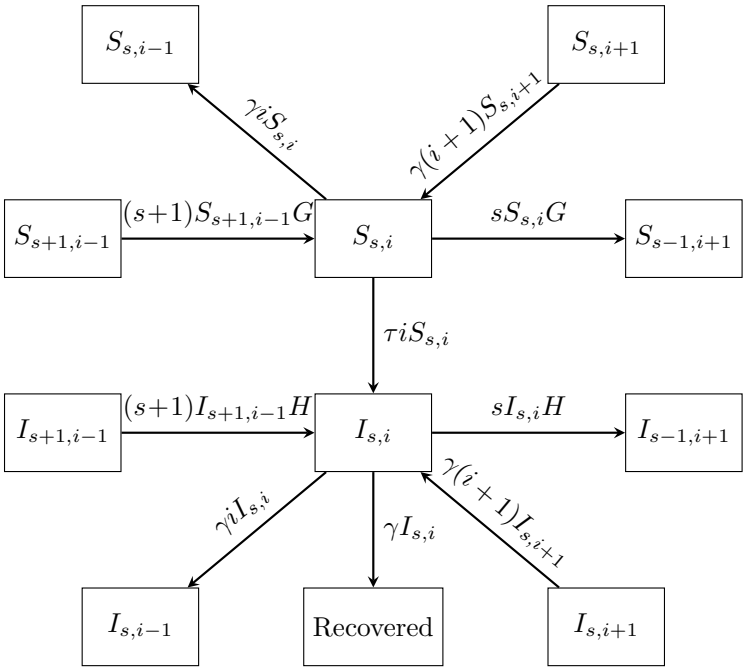

Figure 3.2: Flow diagram representing the state-space of the effective degree model of Lindquist et al. [118] with per-edge transmission rate τ and recovery rate γ. This visualises the flow rates into and out of motifs $S_{s,i}$ (a susceptible node with s susceptible and i infected neighbours) and $I_{s,i}$ (an infected node with s susceptible and i infected neighbours), where $G = \frac{\sum_{j,l}\tau jl[S_{j,l}]}{\sum_{j,l} j[S_{j,l}]}$, and $H = \frac{\sum_{j,l}\tau l^2[S_{j,l}]}{\sum_{j,l} j[I_{j,l}]}$. This figure has been reproduced from [118].

It is noteworthy to remark that the closure in this model is more implicit than that of the pairwise model, i.e. rather than approximating higher order states

with lower order states one makes the assumption that the infectious pressure on the susceptible neighbours of the central node is equal to the population average.

3.4.4 Probability Generating Function (PGF) Methods

Probability generating function (PGF) models [133, 199] provide low-dimensional representations of epidemic dynamics on configuration model networks and have been proved to be asymptotically exact [21, 33, 46, 94].

In the simplest form of PGF model, given in Miller [131], epidemic dynamics can be captured by one differential equation,

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\tau\theta + \gamma(1-\theta) + \tau\frac{\psi'(\theta)}{\psi'(1)} \ . \tag{3.21}$$

The Volz variable θ represents the probability that a 'test node' with one link remains susceptible, and its use is responsible for the massive simplicity of this model. It is expressed in terms of the per-link transmission rate (τ), recovery rate (γ), and PGF of the networks degree distribution ($\psi(\cdot)$). The derivation of this result is highly involved and would serve little use in the context of this section which aims to outline the type of techniques used to construct moment closures for epidemic on networks.

3.5 Dynamics on Clustered Networks

This section introduces extensions of all but the simplest model of the last section such that they better model dynamics on clustered networks. Sub-sections 3.5.1-3.5.2 introduce models from the literature that extend the pairwise and PGF approaches, whilst sub-section 3.5.3 introduces a novel model we developed that extends the effective degree model of Lindquist et al. [118] to the clustered regime.

The extension of the effective degree model addresses the modelling need to be capable of quickly simulating epidemic dynamics on realistic human contact networks - i.e. large, degree heterogeneous and highly clustered networks - more accurately. Fast and accurate simulation enables a rapid modelling response in the case of ongoing/emerging epidemics which can help co-ordinate an effective response.

3.5.1 Pairwise Model

The Kirkwood closure of section 3.4.2 can be extended to clustered networks [102] by introducing a correlation term, $\mathcal{C}_{CA} = \frac{N}{\langle k\rangle}\frac{[CA]}{[A][C]}$, that accounts for the number of transitive links between A and C by measuring the observed $[CA]$ compared to

the number of A–C pairs one would expect given independence between C and A, $\frac{\langle k\rangle[A][C]}{N}$. If $\mathcal{C}_{CA} = 1$ then nodes of type C and nodes of type A are connected at random. Given a clustering coefficient, ϕ, this correlation term is applied to a proportion ϕ of the original Kirkwood closure, giving the clustered Kirkwood closure as,

$$[ABC] \approx \frac{\langle k\rangle - 1}{\langle k\rangle}\frac{[AB][BC]}{[B]}\left((1-\phi) + \phi\,\mathcal{C}_{AC}\right) \ . \tag{3.22}$$

Application of this modified closure to (3.12)-(3.14) yields the clustered pairwise model.

3.5.2 Clustered PGF

House and Keeling [90] introduced the clustered PGF model, which combined the PGF (section 3.4.4) and pairwise (section 3.4.2) models to yield a system of six ODEs which provide a good approximation of epidemic dynamics on degree heterogeneous clustered networks. The first ODE describes the time evolution of the number of effective nodes, $[I]$, and is given by (3.10). The second and third ODEs derive from the PGF approach,

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\tau\frac{[SI]}{N\psi'(\theta)} \tag{3.23}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \tau\left(\frac{\theta\psi''(\theta)}{\psi'(\theta)} + 1\right) - \gamma Y \tag{3.24}$$

The Volz variable θ represents the probability that a 'test node' with one link remains susceptible, and its use is responsible for the massive simplicity of this model. It is expressed in terms of the per-link transmission rate (τ), recovery rate (γ), and PGF of the networks degree distribution ($\psi(\cdot)$). The number of susceptible nodes can be expressed in terms of the PGF, $[S] = N\psi(\theta)$. Y denotes the number of infectious links in the system, $Y = \sum_k k[I_k]$. The three remaining ODEs are gained by combining the pair-level equations of the pairwise model (3.12-3.14) with the following closures the authors derive,

$$[ISS] \approx \frac{[SI][SS]\psi''(\theta)}{N\psi'(\theta)^2}\left((1-\phi) + \phi\langle k\rangle\frac{[SI]}{N^2\theta\psi'(\theta)^2}\right) \ . \tag{3.25}$$

$$[ISI] \approx \frac{[SI]^2\psi''(\theta)}{N\psi'(\theta)^2}\left((1-\phi) + \phi\langle k\rangle N\frac{[II]}{Y^2}\right) \ . \tag{3.26}$$

As with the PGF model of section 3.4.4, derivation is rather involved and not the purpose of this work. All we need are equations above in order to solve the equations for a network of given degree distribution and clustering coefficient. The clustered PGF is the main model that attempts to deal with epidemics on clustered heterogeneous networks and as such serves as a direct comparator to assess the performance of our new approach.

3.5.3 A New Model

The effective degree (ED) degree model (section 3.4.3) approximates the epidemic dynamics on a network of arbitrary degree distribution extremely well; however, in the large network limit the clustering coefficient of the configuration model network it explains tends to zero, which does not realistically reflect real world contact networks. A new set of equations extending the effective degree model to clustered networks is now presented.

This model was first formulated by the author of this thesis [26] as part of a mini-project directly preceding this PhD; however the initial investigation of this model presented was both incorrect and incomplete. Firstly, the equations governing the rate parameters gave an incorrect correlation parameter. Furthermore, the implementation of rate parameters for the model introduced erroneous factors of $\frac{1}{2}$. The analysis was incomplete as model performance during the exponential growth phase was not observed and the error properties were not investigated.

An exact version of the ED model (without closures or mean-field assumptions) may be written as follows,

$$\begin{aligned}\frac{\mathrm{d}[S_{s,i}]}{\mathrm{d}t} =& -\tau i[S_{s,i}] + \gamma\Big((i+1)[S_{s,i+1}] - i[S_{s,i}]\Big) - \tau[ISS_{s,i}]_{\triangle} - \\ & \tau[ISS_{s,i}]_{\wedge} + \tau[ISS_{s+1,i-1}]_{\triangle} + \tau[ISS_{s+1,i-1}]_{\wedge}\ , \end{aligned} \tag{3.27}$$

$$\begin{aligned}\frac{\mathrm{d}[I_{s,i}]}{\mathrm{d}t} =& \tau i[S_{s,i}] + \gamma\Big((i+1)[I_{s,i+1}] - i[I_{s,i}]\Big) - \tau[ISI_{s,i}]_{\triangle} - \tau[ISI_{s,i}]_{\wedge} + \\ & \tau[ISI_{s+1,i-1}]_{\triangle} + \tau[ISI_{s+1,i-1}]_{\wedge} - \gamma[I_{s,i}] + \end{aligned} \tag{3.28}$$

$$\tau\Big((s+1)[I_{s+1,i-1}] - s[I_{s,i}]\Big)\ , \tag{3.29}$$

where we separate out transmission events happening within and outside of the neighbourhood of the star: $[ISS_{s,i}]_{\triangle}$ denotes a closed triple forming a triangle (e.g. edge ① figure 3.3) and $[ISS_{s,i}]_{\wedge}$ denotes an unclosed triple (e.g. edge ② figure 3.3).

In a network with a clustering coefficient of zero then $[ABC]_{\triangle} = 0$ and a

Triangle state	$[SSA_{s,i}]_\triangle$	$[ISA_{s,i}]_\triangle$	$[IIA_{s,i}]_\triangle$
Combinations	$\frac{1}{2}s(s-1)$	is	$\frac{1}{2}i(i-1)$

Table 3.1: The number of possible triangles around a node in state $A_{s,i}$, where $A \in \{S, I\}$, varies depending on the states involved. 'Combinations' gives the number of possible triangles of type 'Triangle state'.

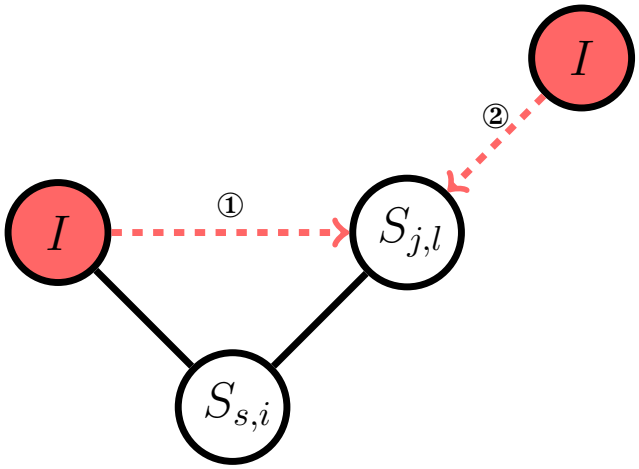

Figure 3.3: Illustration of transmission events occurring within the neighbourhood of a central node ($S_{s,i}$ - with s susceptible and i infected neighbours) and outside of it for the clustered effective degree model. Edge ① denotes a transmission event within the neighbourhood and edge ② denotes a transmission event outside of the neighbourhood of the central $S_{s,i}$. The clustered effective degree model models these two types of transmission event separately unlike the original effective degree model which assumes that the neighbourhood of the central $S_{s,i}$ is locally tree-like.

closure on the $[ABC]_\wedge$ terms will yield the ED model [118]; however an effective closure for clustered networks must be made on both $\triangle$ and $\wedge$ terms.

Given a clustering coefficient, ϕ, there are an expected $\phi k(k-1)/2$ triangles around a given degree-k node. If we decompose the effective degree (c.f. section 3.4.3) of a node, k, into its susceptible and infected neighbours such that $s + i = k$ then Table 3.1 enumerates the expected number of triangles involving different states, given no correlations between states.

$[ISS_{s,i}]_\triangle = \phi is[S_{s,i}]$ as the clustering coefficient, ϕ, gives the expected ratio of edges connecting a node's neighbours together to the maximum possible number of such edges (Table 3.1). Correlations between the states cannot be ignored, therefore

the correlation between nodes of state A and B, $\mathcal{C}_{AB} = \frac{N}{\langle k \rangle} \frac{[AB]}{[A][B]}$, is introduced which is used to account for how many $[AB]$ pairs there are compared to how many one would expect from random mixing, $\langle k \rangle [A] \frac{[B]}{N}$. This yields the closure equation (3.33), and mutatis mutandis (3.35) is obtained.

The original ED closure for the $\wedge$ terms must now be modified to account for the infection events that happen within the $\triangle$ closure. The new closure is achieved by taking the original expressions and, for $A \in \{S, I\}$, using the identity $[ISA]_{\triangle} + [ISA]_{\wedge} = [ISA]$ along with the original and the $\triangle$ closure. For example,

$$[ISS] = \sum_{j,l} [IS_{j,l}S] = \sum_{j,l} jl[S_{j,l}] \,, \tag{3.30}$$

$$[ISS]_{\triangle} = \sum_{j,l} [IS_{j,l}S]_{\triangle} = \sum_{j,l} \phi \, \mathcal{C}_{SI} jl[S_{j,l}] \,, \tag{3.31}$$

$$\implies [ISS]_{\wedge} = \sum_{j,l} [IS_{j,l}S]_{\wedge} = \sum_{j,l} jl(1 - \phi \, \mathcal{C}_{SI})[S_{j,l}] \,. \tag{3.32}$$

Using identity (3.18), one gains the final clustered effective degree closures (3.33-3.36) which when substituted into (3.27-3.28) yield the clustered ED model.

$$[ISS_{s,i}]_{\triangle} = \phi \mathcal{C}_{\mathcal{SI}} \, si[S_{s,i}] \,, \tag{3.33}$$

$$[ISS_{s,i}]_{\wedge} = \frac{\sum_{j,l} jl(1 - \phi \mathcal{C}_{\mathcal{SI}})[S_{j,l}]}{\sum_{j,l} j[S_{j,l}]} s[S_{s,i}] \,, \tag{3.34}$$

$$[ISI_{s,i}]_{\triangle} = \phi \mathcal{C}_{\mathcal{II}} \, si[I_{s,i}] \,, \tag{3.35}$$

$$[ISI_{s,i}]_{\wedge} = \frac{\sum_{j,l} l(l-1)\big(1 - \phi \mathcal{C}_{\mathcal{II}}\big)[S_{j,l}]}{\sum_{j,l} j[I_{j,l}]} s[I_{s,i}] \,. \tag{3.36}$$

3.6 Simulation Study

To test the accuracy of the clustered ED model we perform a simulation study comparing the clustered ED model to the clustered PGF model, and the 'exact' dynamics.

3.6.1 Methodology

First, unclustered networks are generated for three different degree distributions according to the methods presented in section 3.3. The big-V algorithm (figure 3.1 ⑦) was then applied to each unclustered network to generate a series of networks with approximate clustering coefficients $\phi \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$.

Numerical simulation was performed using an individual-based analogue of Gillespie's algorithm (section 2.2.4) [69], which generates a statistically correct trajectory with regard to the master equation of the underlying stochastic process. Simulations were performed for 20 uniquely generated networks of 10^5 nodes with the dynamics being simulated on each network 5 times, for a total of 100 epidemics per network type. To account for the large stochastic variability at the beginning of an epidemic, we shifted the time-origins of each of the 100 epidemics to coincide at the point where 500 individuals are infected before averaging the dynamics, which we then compare to each differential equation model.

Larger network sizes reduce stochastic fluctuations in the simulations. For comparison of the accuracy of two ODE models to the expected prevalence of the 'exact' simulated dynamics, networks of size $N = 10^4$ are sufficient [90]; however we have found that larger networks are required to begin to make an assessment of how the error of our model scales with the clustering coefficient. Therefore, the network size was chosen to be as large as possible whilst still being able to produce sufficiently clustered networks—the big-V rewiring algorithm which we use to generate clustered networks exhibits a critical slowing down of its ability to generate clustering coefficients approaching $\phi = 0.3$ for networks larger than 10^5 nodes. This slowing down arises because the algorithm cannot find 'V' configurations in the network—rewiring one 'V' configuration to create an edge and a triangle results in the removal of another triangle. Generating arbitrarily clustered networks of a given degree distribution remains an open problem within complex networks.

We choose to generate and simulate networks of mean degree four as a rigorous model comparator—closures perform better as $\langle k \rangle$ increases because the behaviour tends towards the simple SIR model which is conceptually an epidemic on a complete graph.

3.6.2 Results

The results in figures 3.4–3.6 plot the epidemic dynamics for moment closure models compared to the 'exact' dynamics along with figure 3.7 which explicitly plots the errors. It is clear that the clustered ED model generally outperforms the clustered PGF approach in terms of errors in expected prevalence, with exceptions to this only occurring after the epidemic peak. For example, the bottom left subplot of figure 3.7 is the only occasion that the cumulative error of the clustered effective degree approach (red line) significantly or consistently exceeds that of the clustered PGF approach (green line); however this only occurs after the epidemic peak.

For 4-regular networks the dynamics of the models (figure 3.4) display sim-

ilar early exponential growth rates (in the figure insets). All models over-estimate prevalence at the epidemic peak but the clustered effective degree model captures the timing of the epidemic peak much better. Note that for regular networks the clustered pairwise and clustered PGF models are identical.

The clustered PGF model over-estimates growth in the exponential phase for highly clustered Poisson degree distributed (Erdős-Rényi) networks (figure 3.5). The clustered effective degree model underestimates but to a lesser extent. Both over-estimate peak prevalence as they did for regular networks - the clustered effective degree model less so - with the clustered effective degree model capturing the timing of the epidemic peak. The exponential growth rate of the clustered pairwise model is completely wrong for all levels of clustering, and due to its unsuitability for degree heterogeneous networks we shall not make further remarks on its performance.

Figure 3.6 shows the results for negative binomial networks - the most heterogeneous of the networks we simulate, and thus the biggest challenge for moment closure models. For small ϕ negative-binomial networks, both the clustered effective degree and clustered PGF models exhibit a similar exponential growth rate; however the clustered effective degree model adapts this rate much more accurately than the clustered PGF model when ϕ is increased. Furthermore, the clustered effective degree model continues to capture the timing and magnitude of the epidemic peak more accurately than the clustered PGF approach, although for $\phi \geq 0.3$ the performance begins to deteriorate.

Work by Pellis et al. [157] argued (based on a rigorous analysis of finite-size networks) that we should expect moment closure to be exact when (i) triangles are not overlapping and (ii) recovery happens after a constant time (or not at all as when $\gamma = 0$). When the clustering coefficient is small, the proportion of overlapping triangles will be $O(\phi^2)$, and so for small values of γ we expect errors in the prediction of prevalence of infection to be $O(\phi^2)$, while for larger values we expect them to be $O(\phi)$. We found that in general, obtaining accurate assessments of absolute error in predictions of prevalence was numerically challenging, and it is likely that deeper theoretical understanding of the source of errors would be required to perform a definitive computational analysis of this question. Nevertheless, we were able to obtain the results shown in figure 3.8, which demonstrate that as expected errors lie between the $O(\phi)$ and $O(\phi^2)$ lines. $\tau = 6$ was chosen, compared to $\tau = 2$ for the rest of the simulation study, to reduce errors arising from the finite length of infectious

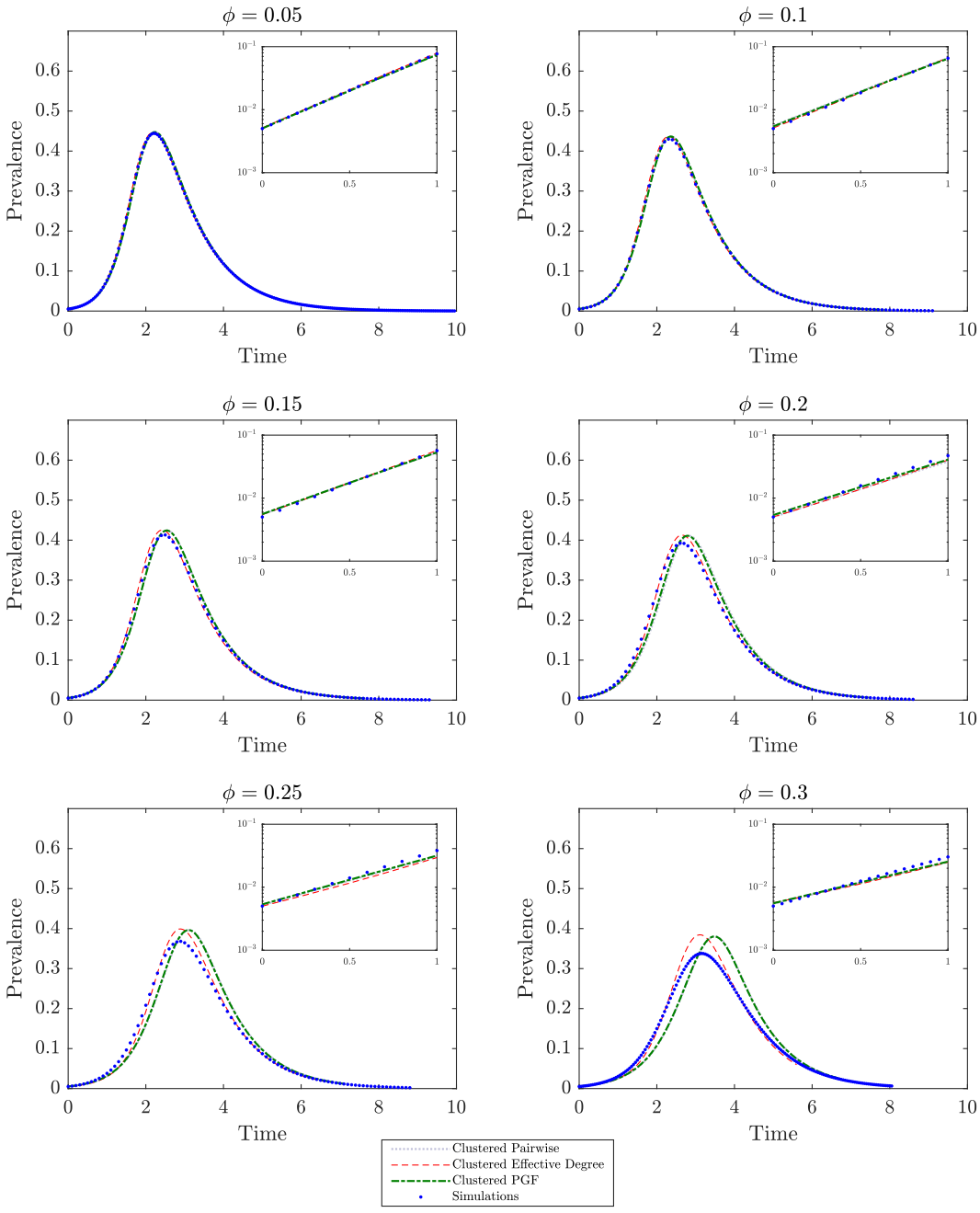

Figure 3.4: Epidemic dynamics on 4-regular networks of $N = 10^5$ nodes, with recovery rate $\gamma = 1$ and per-edge transmission rate $\tau = 2$. Each figure facet represents varying clustering coefficient, ϕ. Prevalence versus time are plotted for the clustered pairwise model (dotted grey), clustered effective degree model (dashed red), clustered PGF model (dot-dashed green), and the mean prevalence of an ensemble of 100 stochastic simulations representing the 'exact' dynamics (blue dots). The inset figures plot log-prevalence against time during the exponential growth phase of the epidemic. On a regular network, the clustered Pairwise and clustered PGF models are identical.

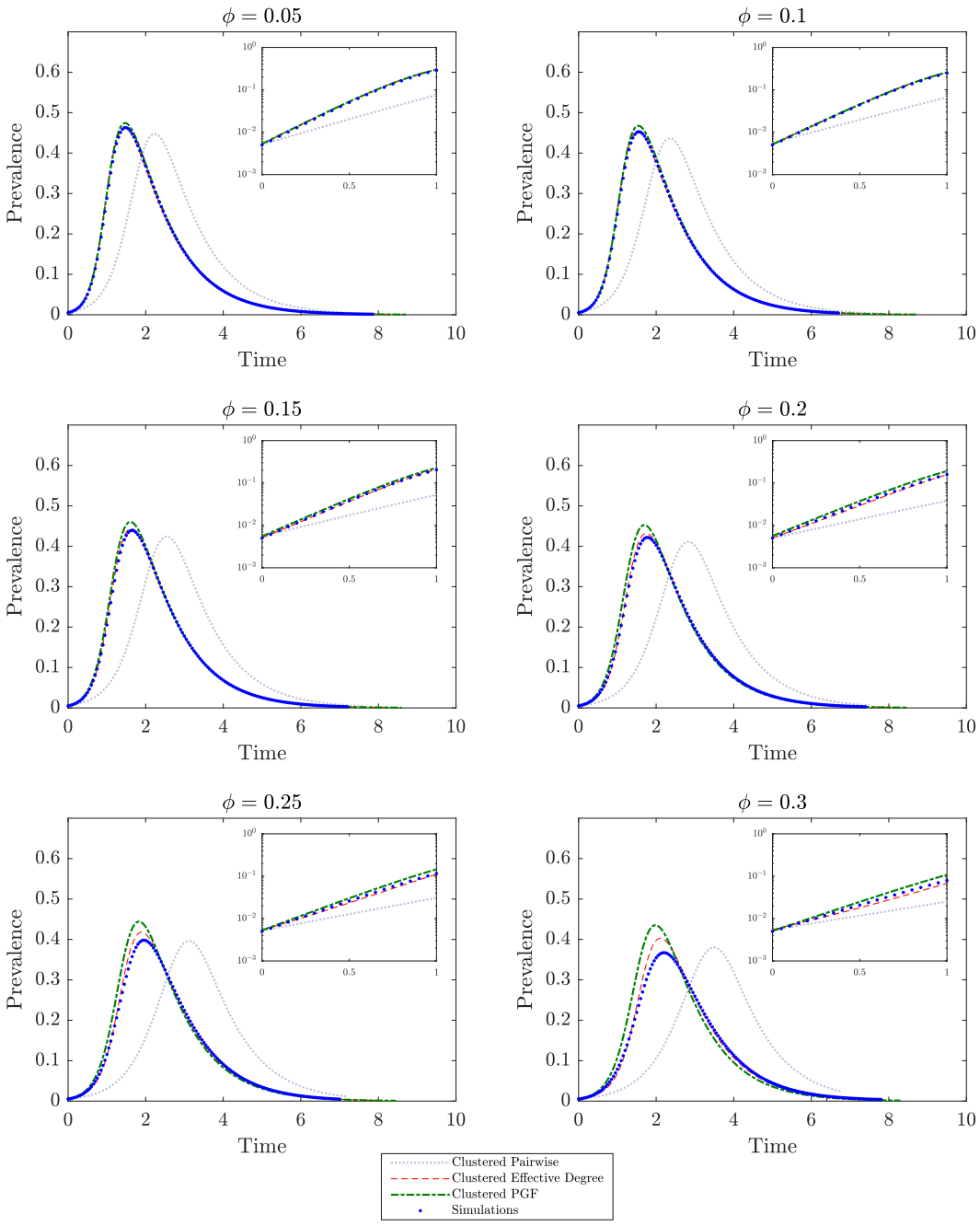

Figure 3.5: Epidemic dynamics on Erdős-Rényi networks of $N = 10^5$ nodes and mean-degree $\langle K \rangle = 4$, with recovery rate $\gamma = 1$ and per-edge transmission rate $\tau = 2$. Each figure facet represents varying clustering coefficient, ϕ. Prevalence versus time are plotted for the clustered pairwise model (dotted grey), clustered effective degree model (dashed red), clustered PGF model (dot-dashed green), and the mean prevalence of an ensemble of 100 stochastic simulations representing the 'exact' dynamics (blue dots). The inset figures plot log-prevalence against time during the exponential growth phase of the epidemic.

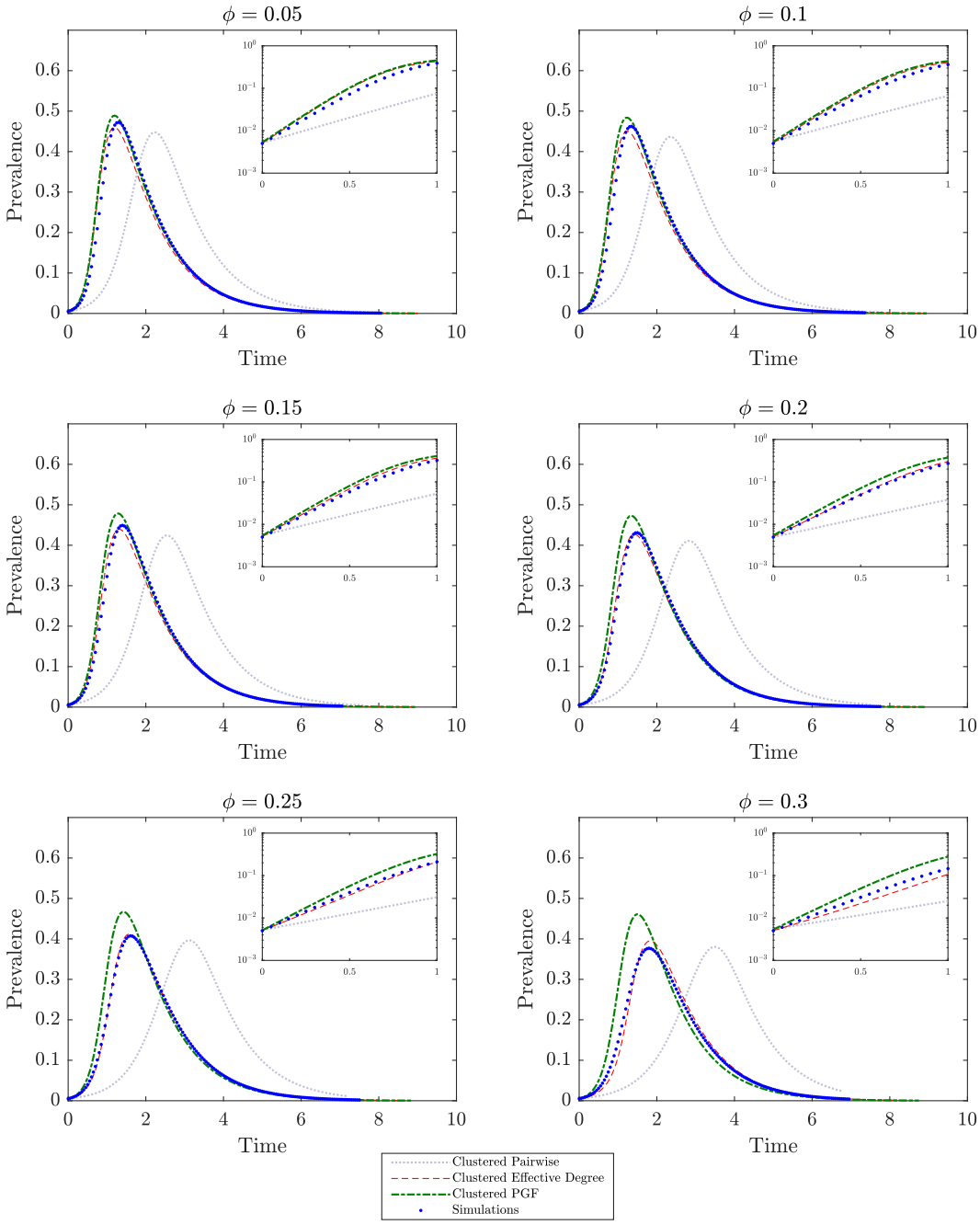

Figure 3.6: Epidemic dynamics on negative binomial networks of $N = 10^5$ nodes with mean-degree $\langle K \rangle = 4$, shape parameter $r = 5$, recovery rate $\gamma = 1$ and per-edge transmission rate $\tau = 2$. Each figure facet represents varying clustering coefficient, ϕ. Prevalence versus time is plotted for the clustered pairwise model (dotted grey), clustered effective degree model (dashed red), clustered PGF model (dot-dashed green), and the mean prevalence of an ensemble of 100 stochastic simulations representing the 'exact' dynamics (blue dots). Inset figures plot log-prevalence against time during the exponential growth phase of the epidemic. We use the negative binomial formulation counting p successes given r failures, such that $\langle K \rangle = \frac{pr}{1-p}$.

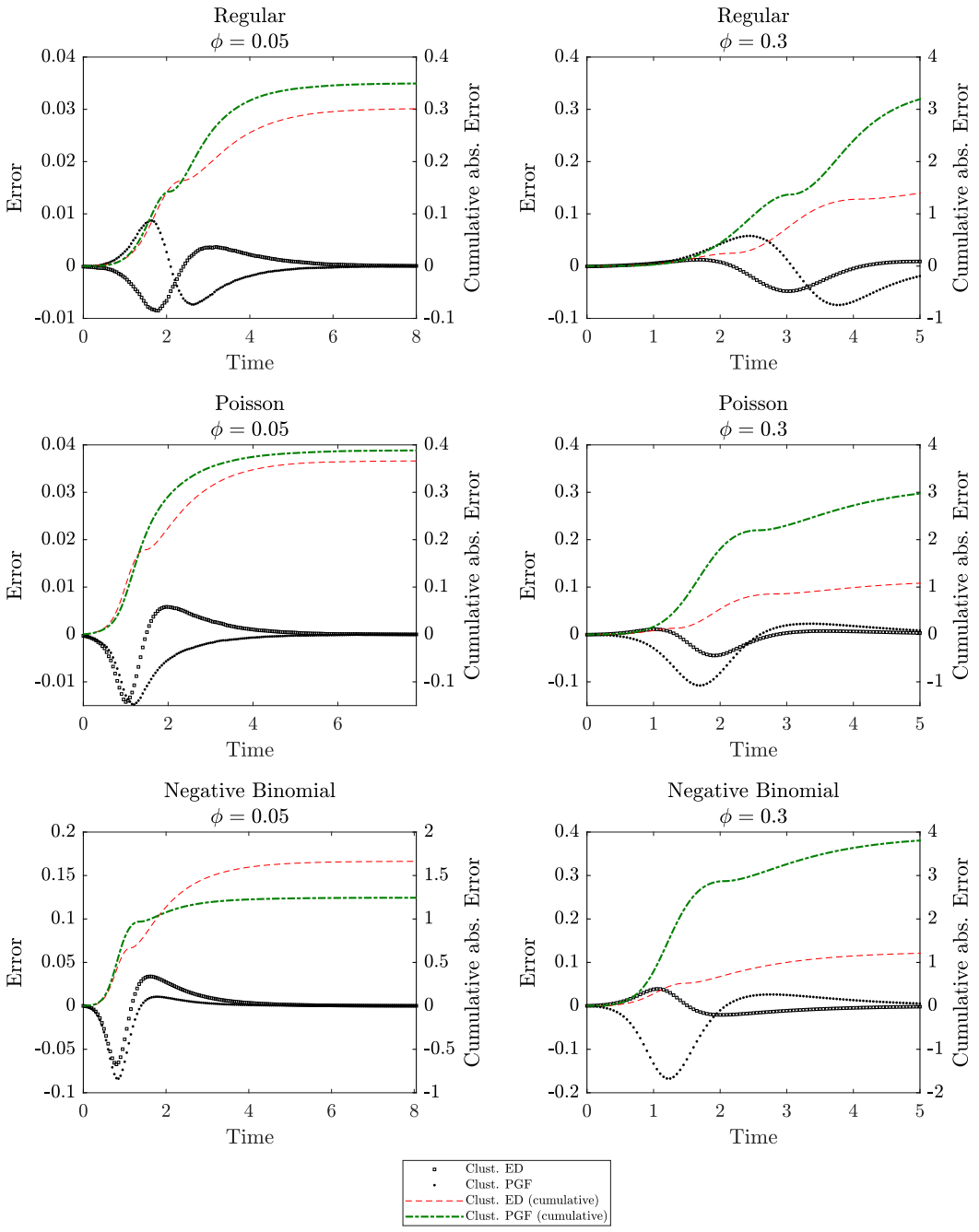

Figure 3.7: Absolute error (left axis) and cumulative absolute error (right axis) over time of the clustered effective degree (ED) and clustered PGF models as compared to the mean prevalence of an ensemble of 100 stochastic simulations representing the 'exact dynamics'. The left (right) column gives the errors for $\phi = 0.05$ ($\phi = 0.3$), with each subsequent row corresponding to Regular, Poisson, and Negative Binomial degree distributed networks. The networks and epidemic parameters, are those of figures 3.4-3.6 – $N = 10^5$, $\langle K \rangle = 4$, $r = 5$, $\tau = 2$, $\gamma = 1$.

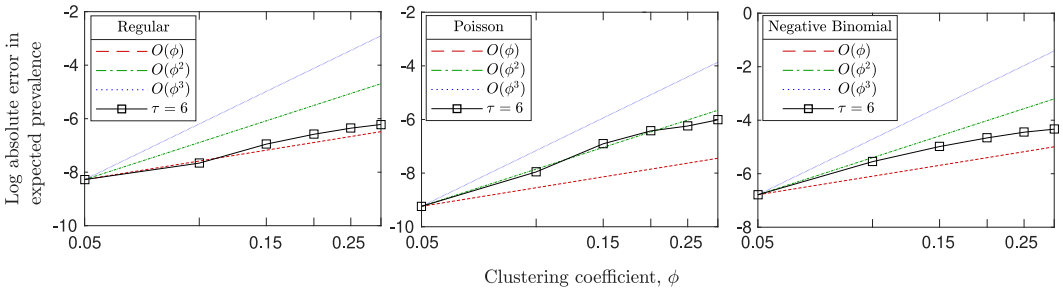

Figure 3.8: Log absolute error between the clustered effective degree model and expected prevalence of 100 stochastic simulations (detailed in section 3.6.1) plotted against errors of order ϕ, ϕ^2, and ϕ^3 plotted in red, green, and blue respectively. Networks were of size $N = 10^5$ with epidemic parameters $\tau = 6, \gamma = 1$.

period which Pellis et al. [157] conjecture are $O(\gamma/\tau)$.

3.7 Discussion

In this chapter, we have introduced the theory behind simulating epidemics on networks using both a stochastic simulation approach and a moment closure ODE approach. We then demonstrated how to combine effective degree approaches to modelling epidemics on heterogeneous networks with moment closures for clustered networks, yielding a new model. Extensive numerical simulations were performed to compare our model and the best model from the literature, the clustered PGF model, showing conclusively that our clustered effective degree model performs best, particularly for highly clustered degree heterogeneous networks. This improved performance does come at increased computational cost - the number of ODEs in the clustered effective degree model scales with the maximum degree of the degree distribution, M, whereas the clustered PGF model has a fixed number of equations. For networks with large maximum degree, M, this means that the effective degree model takes much longer to simulate than the PGF approach (though still several orders of magnitude less than stochastic simulation). Furthermore, for very large M there is the possibility that small numerical errors may affect performance due to numerical errors when solving such a large number of equations.

Numerical analysis suggests that the errors introduced in this approach are better than $O(\phi)$ but worse than $O(\phi^2)$, meaning that potential for improvement remains. In particular, the closure due to Kirkwood and Boggs [108] could be replaced by the maximum entropy method [171] or the improved closure of House and Keeling [88] which preserves the pair-level relationship $\sum_B [AB] = \langle K \rangle [A]$ that the

introduction of the correlation term $\mathcal{C}_{\mathcal{AC}}$ in the clustered Kirkwood closure does not conserve.

In the large ϕ limit ($\phi \approx 1$) we expect graphs to be dominated by cliques [47], whose differential equation limit is better described by equations such as those written in Ball [16]; however this level of clustering appears in a population of households and is not present at the societal scale for which we have developed this models.

One of the strengths of effective degree models are their combined flexibility and accuracy. Given this, other meritorious avenues for future work would be to incorporate waning immunity into the model, and investigate the performance of the clustered effective degree model for these dynamics. In principle, it is possible that degree assortativity could be incorporated into an effective degree model; however it is not clear how to extend the network generation methodology (section 3.3) to incorporate clustered and degree assortative networks with a given degree distribution.

CHAPTER 4

Household driven transmission of soil-transmitted helminths

4.1 Introduction

Soil-transmitted helminthiasis is a type of helminth infection caused by soil-transmitted helminths (STH) - parasitic nematode worms - with human infection occurring through environmental contact with eggs/larvae. More than a billion people, primarily in tropical and subtropical regions, are infected with at least one species of STH with the primary three being roundworms (Ascaris lumbricoides), whipworms (Trichuris trichiura), and hookworms (Necator americanus & Ancylostoma duodenale) [147]. This thesis concerns itself with Ascaris; however the methods developed may also be used not only to model the other types of STH but other more general diseases. Ascaris is the most prevalent of the STHs with some estimating over a billion people infected with Ascaris [44]. Humans are the only major host for Ascaris with the worms residing in the human gastrointestinal system for up to several years. Reproduction occurs in the intestines between adult worms with the eggs entering the external environment along with faeces. Eggs develop in the external environment until they are ingested with Larvae penetrating the intestinal mucosa and migrating via the bloodstream through the lungs, trachea, and pharynx until they are swallowed and reach the adult reproductive stage about 9-11 weeks after ingestion [25]. This life-cycle is summarised in figure 4.1. Ascaris typically receives the most attention out of the three primary species of STH because worm burden - the number of worms within a host - is highest in children aged 5-15 years, and although symptomless in $\sim 85\%$ cases chronic infections result in malnutrition, growth stunting, intellectual retardation, and cognitive deficits.

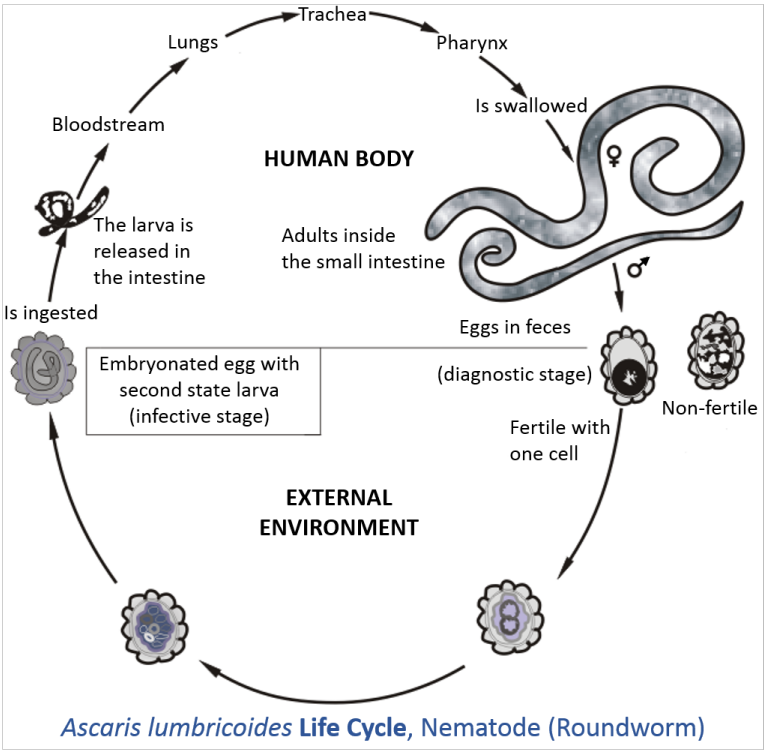

Figure 4.1: Life-cycle of Ascaris. Reproduction occurs in the intestines between adult worms with the eggs entering the external environment along with faeces. Eggs develop in the external environment until they are ingested after which the Larvae penetrate the intestinal mucosa and migrate via the bloodstream through the lungs, trachea, and pharynx until they are swallowed. The adult reproductive stage is reached about 9-11 weeks after ingestion [25]. Image credit: https://bit.ly/2Zr6CdW (Creative Commons 2.0)

Ascaris is treatable with several anthelmintic drugs, such as Mebendazole and Albendazole, which remove worms from the body and causes the death of adult worms over the course of several days [137]; however control is incredibly hard due to the environmental reservoir of infection as each female worm produces thousands ($\sim 200,000$) of eggs per day which can survive for many years in the soil [50]. Treatment of Ascaris with these anthelmintic drugs is known as PCT (Preventative Chemo-Therapy), and is given at frequent (e.g. yearly) intervals to as large a proportion of the eligible population (pregnant women and infants under 1 are not eligible) as possible in what is known as MDA (Mass Drug Administration). An illustrative example of the principles behind MDA is given in figure 4.2.

Intensity of infection is the main epidemiological index used to describe infection because morbidity and transmission rate are directly related to the number

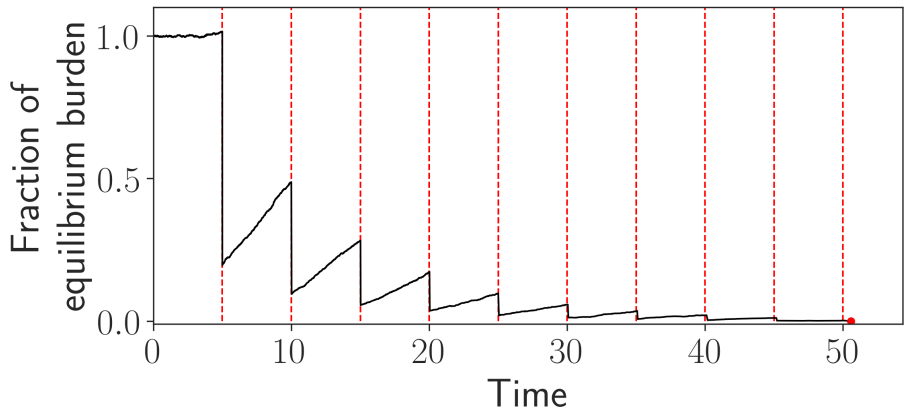

Figure 4.2: Illustrative dynamics of the administration of MDA in a population. Anthelmintics are given to some proportion of the population at fixed intervals (dashed red lines), reducing the number of worms in the system. The number of worms in the system begins to rise towards the equilibrium until another treatment is applied. This continues until either: (i) the "break-point" is reached where transmission cannot be sustained (Ascaris reproduce sexually) and the disease is eliminated (pictured in this scenario); (ii) control is applied indefinitely as the break-point cannot be reached; (iii) efforts to eliminate are ceased and the system returns to equilibrium. This figure was produced for illustrative purposes with a simple simulation, written in Python.

of worms within the host. Intensity is typically measured by the number of eggs per gram of faeces (EPG), historically using the Kato-Katz diagnostic technique [176] - a method of preparing a faecal smear prior to microscope examination and egg counting. Worm burden is not as commonly used as EPG because it is incredibly time-consuming, requiring giving the subject anthelmintics and waiting for worms to be expelled over the course of several days. Whilst EPG is a good intensity measure for the transmission rate - it directly measures the transmissible medium of Ascariasis - there are multiple shortcomings with it as a measure of intensity and the diagnostic techniques used to estimate it. Firstly, only female worms produce eggs meaning a host infected with only male worms (not improbable for a host with only one or two worms) will never have their infection detected by EPG diagnostic techniques. Consequently, even though EPG arguably better captures transmission-rate than worm-burden, in areas with very low prevalence or intensity there may be a large discrepancy between prevalence as measured by worm burden and prevalence as measured by EPG which will impact on the transmission dynamics close to elimination. Two large-scale studies [81, 86] of the relationship between worm burden

and EPG in very-high prevalence areas (both $> 90\%$) found that 6.3% and 6.4% of infections were missed by the EPG microscopy diagnostics used. Furthermore, setting aside the inability of EPG to directly measure male worm burden, EPG remains an imperfect proxy of the true worm burden. The concentration of eggs can vary within a faecal sample and over-time therefore EPG is a noisy measure of intensity of infection at the individual level [80]. The mean concentration of eggs over multiple samples is considered to be representative of worm burden leading WHO to define a light infection with Ascaris as $< 5,000$EPG and a heavy infection as $> 50,000$EPG [136]. Analysis by Hall and Holland [79] found considerable heterogeneity between countries' relationships between EPG and worm burden, likely due to differences in female fecundity and the quality of the diagnostic technique, concluding that EPG was only a reliable measure of the intensity of infection centred around a single site.

Mathematical modelling of the spread of Ascaris has found that the treatment of children alone is insufficient to control Ascaris [6, 189, 190]. Another factor that increases the difficulty of control is that Ascaris is often observed to be unevenly distributed amongst a population with a heavy tail meaning that a large proportion of worms reside in a small proportion of the population which could easily be missed by control measures. Some of the variation [200] can be explained by 'pre-disposition' or the likelihood of hosts with a high macro-parasitic load to be quickly reinfected following treatment. On the other hand, it is also observed that there are 'wormy' households [40, 43, 61, 200]. The effectiveness of targeted control methods will vary depending on the process which generates these heterogeneous worm distributions. Furthermore, age is also a factor with loads appearing to decline post-childhood due to an unknown combination of developed immunity and behavioural change [50, 78].

This leads us to two key research questions which we wish to explore in order to improve the ability to design efficient control strategies. Firstly, is it children or adults who are the major drivers of transmission? Furthermore, what are the roles of households in generating the aggregation of worms that we see?

4.2 Data

The data we will use to answer our previously stated research questions are from a study performed in south west Nigeria by Asaolu et al. [11]. The authors compared 'mass', 'targeted' and 'selective' PCT (Preventative Chemo-Therapy) with Levamisole in three communities in Oyo State, Nigeria. In addition, a fourth community was measured but no control was performed until after the study. Each village lies within 10.5-27km of the Obafemi Awolowo University campus (figure 4.3)

at Ile-Ife, and were selective based on the basis of their similar socio-economic profiles and co-operation. Village residents were subsistence and cocoa farmers, their children attended a primary school located in their own village. Dwellings in the study villages were built of mud-wall and roofed with corrugated iron sheets. Due to a lack of infrastructure, refuse and human faeces were dumped in the bush behind each house. Ponds and streams located near the village served as a water-source.

Figure 4.3: Map of modern day Nigeria with two circles, centred at Obafemi Awolowo University campus, between which the four villages in the study lie. The circles are of radius 10.5km and 27km. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

'Selective' control targets the most heavily infected 20%; 'Targeted' control targets children aged 2-15 years; and 'Mass' control targets everyone except infants <1 year of age and pregnant women. The Villages were isolated from one another with no evidence of previous STH control initiatives, and Levamisole was given at 3-monthly intervals for 1 year. Whilst multiple measurements were taken we only have access to the pre-treatment measurement stage which measures the age, sex, household membership, and EPG (eggs per gram of faeces) count for Ascaris, Trichuris, and Hookworm. EPG measurements were performed by mixing each stool thoroughly and fixing a portion in 10% aqueous formaldehyde before examining using a modified

Kato-Katz procedure [146]. The pre-treatment faecal samples were taken during January and February 1989.

First, we begin by briefly exploring the dataset to try to extract modelling insights. Figure 4.4 plots the proportion of a given household that are infected with Ascaris. We observe a 'U'-shape in the prevalence curve between 0.5 and 1.0, and moderate heterogeneity between villages. Figure 4.5 shows the age distribution of the communities binned at intervals of one and five years. A larger proportion of the population are younger than 20 compared to a more developed Country. In addition, we observe a censoring effect for adult ages where multiples of 5 and 10 are much more likely, probably as participants only give a rough estimate of their age.

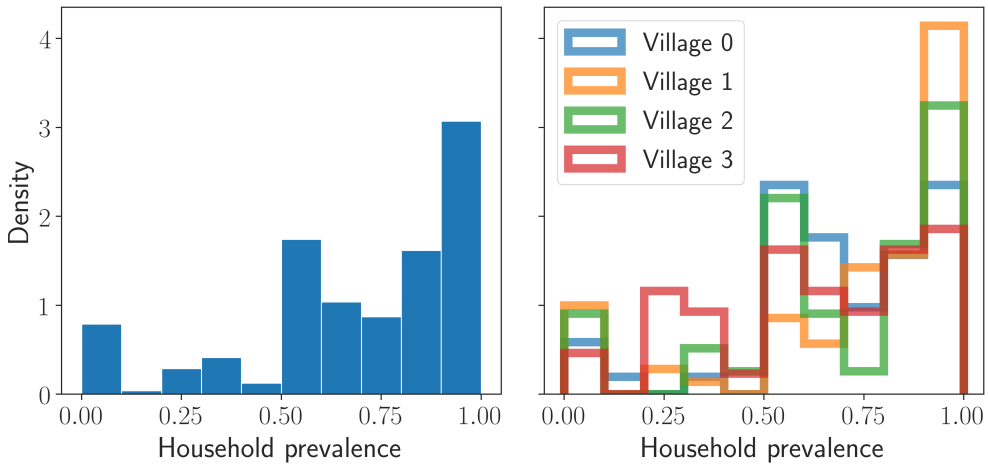

Figure 4.4: Proportion of a given household infected for all villages (left), and stratified by village (right). A small number of households have zero or very low prevalence, but most have a prevalence of at least 0.5. The prevalence distribution above 0.5 exhibits a 'U'-shaped curve.

Another departure from developed countries are the distribution of household sizes (figure 4.6) which are incredibly large. If indeed the spread of Ascaris is facilitated by the closeness of a household unit then large households could spread infection rapidly depending on whether the epidemic dynamics are frequency or density dependent.

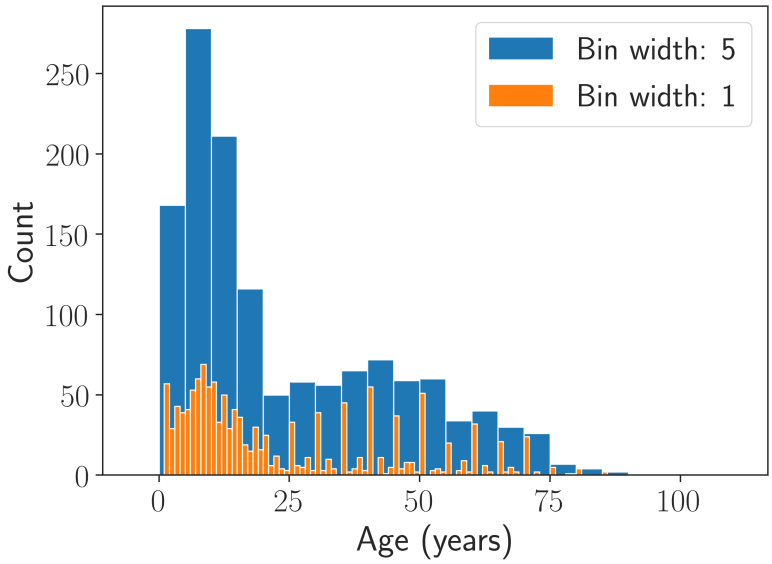

Figure 4.5: Age distribution of study participants. The blue bars are binned at intervals of 5 years and show the general patterns of age distribution. The orange bars are binned at yearly intervals and reveal a censoring effect for adult ages where multiples of 5 are much more likely.

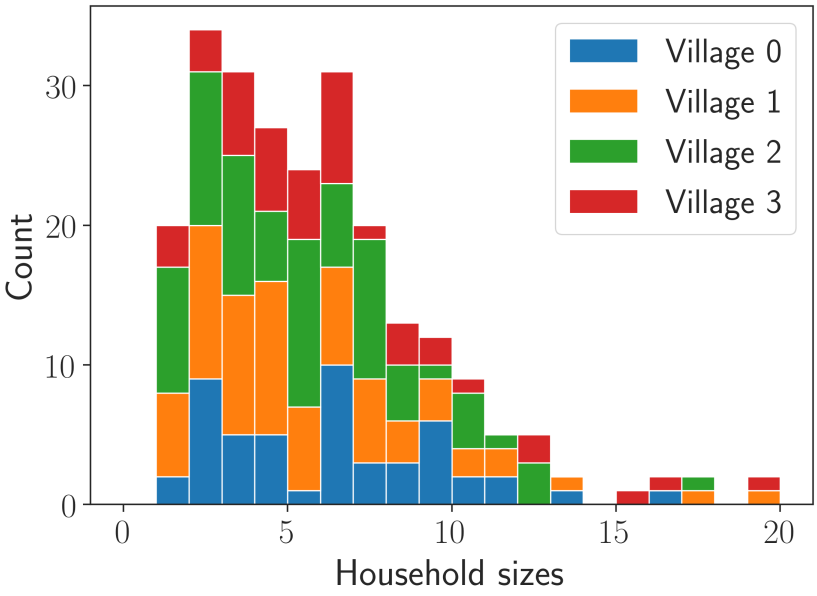

Figure 4.6: Distribution of household sizes split by village.

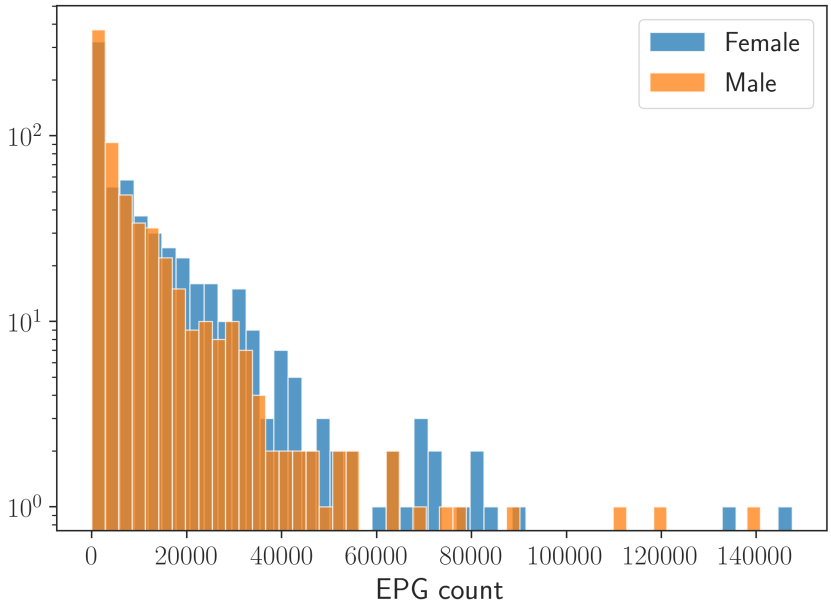

Figure 4.7: Ascaris eggs per gram of faeces (EPG) split by gender. The two distributions are broadly similar, with the distribution for Females exhibiting a heavier tail.

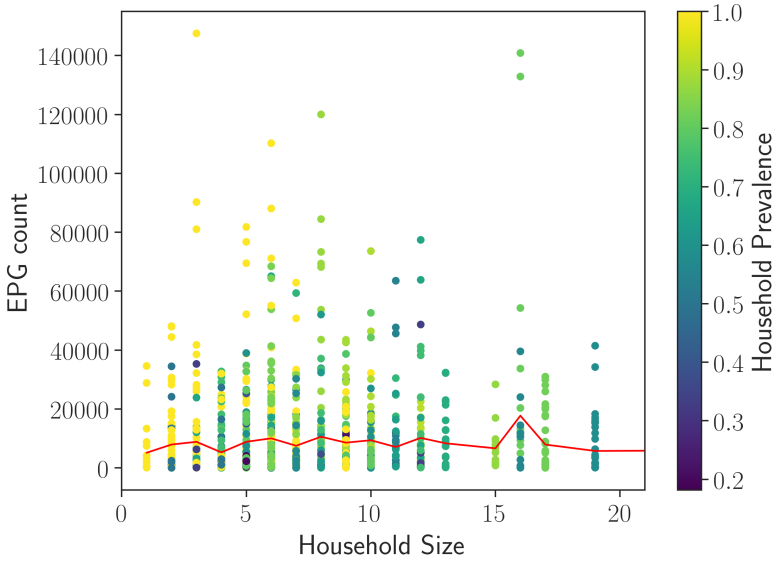

Figure 4.8: Size of the household of a resident plotted against their Ascaris eggs per gram of faeces (EPG) count, coloured by the prevalence within their household. The red line denotes the mean EPG for each household size.

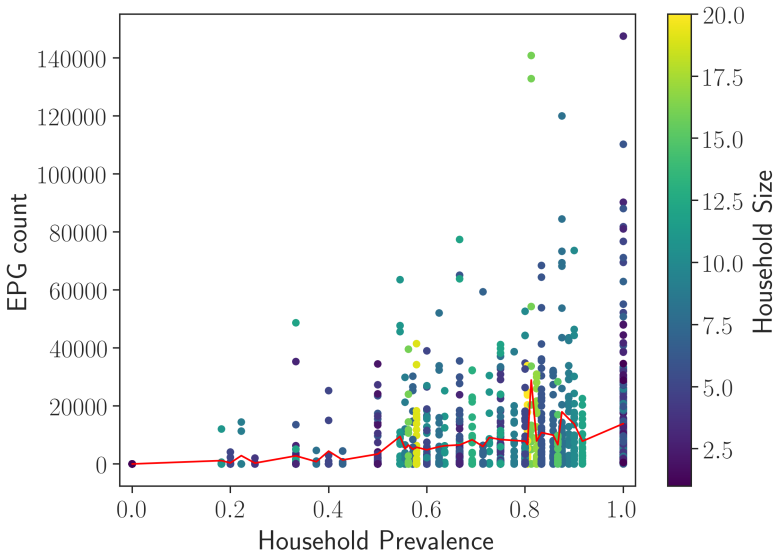

Figure 4.9: The prevalence within the household of an individual plotted against their Ascaris eggs per gram of faeces (EPG) count, coloured by Household size. The red line denotes the average EPG at a given household prevalence.

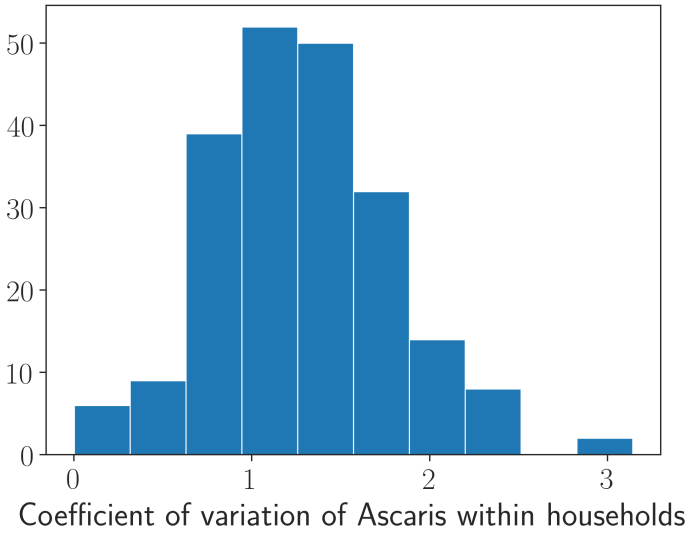

Figure 4.10: The distribution of the coefficient of variation of Ascaris eggs per gram of faeces (EPG) count within households.

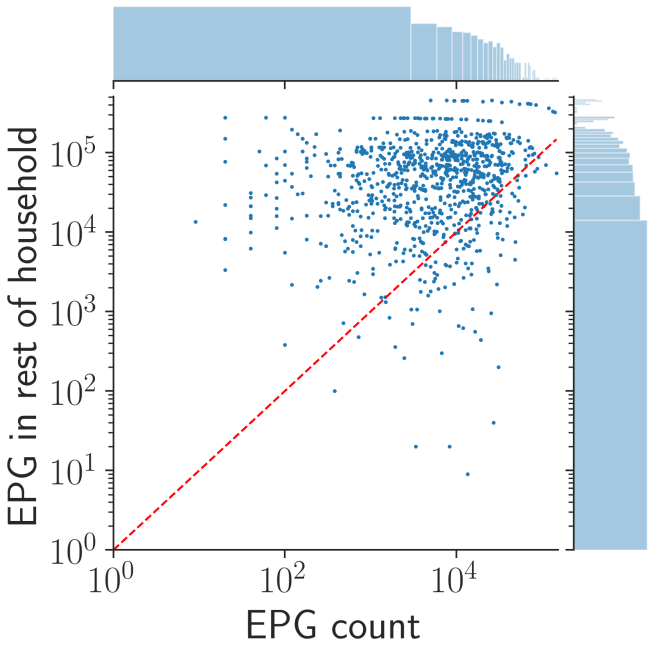

Figure 4.11: Joint distribution of the Ascaris eggs per gram of faeces (EPG) count of individuals against the total EPG of other individuals in their household. The two histograms give the marginal distributions of each axis (logarithmic density). The dashed red line corresponds to the individual having the same EPG as the rest of the household.

While no observable difference in prevalence is evident, there does appear to be a difference in EPG count between Males and Females with the Female distribution having a higher mean due a heavier tail (figure 4.7). A two-sample Kolmogorov-Smirnov test suggests that the two distributions are not the same ($p = 2.76 \times 10^{-4}$).

As previously mentioned the large household sizes could have a large effect on the spread of Ascaris if the epidemic dynamics were density dependent. Fortunately, figure 4.8 suggests that the dynamics are frequency dependent as there doesn't appear to be a difference in the distribution of Ascaris EPG across household sizes.

A clearer relationship between the distribution of EPG and the prevalence within a Household is present (figure 4.9) with a larger household prevalence leading to a broader distribution of EPG counts. In addition, household size does not seem to be a factor in this relationship.

Finally, figure 4.10 demonstrates the over-dispersion of EPG within households as the coefficient of variation for the majority of households is greater than unity. Figure 4.11 tries to gain insight into the structure of this over-dispersion by plotting the joint distribution between the Ascaris EPG of an individual and the total EPG in the rest of their household (excluding that individual). There is no discernible relationship, perhaps because this plot censors any individuals who do not have an infection, though we observe that the minority of individuals have a higher EPG than the rest of their households!

With the key features of the dataset explored we now proceed to begin investigating the aggregation of Ascaris in more principled ways using statistical modelling.

4.3 Hierarchical model of Ascaris infection in a population

Many statistical applications involve multiple parameters that are related, implying a joint probability model should reflect this dependence. For example, in a study of the aggregation parameter of worms in a population, with the occupants of village k having aggregation parameter θ_k, it is not unreasonable to assume that the estimates of θ_k's are related to one another as they share the same biological process. This is naturally achieved if we view the θ_k's as a sample from a common population distribution. We model this problem hierarchically with observations modelled conditionally on certain parameters, called hyper-parameters, which themselves are given a probabilistic nature. Our nested hierarchy consists of villages, households, and individuals. One key advantage of tackling the problem this way is that non-hierarchical models are ill suited for hierarchical data: with few parameters

they cannot capture the underlying complexity of the problem, whereas with many parameters they tend to overfit [65]. Hierarchical models can sidestep this issue by using a population distribution to structure some dependence into the model, avoiding overfitting by constraining extreme values within the population distribution, and shrinking the estimates for units for which there may be little data to the population mean.

To illustrate this point, lets revisit the example mentioned above with aggregation parameter θ_k. One could model the aggregation of the EPG, y_{ijk}, of villager i in household j in village k with a negative binomial in 3 different ways:

$$y_{ijk} \sim \text{NB}(\mu_{ij}, \theta) \tag{4.1}$$

$$y_{ijk} \sim \text{NB}(\mu_{ij}, \theta_k); \tag{4.2}$$

$$y_{ijk} \sim \text{NB}(\mu_{ij}, \theta_k); \theta_k \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2) \tag{4.3}$$

μ_{ij} is some log-linear effects model based on individual and household covariates which is not relevant to this explanation. The first is a fully pooled (non-hierarchical) model, where we have too few parameters as we assume that all villages have the same aggregation parameter. This is an oversimplification as we have no reason a priori to believe this, particularly given figures 4.4 and 4.6 observe heterogeneity in prevalence and household sizes across villages which are likely to affect aggregation. The second is an unpooled model with too many parameters to independently fit, one for every (of potentially many) villages which can lead to overfitting as each village (or whatever hierarchical unit is being considered for the problem at hand) may not contain enough data to accurately estimate an aggregation parameter for each independently. This approach amounts to fitting many separate regressions. The third is a 'partially pooled' hierarchical model which complements the previous two approaches: there are more parameters than the fully pooled model, allowing greater complexity to be captured by the model; however the parameters are not independently estimated but come from some population distribution specified by hyper-parameters μ_θ and σ_θ^2 which enables the sharing of information between villages. This approach is in fact a generalisation of the first two. As $\sigma_\theta^2 \to 0$, model 3 becomes equivalent to model 1, with $\theta_j = \mu_\theta \forall j$. Similarly as $\sigma_\theta^2 \to \infty$, model 3 becomes equivalent to model 2.

We perform a hierarchical negative binomial regression (4.5)-(4.11) on our dataset, testing several model formulations and comparing their goodness of fit. We consider a similar structure to our illustrative example in (4.3) above. We assume

the underlying distribution to be a zero-inflated negative binomial (ZINB) distribution, i.e. a mixture model where with probability $(1-\psi)$ the EPG is zero, and with probability ψ the EPG is drawn from a negative binomial distribution - this captures the aggregation of Ascaris that closely approximates a negative binomial [2, 5]. The negative binomial distribution arises as a mixture of a Poisson and Gamma distribution (4.4) with the Poisson rate parameter distributed as a Gamma random variable - in other words the number of worms (or EPG) someone has is based upon some heterogeneous Poisson rate that varies e.g. due to a difference in exposure or some pre-disposition to infection.

$$f_{\mathrm{NB}(\mu,\alpha)}(n) = \int_0^\infty f_{\mathrm{Poisson}(\mu)}(n) f_{\mathrm{Gamma}(\alpha,\alpha/\mu)}(\mu) d\mu \tag{4.4}$$

$$y_{ijk} \sim \mathrm{ZINB}(\mu_{ij}, \psi_{ik}, \alpha_k) \tag{4.5}$$

$$\ln(\mu_{ij}) = \beta_{0j} + \beta_1 x_{1j} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_{2+A} x_{(2+A)i} \tag{4.6}$$

$$\beta_{0j} \sim \mathrm{Normal}(\mu = \mu_{HH}, \sigma = \sigma_{HH}) \tag{4.7}$$

$$\beta_1, \beta_2 \sim \mathrm{Normal}(\mu = 0, \sigma = 5) \tag{4.8}$$

$$\beta_3, \ldots, \beta_{2+A} \sim \mathrm{Normal}(\mu = 0, \sigma = \sigma_A) \tag{4.9}$$

$$\alpha_k, \mu_{HH}, \sigma_{HH}, \sigma_A \sim \mathrm{HalfCauchy}(5) \tag{4.10}$$

$$\psi_{ik} \sim \mathrm{Uniform}(0, 1) \tag{4.11}$$

When performing inference β_{0j} is given a population mean of zero and this mean is inferred separately in what is called a non-centred parameterisation. This has been shown in the literature to provide a posterior geometry that is much more easily explored [155]. x_{1j} denotes the prevalence within household j, x_{2i} is an indicator variable for gender that is one if Female, $x_3 \ldots x_{2+A}$ are indicator variables for being in one of A age groups. We choose $A = 13$ with age groups (0-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-16, 17-26, 27-36, 37-46, 47+) as in Walker et al. [200].
Aggregation α_k is fitted independently for each village, allowing for a difference in aggregation over villages. Similarly, ψ_{ik} allows for variation over villages; however it also allows for variation based on the age group of the individual i. This is because the prevalence varies greatly with age (figure 4.13) and fitting a model without accounting for this heterogeneity constrains the model in such a way that it yields a joint distribution between the aggregation and prevalence parameters that induces a false linear relationship (figure 4.12) where a higher prevalence leads to lower

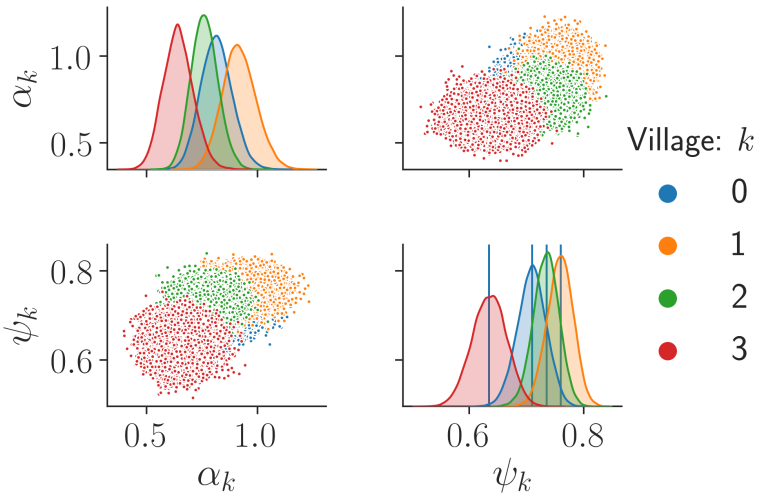

Figure 4.12: Joint distribution of aggregation (α_k) and the prevalence (ψ_k) for each village for a model where aggregation and prevalence vary across villages but prevalence does not vary with age. This model induces a false linear relationship between aggregation and prevalence due to the fact that prevalence varies with age.

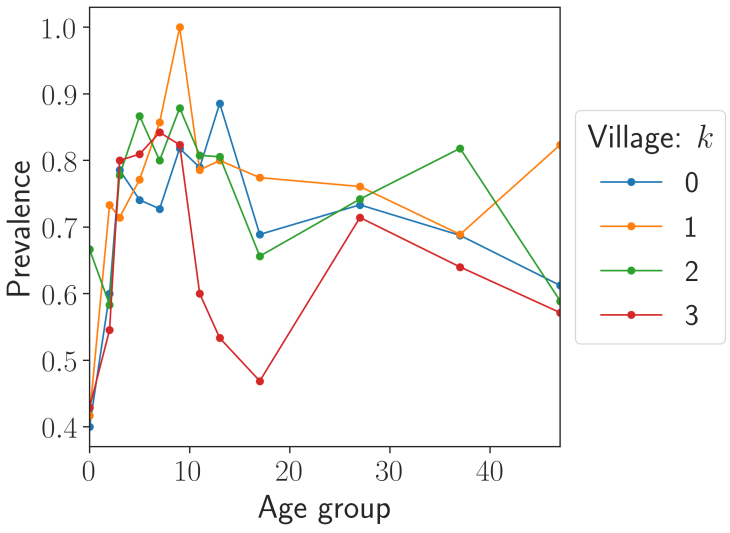

Figure 4.13: Relationship between age and prevalence stratified by Village. Ages are grouped into 13 groups (0-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-16, 17-26, 27-36, 37-46, 47+) as in Walker et al. [200].

aggregation (higher α) which disappears when prevalence is allowed to vary with age.

	LOO	pLOO	dLOO	weight
Age + Sex + HH + HH prevalence; ψ_{ik}	21201.5	124.69	0	0.56
Age + Sex + HH + HH prevalence	21210	84.59	8.44	0.25
Age + Sex + HH	21216.6	86.28	15.04	0.18
Age + HH	21245.7	85.18	44.13	0
Age + Sex + HH prevalence	21275.7	22.73	74.21	0
Age + Sex + HH + HH prevalence; $\psi = 0$	21954.2	15.33	752.65	0.01

Table 4.1: Model comparison. LOO denotes the LOO score (lower is better), pLOO denotes the model complexity penalty, dLOO denotes the difference between a model and the best performing model, and weight denotes the model weighting using model stacking [206].

We fit our model with MCMC using the probabilistic programming package pyMC3's [175] implementation of the gradient based No U-Turn Sampling (NUTS) algorithm (see section 6). We compare multiple variants of our model with various features excluded to check a simpler model would not be more permissible, this is performed using PSIS-LOO (section 2.3.4.2) with the results shown in table 4.1. A traceplot of the fit and convergence diagnostics may be found in figure 4.15.

Figure 4.14 shows the fit of the model using posterior predictive checks showing both the posterior means of the mean EPG and standard deviation of EPG against the data. In addition to seeing that the model fits both of these well, we see that a regular (not zero-inflated) negative binomial distribution captures well the mean EPG count but completely fails to model the standard deviation of EPG - this is also reflected in it having the lowest LOO score.

4.4 Discussion

We have shown that Ascaris EPG count can be well approximated by a zero-inflated negative binomial distribution with both age and household having significant effects. The large household effects provide evidence that households indeed play an important role in generating aggregation, though our exploration of the data suggest that the size of the household does not appear to have an effect on EPG. The main limitation of this modelling approach is that empirical models are conditioned on the independence of observations which is violated by a spreading process. Therefore,

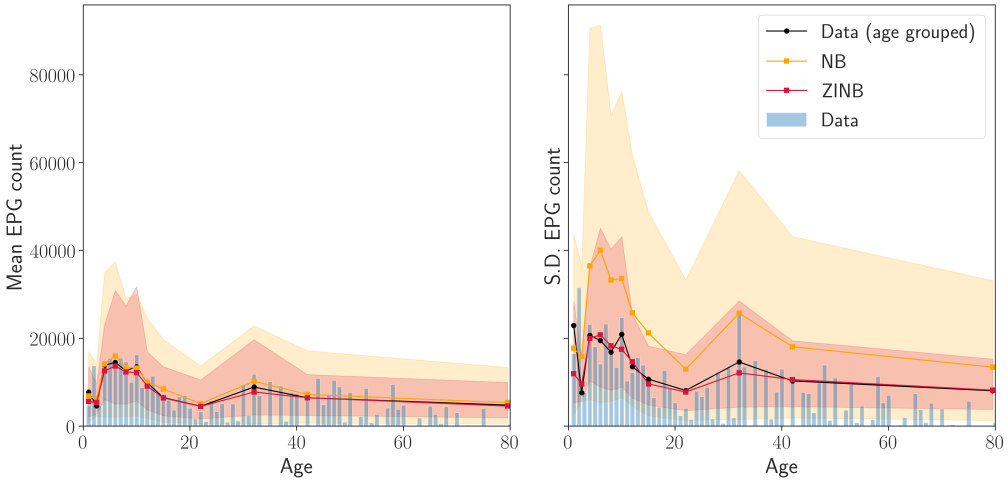

Figure 4.14: Posterior Predictive mean (left) and standard deviation (right) for our full model (ZINB), and our model with $\psi = 0$ (NB). Posterior predictive intervals are plotted with the posterior mean, and 95% Bayesian credible intervals. The age grouped data in black show the age-group data that the model was fitted to, and the blue bars show the data before age-grouping.

while these models are insightful, flexible, and relatively straightforward to fit they do not capture the most important component of a spreading process. In the next chapter we develop a methodology to specify and fit a general transmission model which accounts for both age and household structure. This more dynamic analysis should be more capable of answering the question of whether it is children or adults who are the major drivers of transmission by inferring quantities such as the infectiousness of different age classes and comparing them. Further work on empirical models could focus on the link between EPG and worm burden. With access to data that measures both EPG and worm burden in a population, we could augment the model with an extra process that given an observation of the EPG count of an individual and their individual covariates, such as age and gender, yields a distribution of worm burden which can be fed into the rest of the model. This would mitigate the limiting assumption that EPG is directly proportional to worm burden. Finally, yet another direction further analysis could take is to analyse the EPG counts for Trichuris and hookworm, which we have not analysed. A multi-variate regression model of the three EPG counts could provide insight on the co-morbidity of different STH infections at the household level. Non-parametric techniques, such as Gaussian process regression, may prove promising for such an analysis.

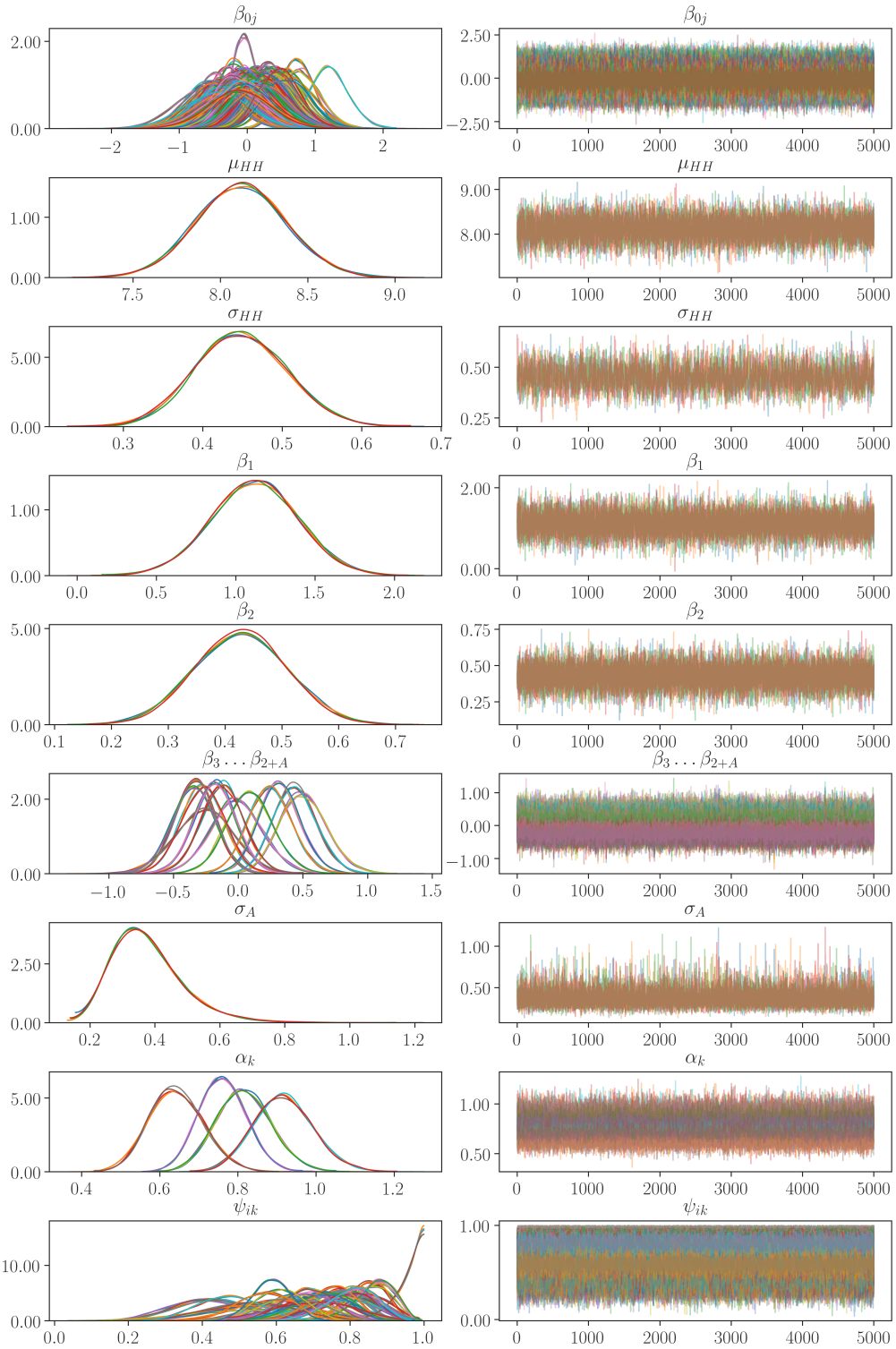

Figure 4.15: Posterior traceplot for the zero-inflated negative binomial model. The model was tuned for 5000 iterations, and then 5000 samples were taken from each of 4 chains. $\widehat{R} \in [0.9999, 1.0013]$, and the smallest number of effective samples was > 2000, with no divergences observed indicating a well converged and sampled model.

CHAPTER 5

Household epidemic models for STH

5.1 Introduction

The previous chapter demonstrated that household structure plays an important role in determining the prevalence and intensity of Ascaris infection. There are two key limitations to performing this analysis with an empirical model, namely that non-independence of observations cannot not modelled, and that forward projections of the transmission dynamics under different interventions cannot be performed to determine the effectiveness of various control strategies.

This chapter aims to bridge the gap between stochastic meta-population epidemic models and previous disease modelling of Ascaris which has focused on either Anderson & May type macro-parasite models (section 2.1.4) or empirical statistical analyses (such as that of the previous chapter). The necessity of incorporating the meta-population (i.e. household) structure is self-evident in this context, but it is equally crucial that this is done in a stochastic setting due to both the small sizes of the households and the fact that in order to make statements about the probability of extinction under some control measure one requires the ability to accurately capture the large stochastic fluctuations that occur close to the transmission breakpoint. The key challenge here lies in the trade-off between specifying a stochastic "household model" that is complex enough to capture both age dynamics and multiple levels of infection whilst simultaneously being amenable to inference by having a tractable likelihood that does not take long to compute. As we shall see, as a model becomes increasingly complex then inference very quickly becomes infeasible. This is due to both the time taken to calculate the likelihood increasing polynomially in both the number of age and infection levels, and due to the increasing number of model parameters needing to be inferred from a small dataset.

Household models [18] mean the key unit of structure in the population (the household) can be accurately modelled and parameterised using data whilst maintaining mathematical tractability. In reality a network model with community structure would provide a more realistic model though its high dimensional nature means that such a model would not be amenable to inference and would be incredibly hard to collect relevant epidemiological data for. This structure is simply modelled by considering two levels of mixing with transmission occurring faster between individuals in the household than transmission between individuals from different households. Typically household models have individuals with a homogeneous risk of infection from outside the household but consider more complex dynamics within the household.

A key feature and motivating factor behind the use of stochastic household models is their ability to generate "U"-shaped prevalence distributions where there is a high probability of the prevalence being both high and low. For instance, figure 5.1 demonstrates this by considering a household with stochastic SIS dynamics (section 2.1.2) where β governs the within household spread, ϵ represents import of infection into the household, and $\beta \gg \epsilon$. This matches well with the shape of the prevalence curve that we observed for our data in figure 4.4.

The overwhelming majority of the literature relating to performing inference on household models such as [28, 30, 39, 42, 138] focus on SIR-type dynamics because the lack of recovery (or the slow time scale of recovery in models with waning immunity) allows the use of several methods which cannot be applied for models with immediate reinfection. House et al. [92] give a comprehensive overview of numerous ways to calculate the final size distribution for SIR-type dynamics, such as Sellke's method [177] and Bailey's method [14, 140] demonstrating how the deployment of appropriate numerical methods can extend the reach of machine-precision methods for doing Bayesian inference; however it does not discuss these methods or others in the context of SIS-type dynamics. This 'gap' in the literature is what the following two chapters attempt to fill, using the open questions around the role of households in the transmission and control of Ascaris as a guiding example.

In this chapter we introduce and extend the SIS household model [173] and apply an approximation in order to be able to write down the likelihood of this model which allows us to perform Bayesian inference without having to resort to likelihood-free methods which have been demonstrated to be susceptible to significant bias without careful choice of both algorithm and summary statistic [57]. Having inferred the model posterior, we estimate the ratio of within versus between house-

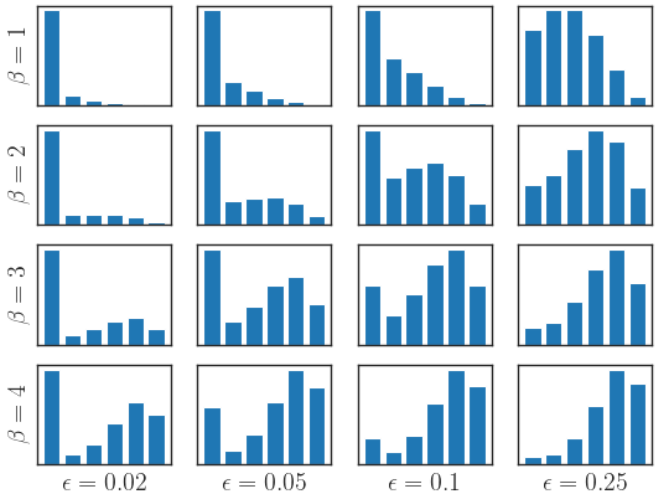

Figure 5.1: Stationary distribution of the number of individuals infected in a stochastic SIS model ($N = 5$) for varying transmission rate β and external force of infection rate ϵ. With β and ϵ both low, there is a high probability of zero prevalence; with β and ϵ both high, the stationary distribution is uni-modal with the prevalence most likely to be medium-high; however, for many intermediate values the distribution is "U" shaped, with a high probability of the prevalence being both high and low.

hold transmission events, perform forward simulations of various control strategies in order to estimate a Bayesian distribution of the time to extinction under various coverage levels and treatment intervals in order to remark on their effectiveness.

5.2 Stochastic SIS household model

To illustrate our use of stochastic household models we begin with the stochastic SIS model [173] (with no age structure) as a simple model for prevalence of Ascaris (section 2.1.2).

5.2.1 Construction of model

We begin by considering one household unit in isolation. The household contains N people who are either susceptible (state S) or infected with worms (state I). Infections are imported into the household according to some time-varying rate $\lambda_g(t)$ and an infectious individual within the household exerts infectious pressure on a susceptible individual at rate $\beta/(N-1)^\alpha$. β is the within-household transmission

rate, and $\alpha \in [0, 1]$ governs the extent to which transmission is frequency-dependent ($\alpha = 1$) or density-dependent ($\alpha = 0$). Typically the assumption that within household transmission is much stronger than between household transmission is made - exactly the relationship we wish investigate in the data from the previous chapter. Proceeding as in section 2.1.2, the Master/Kolmogorov equations of this model are specified by writing down an ODE for the probability of finding the population in every possible state. This set of $N + 1$ ODEs are almost identical to those given in (2.6) albeit with two minor generalisations. Firstly, what was previously a constant external force of infection, ϵ, is now a time varying rate $\lambda_g(t)$ to account for time-varying infectious pressure from other households. The second generalisation is the addition of α to the term governing the infectious pressure an infected individual asserts on an individual in the household - previously α was not defined and was implicitly taken to be unity. The corresponding dynamics are given in (5.1) in their deterministic form (for ease of readability).

$$\frac{dI}{dt} = \left(\frac{\beta I}{(N-1)^\alpha} + \lambda_g(t) \right) S - \gamma I \tag{5.1}$$

The interaction between m different households occurs through the function $\lambda_g(t) = \delta J(t)$, where δ denotes the external infectious pressure upon a susceptible individual in a fully infected population, $J(t) = \frac{\sum_{h=1}^{m} I_h(t)}{\sum_{h=1}^{m} N_h}$ is the global prevalence in the population, $I_h(t)$ denotes the number infected in household h, and N_h is the size of household h.

As the number of households being considered becomes large ($m \to \infty$) and as the system reaches equilibrium ($t \to \infty$) then the global prevalence converges to a fixed value J^*. This results in the external force of infection becoming constant, $\lambda_g(t) \to \delta J^* = \epsilon$. The system of m interacting households which were before a set of time inhomogeneous Markov processes have now been reduced to a set of independent time homogeneous Markov processes whose stationary distributions may be efficiently evaluated using either the matrix exponential (section 2.2.2 - for large enough t) or by finding the eigenvector corresponding the zero eigenvalue (section 2.2.3.2).

Throughout the remainder of this chapter we make the assumption that we are dealing with a constant external force of infection, $\lambda_g(t) = \epsilon$. This is valid if the cross-sectional observation of our data (section 4.2) was collected from the system in equilibrium ($\frac{d\mathbf{p}}{dt} = 0$), and the population is large enough that fluctuations in ϵ are small - in reality m does not have to be much larger than 25. This approximation now allows us to calculate the likelihood of our household model as a sum of independent

terms of the stationary distributions obtained by computing the null eigenvector of the transition matrices of each household.

Given that we are now able to compute the likelihood of the data given a set of parameters to machine-precision we can use Markov Chain Monte Carlo (MCMC - section 2.3.3) methods to infer the posterior distribution of the model parameters - it is necessary to set $\gamma = 1$ in order to set the time-scale of the system otherwise the model would not be identifiable. Calculation of the posterior distribution facilitates analysis of the ratio of between household versus within household spread, and simulation of the effectiveness of various control strategies (which may exploit the inferred dynamics) using samples from the posterior. Simulation of control strategies from the posterior in turn allows the calculation of the posterior distribution of extinction times.

Whilst an incredibly simplistic model compared to the complex biology of Ascaris, the stochastic household SIS model provides many benefits over an empirical approach, such as a generalised linear mixed model (GLMM), when analysing household clustering of infections. For example, the ability to consider a caricature of the process generating the data rather than a statistical description assuming independence of observations. However, the model is obviously too simple to accurately answer all the questions we may want to ask, such as how the interaction between adults and children (who exhibit different worm distributions) affects the dynamics. It is therefore necessary to extend this modelling approach such that it captures these more realistic infection dynamics.

5.3 Defining a general model and multi-dimensional Markov chains

Extending the stochastic household SIS model to a model with A age classes and R infection levels of increasing severity requires a generic specification that is capable of generating the appropriate set of ODEs, the corresponding transition matrix of the stochastic process, and a two way mapping between the multi-dimensional model state space and the one-dimensional index of the transition matrix and stationary distribution. Once we have a generic specification we can perform inference and forward simulations on a model of arbitrary A and R - at least until A and R become so large that inference becomes computationally infeasible due to the size of the stochastic transition matrix scaling polynomially in both A and R.

We may define a model state by the matrix $\boldsymbol{X}$:

$$\mathbf{X} = \begin{pmatrix} S_1 & \cdots & S_A \\ I_{11} & \cdots & I_{1A} \\ \vdots & \ddots & \vdots \\ I_{R1} & \cdots & I_{RA} \end{pmatrix} = \begin{pmatrix} \mathbf{s} \\ \mathbf{I} \end{pmatrix} \in \mathbb{N}^{R+1 \times A} \tag{5.2}$$

$$\text{vec}([\mathbf{x}_1^T, \cdots, \mathbf{x}_n^T]) = [\mathbf{x}_1, \cdots, \mathbf{x}_n]^T \tag{5.3}$$

$$N = \|\text{vec}(\boldsymbol{X})\|_1 \tag{5.4}$$

Where S_a is the number of susceptibles in the a^{th} age class and I_{ra} is the number infected in the r^{th} level of infection of infection severity in age class a. Our transmission (β) and recovery (γ) parameters from the SIS model now become matrices with different parameters between each age-risk class combination,

$$\boldsymbol{\beta} = \begin{pmatrix} 0 & \cdots & 0 \\ \beta_{11} & \cdots & \beta_{1A} \\ \vdots & \ddots & \vdots \\ \beta_{R1} & \cdots & \beta_{RA} \end{pmatrix} \in {\mathbb{R}_+}^{R+1 \times A} \tag{5.5}$$

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & \cdots & 0 \\ \gamma_{11} & \cdots & \gamma_{1A} \\ \vdots & \ddots & \vdots \\ \gamma_{R1} & \cdots & \gamma_{RA} \end{pmatrix} \in {\mathbb{R}_+}^{R+1 \times A} \tag{5.6}$$

The first row of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are zero as susceptibles neither transmit or recover. In the last section we noted that setting γ equal to unity defines the time-scale of the system if we are only interested in the stationary distribution. Here we may do the same, for example by setting γ_{11} equal to unity. In addition to adding a more stratified age and infection severity structure, we also add a susceptibility matrix, $\boldsymbol{\rho} = [\rho_{ra}]$. ρ_{ra} denotes the susceptibility of an individual in age class a to the r^{th} stage of infection (given of course that they are presently in the $(r-1)^{th}$ stage of infection). For example, if $\rho_{11} = 2$ and $\rho_{12} = 1$ then the first age class is twice as susceptible to the first level of infection as the second age class under the same force of infection. This addition is a simplistic representation of complex latent/unmeasured processes such as behavioural differences or natural history of the disease considered. The last row of $\boldsymbol{\rho}$ is zero as maximally infected individuals have no higher level of infection to be susceptible to.

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1A} \\ \vdots & \ddots & \vdots \\ \rho_{R1} & \cdots & \rho_{RA} \\ 0 & \cdots & 0 \end{pmatrix} \in {\mathbb{R}_+}^{R+1 \times A} \tag{5.7}$$

Given a state, $\text{vec}(\mathbf{X})$, we can write down the rate of change of the probability of being in this state, $P_{\text{vec}(\mathbf{X})}$, under our generic model:

$$\begin{aligned} \frac{dP_{\text{vec}(\mathbf{X})}}{dt} = \sum_{a=1}^{A}\sum_{r=1}^{R} \Big(& \boldsymbol{\Lambda}\rho_{(r-1),a}\big(\mathbf{X} + \boldsymbol{\Theta}(r-1,a)\big)_{(r-1,a)} P_{\text{vec}(\mathbf{X}+\boldsymbol{\Theta}(r-1,a))} \\ & - \boldsymbol{\Lambda}\rho_{r,a}\mathbf{X}_{r,a}P_{\text{vec}(\mathbf{X})} \\ & + \gamma_{(r+1),a}\big(\mathbf{X} - \boldsymbol{\Theta}(r+1,a)\big)_{(r+1),a} P_{\text{vec}(\mathbf{X}-\boldsymbol{\Theta}(r+1,a))} \\ & - \gamma_{r,a}\mathbf{X}_{r,a}P_{\text{vec}(\mathbf{X})} \Big) \end{aligned} \tag{5.8}$$

$$\boldsymbol{\Lambda} = \frac{\boldsymbol{\beta}\cdot\mathbf{X}}{(N-1)^{\alpha}} + \epsilon \tag{5.9}$$

$$\boldsymbol{\Theta}(r,a) = \boldsymbol{\Theta}(r,a;R,a) = (\hat{\mathbf{e}}^T_{R;(r-1)}\cdot\hat{\mathbf{e}}_{A;a}) - (\hat{\mathbf{e}}^T_{R;r}\cdot\hat{\mathbf{e}}_{A;a}) \tag{5.10}$$

$$\hat{\mathbf{e}}_{X;0} = [0,\cdots,0] \in \{0,1\}^X \tag{5.11}$$

$\boldsymbol{\Lambda}$ gives the total force of infection. $\hat{\mathbf{e}}_{X;i}$ is the indicator function of length X with a one in the i^{th} position. $\boldsymbol{\Theta}(r,a)$ corresponds to an increase of one person in the $(r-1,a)^{th}$ state and a decrease of one person in the $(r,a)^{th}$ state, for example with $(R,A,r,a) = (3,2,2,1)$:

$$\boldsymbol{\Theta}(r,a) = \boldsymbol{\Theta}(2,1) = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 0 \end{pmatrix} \tag{5.12}$$

$$\mathbf{X} + \boldsymbol{\Theta}(2,1) = \begin{pmatrix} S_1+1 & S_2 \\ I_{11}-1 & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \tag{5.13}$$

Using this notation then a transition from $\mathbf{X} + \boldsymbol{\Theta}(2,1)$ to $\mathbf{X}$ denotes the transition $(S_1+1, I_{11}-1, I_{21}, S_2, I_{12}, I_{22}) \to (S_1, I_{11}, I_{21}, S_2, I_{12}, I_{22})$.

Given that we have now written down the master equations for a general

household stochastic transmission model with A age classes and R infection levels with susceptibility ($[SI_RS]_A$), we must now ponder how to construct the transition matrix ($\boldsymbol{Q}$) for this general stochastic process.

5.3.1 Remarks on model formulation

Given we are interested in calculating the stationary distribution of a household model we must first show that one exists. A stationary distribution for a Markov chain exists (section 2.2.3.1) if and only if it is irreducible and every state is recurrent with a finite mean recurrence time. We argue without proof that this generic model specification has a stationary distribution provided all parameters are strictly positive. Given both the discrete, finite state space and the strictly positive parameters it is trivial to see from the master equations that every state is reachable from every other in a finite time.

As we formulated this generalised household model in the context of Ascaris we have labelled the demographic and infectious levels as 'Age' and 'Risk' levels respectively; however these apply more generally. For example, instead of representing age levels we could have different species, different genders etc. Furthermore, the infectious levels could all correspond to one true level of infection but with the model levels representing a variability in infectiousness such as through a latent period or constant infection but with a non-Markovian infectious period. Finally $\boldsymbol{\gamma} = \mathbf{0}$ yields SI-type dynamics.

5.4 Machine Implementation

In addition to a mathematical description of the model we must consider how to describe the model such that a computer may solve it. This section describes efficient ways of mapping from a model state, $\boldsymbol{X}$, to the integer index of the Markov chain state-space, and a way to map from the integer index back to a model state. We need to map from a model state to an index such that we can construct the transition matrix of the Markov chain using the rates defined by (5.8), and we need to map from an integer index back to an infection state so that we can efficiently compute posterior distributions for variables of interest - e.g. to compute the posterior probability that there are more than, say, one person infected in a household then we need to know which entries in the posterior distribution correspond to such infection states. Efficient python implementations of the mappings specified in this section can be found at `https://github.com/bishax/snepits`.

5.4.1 Mapping model state to natural numbers

The first consideration is how to encode our model states, $\boldsymbol{X}$, as rows/columns of the transition matrix. We choose a lexicographic ordering on the cartesian product $\boldsymbol{x}_1 \times \boldsymbol{x}_2 \times \cdots \times \boldsymbol{x}_A$ where $\boldsymbol{x}_i$ represents the states of age group i as in (5.2). Furthermore, for R risk levels then each age group of population n has a lexicographical ordering on the n-subsets of an $R+1$-set imposed on it.

For example, we shall see that given $A = 2, R = 2, N_1 = 1, N_2 = 2$ the ordering of the states $(S_1, I_{11}, I_{12}, S_2, I_{12}, I_{22})$ begins as:

$100200, 100110, 100101, 100020, 100011, 100002, 010200, 010110,$

$010101, 010020, 010011, 010002, \ldots, 001020, 001011, 001002.$

Starting with one age class ($A = 1$) and for general $R \in \mathbb{Z}_+$, one may find the element by,

$$\text{element} = \sum_{r=1}^{R} \frac{1}{r!} \left(\prod_{j=0}^{r-1} \left(\sum_{i=R-r+1}^{R} I_{1i} \right) + j \right) \tag{5.14}$$

Where noting that,

$$\frac{1}{r!} \left(\prod_{j=0}^{r-1} \left(\sum_{i=R-r+1}^{R} I_{1i} \right) + j \right) = \binom{\sum_{i=R-r+1}^{R} (I_{1i}) + r - 1}{r} \tag{5.15}$$

Allows simplification to,

$$\text{element} = \sum_{r=1}^{R} \binom{\sum_{i=R-r+1}^{R} (I_{1i}) + r - 1}{r} = \sum_{r=1}^{R} \binom{\iota(1,r)}{r} \tag{5.16}$$

Where,

$$\iota(a, r) = \sum_{i=R-r+1}^{R} (I_{1i}) + r - 1 \tag{5.17}$$

This may now be generalised to the case where $A \in \mathbb{Z}_+$:

$\binom{N_b+R}{N_b} = \left(\binom{R+1}{N_b} \right)$ represents the number of N_b-multisets from a $(R+1)$-set, or the size of the state matrix for one age class of size N_b and R infectious levels (and a susceptible class).

$stepsize(a) = \prod_{b=a+1}^{A} \binom{N_b+R}{N_b}$ represents the product of the sub-matrix sizes of age classes $a+1, \ldots, A$ such that the next lexicographic ordering of the state will result in an increase of this amount for the index.

To find the element contribution for the a^{th} age class we take the element contribution

as if there was only one age class $\sum_{r=1}^{R} \binom{\iota(a,r)}{r}$ and multiply it by the product of the sub-matrix sizes of age classes $a+1, \ldots, A$. Summing up the contributions for each age class yields:

$$\text{element} = \sum_{a=1}^{A} \left(\prod_{b=a+1}^{A} \binom{N_b + R}{N_b} \right) \sum_{r=1}^{R} \binom{\iota(a,r)}{r} \tag{5.18}$$

5.4.2 Mapping natural numbers to model state

The ordering defined by the mapping from model state to (unsigned) integer in the previous section is such that it is not possible to simply write down an analytical expression for the inverse; however we may efficiently enumerate all states in lexicographical order and create a hash-table for efficient lookup.

To map our integer index back to a model state we will make use of a mapping between the binary numbers with R 1-bits set and our model state for one age-group ($A = 1$). In this mapping a 0 corresponds to one individual in the "bins" which are separated by 1's. The number of individuals in the currently considered age class is equal to the number of 0's in the binary representation and the number of 1's corresponds to the number of infection levels.

For example the binary representation of 22 is 0010110, which would represent two individuals in the first "bin" as we have two 0's at the start. The 1 denotes that we move to the next bin where there is one individual (one 0 before the next 1). In the third bin there are no individual as we have sequential 1's. In the final bin we have one individual.

This representation is convenient because we can very efficiently compute the lexicographically next bit permutation using bit-wise operations. The first permutation will be $v = 2^R$ which corresponds to N 0-bits followed by R 1-bits (representing a fully susceptible population). From this start point we can calculate the next permutations as follows:

$$t = v | (v - 1) \tag{5.19}$$

$$z = \#\text{ trailing zeros of v} = \_\_builtin\_ctz(v) \text{ for GCC (x86)} \tag{5.20}$$

$$w = (t+1) | \left(((\sim t \,\&\, -\sim t) - 1) >> (z+1) \right) \tag{5.21}$$

Where $|$, $\&$, and $\sim$ denote bit-wise or, bit-wise and, and bit-wise complement respectively. N-bit signed integers are assumed to be represented using two's

complement. For example, the two's complement of the binary number $111b$ is $001b$ because $111b + 001b = 1000b$.

A step by step example illustrates the effect of these operations:

$$\begin{aligned}
v &= 10010110\underline{0000} \text{ Represents: (0, 2, 1, 0, 4)} \\
t = v|(v-1) &= 1001\underline{011}\overline{1111} \text{ Flip trailing 0's} \\
t+1 &= 1001\underline{1}\overline{000000} \text{ Flip trailing 1's and rightmost 0} \\
\sim t \& - \sim t &= 0000\underline{1}\overline{000000} \text{ Rightmost bit is only bit of } t+1 \\
t' = (\sim t \& - \sim t) - 1 &= 0000\underline{0}\overline{111111} \text{ Flip trailing 0's and rightmost 1} \\
t' >> z+1 &= 0000\overline{000000}1 \text{ Flip } z+1 \text{ leftmost 1's} \\
(t+1)|(t' >> z+1) &= 10011000001
\end{aligned}$$

This enables the efficient enumeration of numbers with R 1-bits set in lexicographic order, allowing the construction of a hash-table for fast lookup with the natural number denoting the ordered index as the key and the binary representation of the state vectors being the values. We build a hash-table for each unique value of the age-class populations $[N_1, \ldots, N_A]$ and denote the hash-table for an age-class population of size n by h_n, such that $h_n[i]$ will give the state corresponding to the i^{th} lexicographical permutation.

All that remains is to extend this method from the ability to compute the index corresponding to a model for $A = 1, R \in \mathbb{Z}_+$ to one that computes the index corresponding to a model for $A, R \in \mathbb{Z}_+$. This is achieved by exploiting the block structure introduced by the ordering of the previous subsection. $\binom{N_a+R}{N_a}$ gives the sub-matrix size for age class a, and as in the previous section $stepsize(a)$ represents the product of these sub-matrix sizes for age classes $a+1, \ldots, A$.

Given an index i, we can calculate an index for age a, i_a, representing the index for a model of only one age class containing N_a individuals and use this index along with the hash-table to calculate the state vector for the a^{th} age class:

$$i_a = i//stepsize(a))\%subsize(a, R) \tag{5.22}$$

$$[S_a, I_{a1}, \ldots, I_{aR}] = f(h_{N_a}[i_a]) \tag{5.23}$$

Where $f(\cdot)$ is the function that converts from a binary representation to a state vector which can be implemented using similar bit twiddling hacks as before.

Putting this all together we can calculate the model state from the index

efficiently using algorithm 3.

Algorithm 3 Calculate model state from index

Input: i, A, R, $N = [N_1, \ldots, N_a]$
for a=1,2, ..., A **do**
    $v =$ calculate_all_bit_perms(N_a, R)
    if h_{N_a} undefined **then** ▷ create hash-table if it doesn't exist
        for i=1,2, ..., length(v) **do**
            $h_{N_a}[i] = v[i]$
        end for
    end if
    $subsize(a) = \binom{N_a+R}{N_a}$
    $stepsize(a) = \prod_{b=a+1}^{A} \binom{N_b+R}{N_b}$
    $i_a = i//stepsize(a))\%subsize(a)$
    $[S_a, I_{a1}, \ldots, I_{aR}] = f(h_{N_a}[i_a])$
end for

5.5 Modelling of worm prevalence

We begin by modelling the prevalence of Ascaris in the four villages in Nigeria using the simplest specification of our model with $A = 1$, $R = 1$.

5.5.1 Demonstrating validity of independence approximation

As previously mentioned, the fact that households may be treated independently at equilibrium is only exact in the limit of infinite households ($m \to \infty$); therefore we must justify the assumption that this approximately holds for the much smaller populations (several hundred people per village) in our dataset.

Stochastic simulations of the long-term dynamics of the full time-dependent external infection model using the Gillespie algorithm (algorithm 1) can be performed for various model parameters and compared to the calculated stationary distributions of households assuming $\epsilon = \delta\hat{J}^* \approx \delta J^*$. As defined in section 5.2.1, δ is the external infectious pressure on a susceptible individual in a fully infected population, J^* is the prevalence at equilibrium in the limit of infinite households, and $\hat{J}^*$ is the average prevalence resulting from the stochastic simulations.

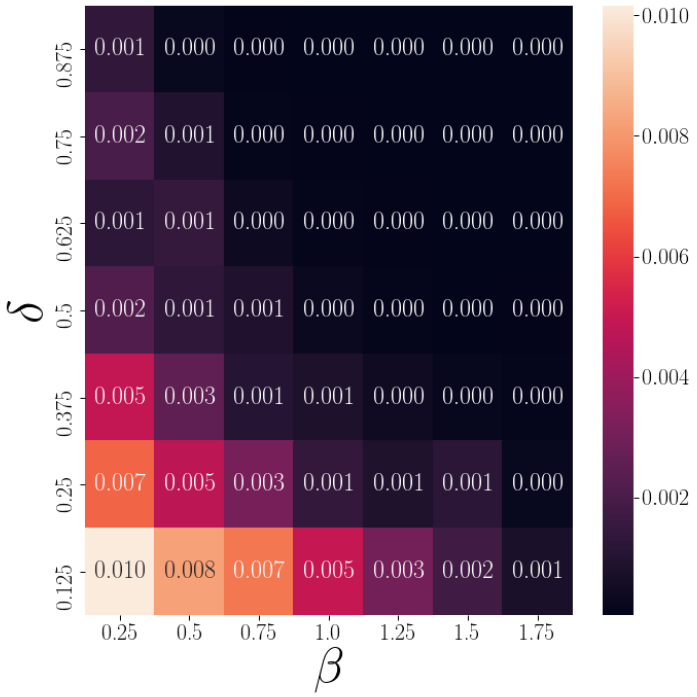

Figure 5.2: Mean squared error of the stationary distribution calculated using the independence approximation when compared to the fully simulated dynamics for varying within-household transmission parameter, β, and between-household transmission parameter, δ, with $m = 50$ households and $\alpha = 1$.

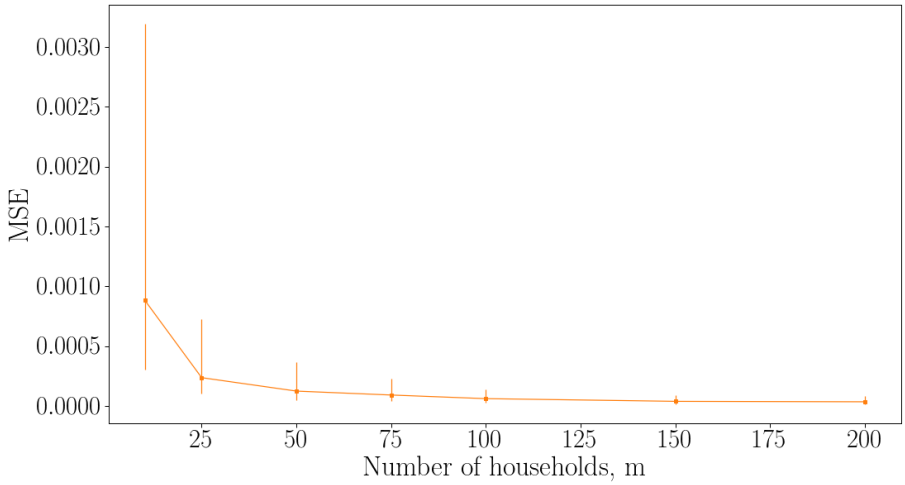

Figure 5.3: Mean squared error of the stationary distribution calculated using the independence approximation when compared to the fully simulated dynamics for a varying number of households m and fixed transmission parameters $\beta = 2, \delta = 0.3, \alpha = 1$.

Figures 5.2-5.4 show the mean squared error between the stationary distribution of the exactly simulated dynamics (averaged over 25 simulations lasting for $t = 150\gamma$) and our independence approximation. We observe that the error is small even for a modest number of households (~ 25). The errors are larger the smaller β or ϵ are because this increases the probability of stochastic extinction of all households simultaneously which cannot happen with the fixed external force of infection our model assumes; however from our data we know we are in an invasion regime so this breakdown does not concern us.

5.5.2 Results

Now we have demonstrated that the independence approximation holds well for our dataset we fit the model independently to each village using the adaptive Metropolis Hastings algorithm detailed in section 2.3.3.3. Figures A.1-A.4 and table 5.1 give traceplots and convergence diagnostics demonstrating the structure of the posterior and that the Markov Chain is well mixed, and figure 5.5 compares parameters across the four villages.

There is large uncertainty in the model parameters and a large discrepancy between villages which indicates that either villages are very different or we are not

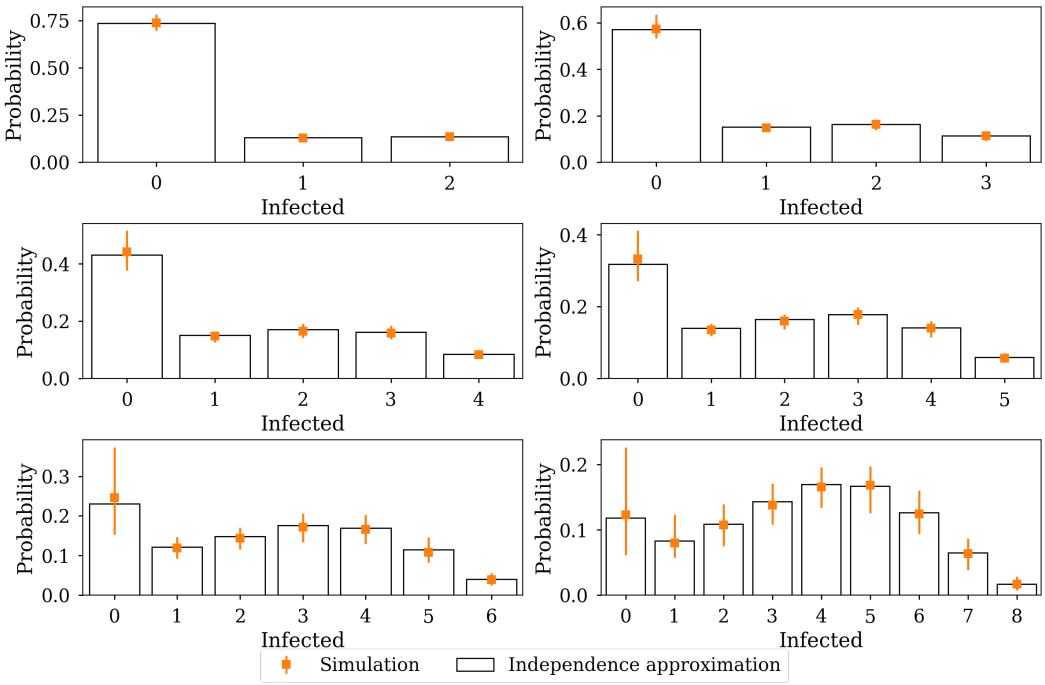

Figure 5.4: Comparison of the fully simulated (time averaged) dynamics versus the household independence approximation in terms of the probability distribution of the number infected in a household of a given size. The comparison was performed for Village 2 with transmission parameters $\beta = 2, \delta = 0.3, \alpha = 1$ ($m = 57, N = 180$).

capturing some important process. The work of the previous chapter tells us that the villages do not appear incredibly different but that the difference between adults and children (which this model does not capture) is incredibly important.

Figure 5.6 shows the posterior predictive distribution for the smallest six household sizes in Village 2. The posterior predictive check shows a good match between the models and the data. As we scale up A and R in our generic model specification then this visual posterior predictive checking becomes infeasible because the number of age and risk levels will split the households into further demographic categories, which would result in more subplots and thus a lot of household types with only one data point (like the last subplot).

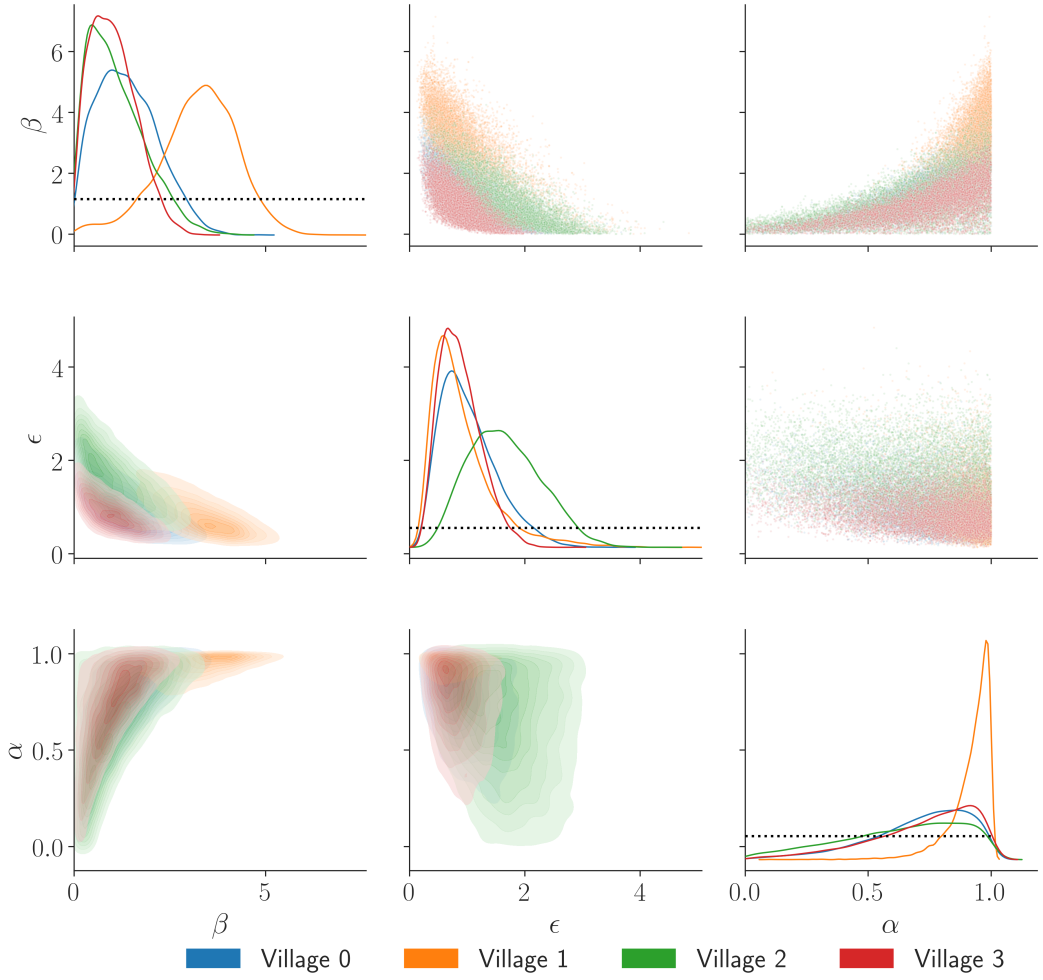

Figure 5.5: Posterior distributions of the $[SI_1S]_1$ household model compared across four villages, where β, ϵ, α refer to the within-household transmission, between-household transmission, and frequency dependence parameters respectively. The diagonals give kernel density estimates of the posterior distributions, and the off-diagonals give the pairwise distributions of parameters both as a scatter-plot (above diagonal) and bi-variate kernel density estimate (below diagonal). Wide uniform priors are plotted as dashed black lines.

5.6 Modelling of worm prevalence with age and risk structure

The heterogeneity between the posteriors of villages in the last section which was not observed by our empirical models in chapter 4 suggested the need to incorporate more structured dynamics into our models. The principal candidate for this based

	Village	Mean	s.d.	MC error	HPD 2.5	HPD 97.5	n_{eff}	$\widehat{R}$
α	0	0.71	0.21	0.00	0.30	1.00	6075.95	1.0
	1	0.92	0.09	0.00	0.75	1.00	4395.18	1.0
	2	0.63	0.25	0.00	0.15	1.00	8124.09	1.0
	3	0.72	0.21	0.00	0.28	1.00	7298.10	1.0
β	0	1.40	0.79	0.01	0.04	2.85	5550.94	1.0
	1	3.24	1.02	0.02	1.21	5.30	4279.77	1.0
	2	1.11	0.75	0.01	0.00	2.54	6507.49	1.0
	3	1.02	0.60	0.01	0.00	2.10	6501.33	1.0
ϵ	0	1.03	0.49	0.01	0.23	1.99	5544.38	1.0
	1	0.92	0.56	0.01	0.18	2.06	3767.85	1.0
	2	1.65	0.62	0.01	0.57	2.87	6758.75	1.0
	3	0.87	0.36	0.00	0.26	1.59	6268.34	1.0

Table 5.1: Posterior summary statistics for the $[SI_1S]_1$ household model compared across four villages, where β, ϵ, α refer to the within-household transmission, between-household transmission, and frequency dependence parameters respectively. Mean and s.d. give the posterior mean and standard deviation; MC error gives the monte carlo error; HPD 2.5 & HPD 97.5 refer to highest posterior density intervals; n_{eff} gives the number of effective samples; and $\widehat{R}$ gives the Gelman-Rubin convergence diagnostic.

on our prior knowledge of STH dynamics, analysis of chapter 4, and the research questions we are trying to answer is the need for incorporating age and infection level dynamics. Therefore we now analyse a model with two age categories and two infection levels, $[SI_2S]_2$ ($A = 2$, $R = 2$), with the first age category being individuals younger than 16 years of age ('Children'), and the two infection levels corresponding to EPG counts that are below and above the population mean.

Figure 5.7 summarises the states and possible transitions for this model. In addition to this we assume recovery rates for both age classes (A and C) are the same with the transition from $I_{XL} \to S_X$ (for $X = \{C, A\}$) happening at rate γ and the transition $I_{XH} \to I_{XL}$ occurs at rate γ_H. We take $\gamma = 1$, as with the stochastic SIS household model, in order to set the time scale of the model so that it is identifiable. The transitions $S_X \to I_{XL}, I_{XL} \to I_{XH}$ occur at rate $\Lambda\rho_X$ where $\rho_A = 1$ and ρ_C is to be inferred.

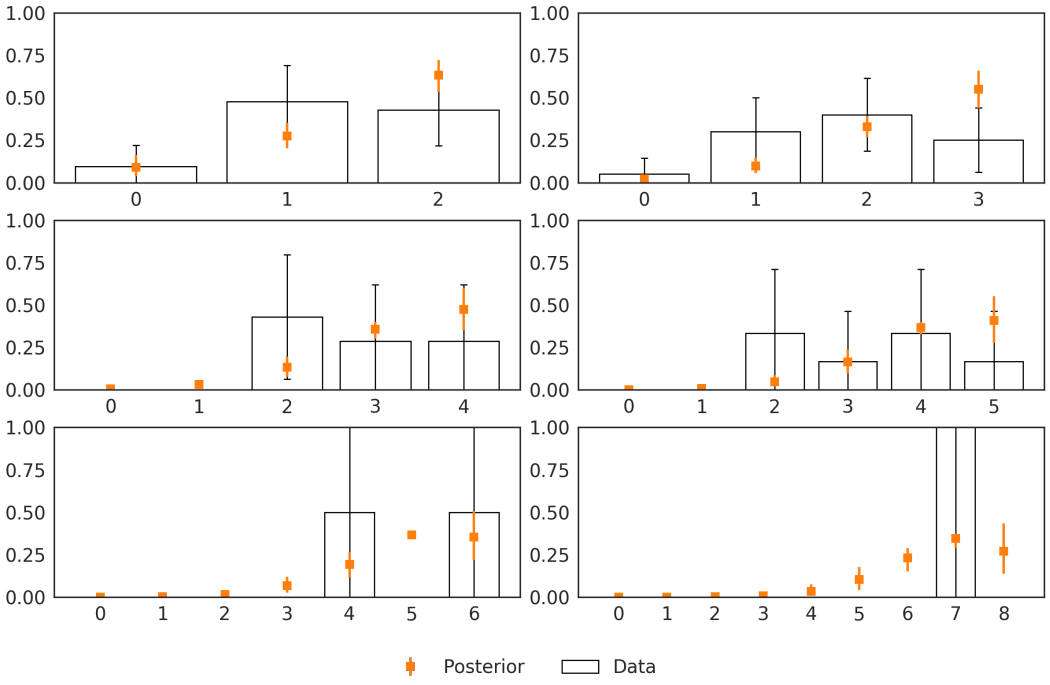

Figure 5.6: Posterior predictive plot of the probability distribution of the number infected in a household of a given size for village 2 under the $[SI_1S]_1$ household model. Uncertainty in the posterior is the 95% BCI, uncertainty in the data is given by Jeffrey intervals [37]. The small number of households of each size limit the use of this check.

5.6.1 Results

We perform MCMC on this model in an identical manner to the previous subsection. Standard log-normal priors were chosen for ρ_C and γ_H to give equal weighting to these parameters being higher or lower than $\rho_A = 1$ and $\gamma = 1$ respectively. A Beta($\alpha = 10, \beta = 1$) prior was put on α to reflect our prior that frequency dependent transmission is more likely (figure 4.8). Wide half-normal priors with a standard deviation of 5 were placed on the rest of the parameters to reflect our relative ignorance of these parameters.

The posterior distributions are given in figure 5.8; traceplots for each village are plotted in figures A.5-A.8; and posterior statistics and convergence diagnostics are summarised in table A.1. The joint posterior distribution across villages is shown in figure A.9 allowing a comparison across villages. There are multiple discrepancies between the inferred parameters, the most notable being that villages 1 & 2 (green and orange) have much larger estimates of both ρ_C and γ_H.

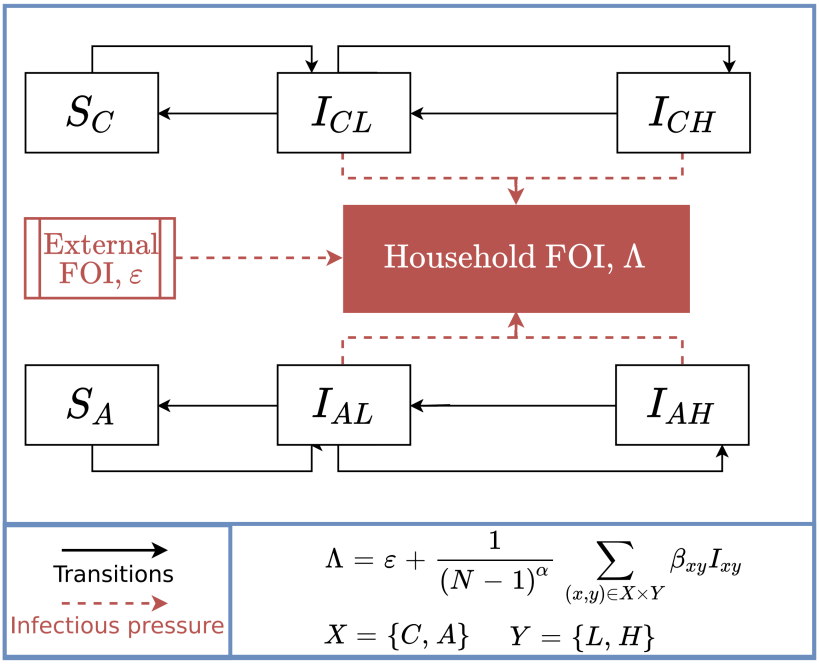

Figure 5.7: Model flow diagram for a $[SI_2S]_2$ household model. States with sub-scripts of A and C denote adult and child age classes and sub-scripts of L and H denote 'low' and 'high' levels of infection. Infected individuals in the household contribute to the household force of infection (FOI), Λ, along with an external FOI which comes from infected individuals external to the household.

By sampling from the posterior distribution one may calculate the ratio of within household to between household infections for each village (figure 5.9). The results show striking evidence of households driving the infections under this model - all villages have at least 79% of infections occurring from within the household.

5.6.2 Forward simulation

Having found evidence for strong within household transmission (at least under the assumptions of our model) we now proceed to answer the impact that these dynamics will have on different MDA strategies. By drawing samples from the posterior and running our dynamics forward in time for a large ensemble of Monte-Carlo simulations (using the Gillespie algorithm - algorithm 1) we may run *in silico* experiments of the distribution of elimination times under these different MDA strategies.

In these experiments we assume the efficacy of our simulated anthelmintic is 88% [104] wherever we draw conclusions from data (we use different efficacies for expository purposes) and vary both treatment interval (measured in units of γ as

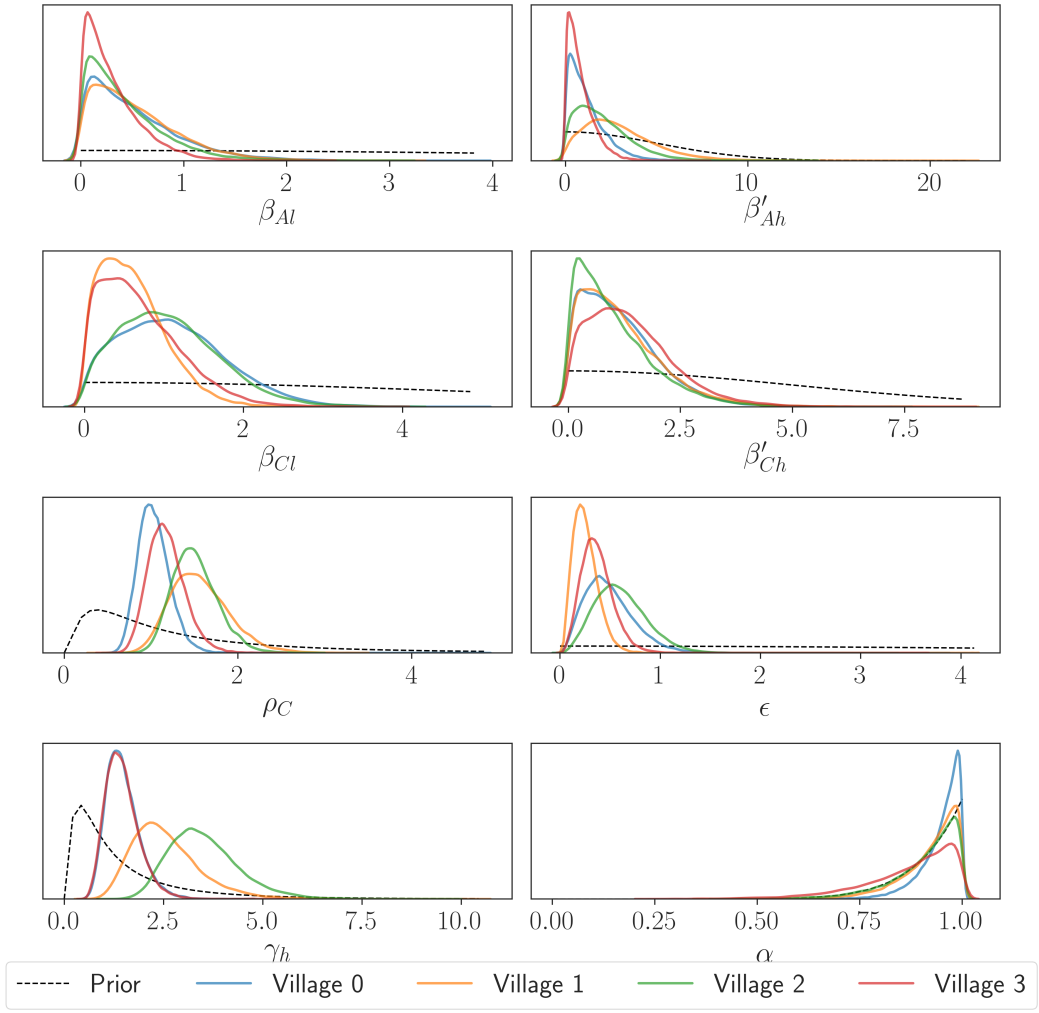

Figure 5.8: Posterior distribution of the $[SI_2S]_2$ household model compared across villages. β_{Al} (β_{Cl}) is transmission rate of an adult (child) at the lowest infection level; β'_{Ah} (β'_{Ch}) is the transmission rate of an adult (child) at the highest infection level minus β_{Al} (β_{Cl}); ρ_C is the relative susceptibility of the child class compared to the adult class; ϵ is the external force of infection; γ_h is the recovery rate of both age classes from the highest infection level to the lowest infection level; and α is the frequency dependence parameter. Priors are plotted as dashed black lines.

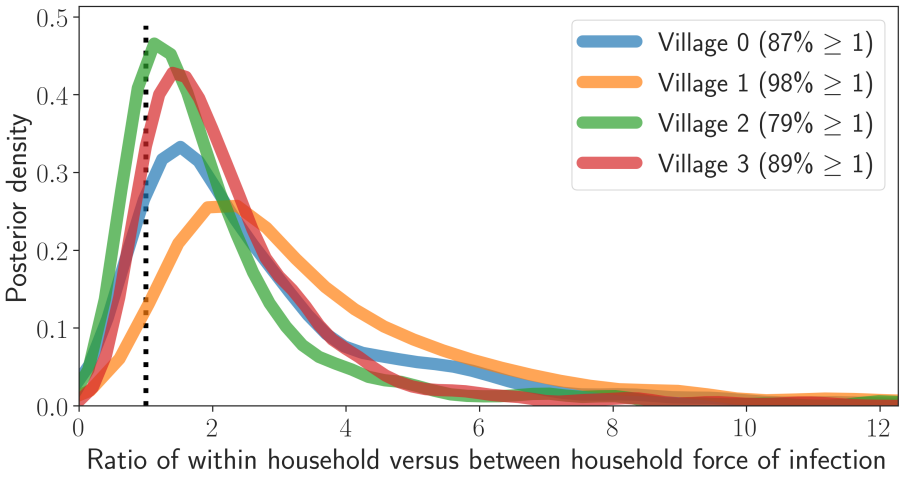

Figure 5.9: Posterior distribution of the ratio of within household infections to between household infections across villages for the chosen $[SI_2S]_2$ dynamics. The number in the brackets of the legend entries give the percentage of samples that are greater than one (correspond to more within household infections than between household infections).

this is the timescale of our system) and coverage levels. Six control strategies are considered: 'Random' control where the population is treated completely randomly (without replacement); 'Household' where households are treated in order of prevalence until coverage is reached; 'Targeted' where children are treated first, followed by random adults until coverage is reached; 'Targeted + Household' where children are treated first, followed by implementing the 'Household' strategy until coverage is reached; 'Household (baseline)' where the 'Household' strategy is implemented but the households that are treated are always based on the most prevalent households in the first MDA round of the simulation; 'Targeted + Household (baseline)' where the 'Targeted + Household' strategy is implemented but the households that are based on the first MDA round.

The motivation behind the baseline methods are that the regular household strategies rely on the assumption that we know which are the most prevalent households during every MDA round. This is not a far-fetched assumption for the 'Targeted + Household' strategy as MDA programmes are often based around de-worming through school programmes therefore when the 'Targeted' part of the treatment is being given to school children then the EPG count (or physical number of worms expelled) of school children could be used to gain knowledge of which households to target [76]. For the 'Household' strategy there is no way to estimate

the most prevalent households without a measurement process that would approximate the same amount of effort as obtaining 100% coverage (at which point all MDA strategies) are going to be equally effective!

Figures 5.10-5.11 show example epidemic dynamics (i.e. fictional parameters) for high efficacy, low coverage (Figure 5.10) and low efficacy, high coverage (Figure 5.11) scenarios for each MDA strategy. We observe in the low efficacy scenario, the effects of the control strategies become much more similar as the drop in efficacy brings all methods closer to the random strategy because any structure/dynamics being exploited have a higher chance of failure due to the poor efficacy. The high efficacy scenario results in the household based methods being much more effective as the highest burden patients are being treated with a very high success rate.

Figures 5.12-5.13 show the posterior mean time to extinction and the posterior probability of extinction for each MDA strategy over a variety of coverage and treatment interval assumptions. 'Household' strategy is the highest performer closely followed by 'Targeted + Household' with both consistently having the lowest mean time to extinction and the highest probabilities of extinction occurring within 250γ.

Considering the full 'posterior' distribution of extinction times can shed light on why the 'Baseline' versions of household models perform so poorly. Figure 5.14 shows this distribution for a coverage of 70% and treatment interval 4γ and shows that the 'Baseline' strategies have a reasonable probability of reaching extinction early but have very heavy tails as if they do not reach extinction within the first couple of rounds of treatment then the 'Baseline' households they are treating (the ones that started with the highest prevalence) are unlikely to still be the highest prevalence households as our model does not capture pre-disposition to infection. This results in later rounds of treatment being no more effective than the random strategy. These strategies demonstrate an absolute worst case scenario for household strategies because in reality households are more likely to remain high prevalence due to factors such as genetic pre-disposition.

5.7 Discussion

In this chapter we have developed a generic methodology for specifying, solving, and performing inference on stochastic household models with an arbitrary number of

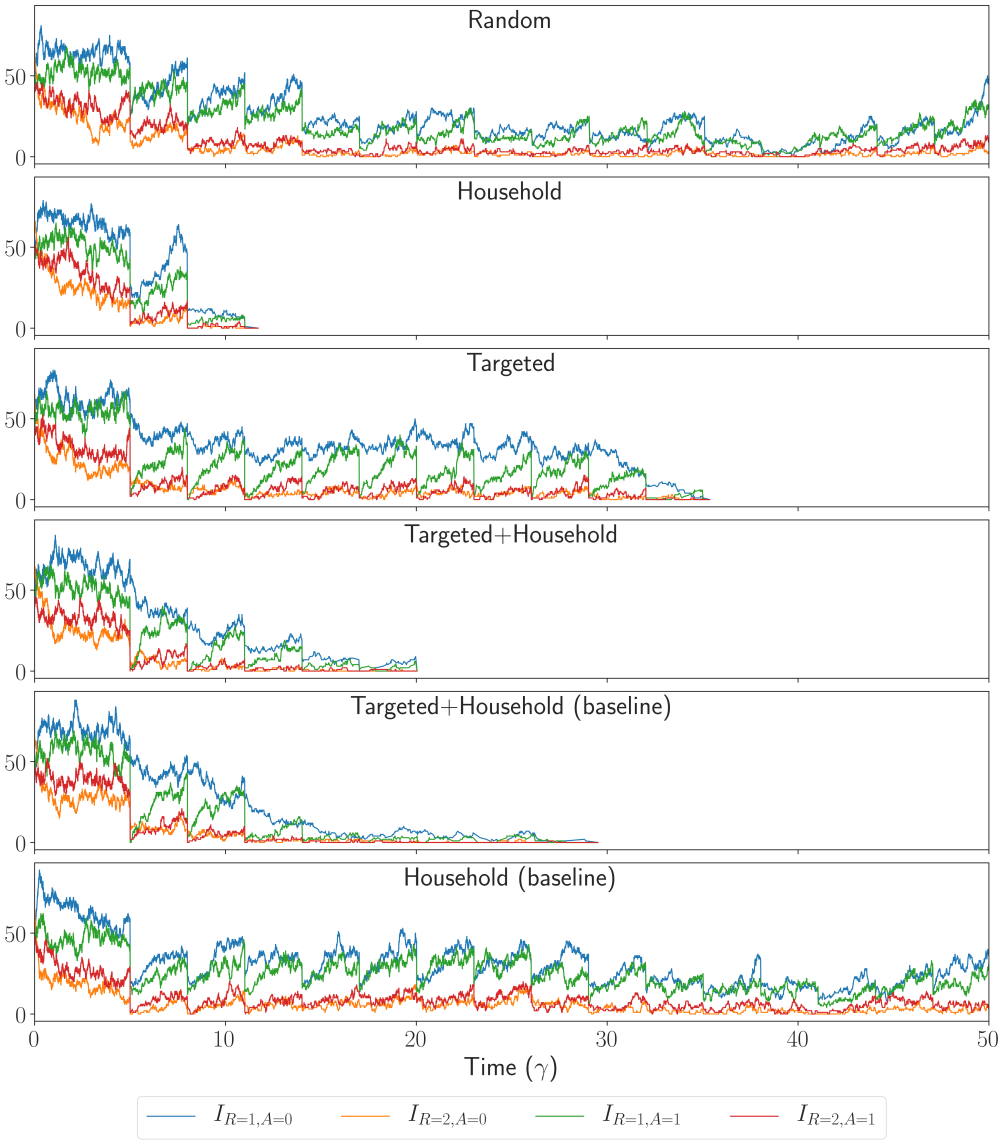

Figure 5.10: Example simulated $[SI_2S]_2$ dynamics of MDA strategies for an efficacy of 95%, population coverage of 60% and treatment interval of 3γ. 'Random' treats the population at random; 'Household' treats households in order of prevalence; 'Targeted' treats children at first, then at random; 'Targeted+Household' treats children at first, then treats households in order of prevalence; 'Targeted+Household (baseline)' treats children at first, then treats households in order of their prevalence when first measured; and 'Household (baseline)' treats households in order of their prevalence when first measured. The four lines give the number infected in the different age and risk level combinations.

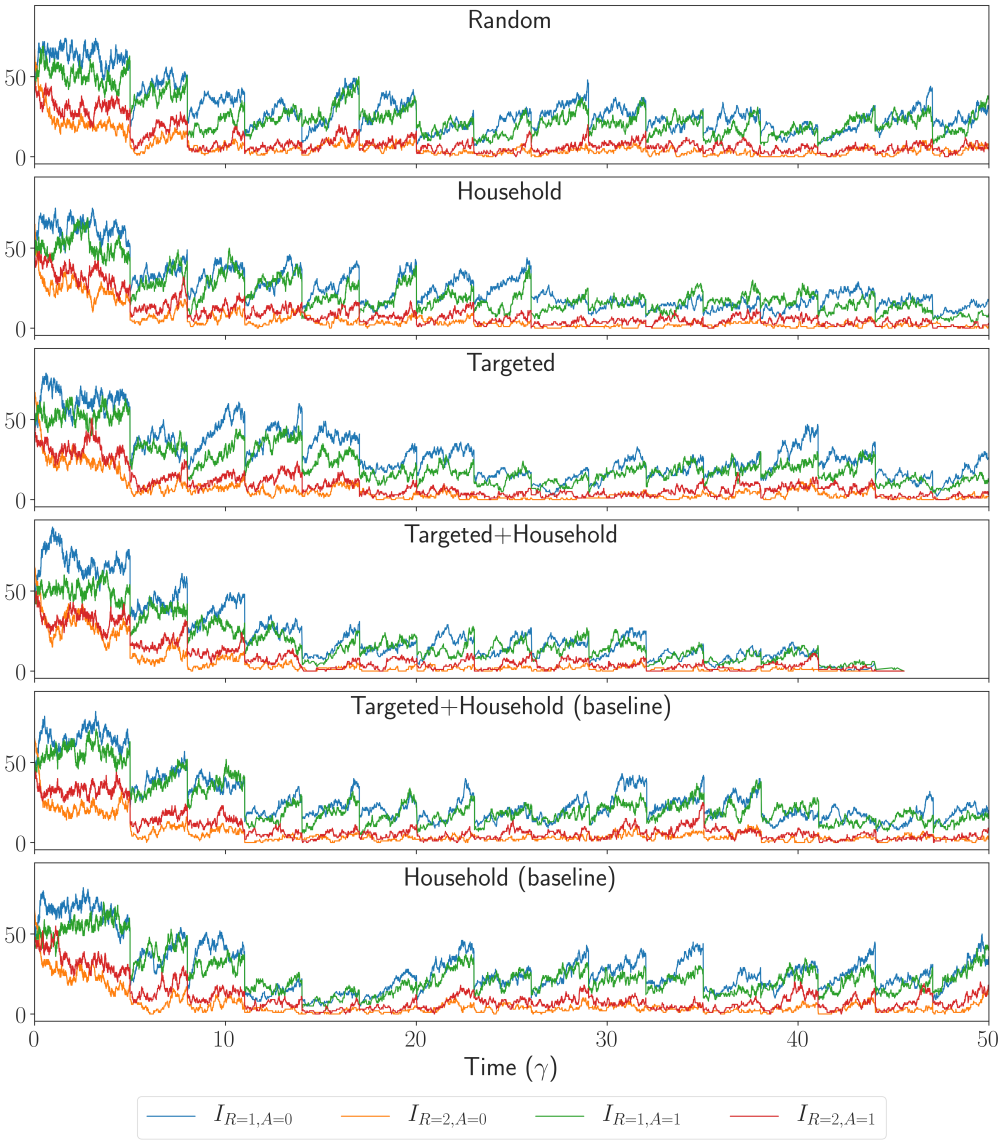

Figure 5.11: Example simulated $[SI_2S]_2$ dynamics of MDA strategies for an efficacy of 60%, population coverage of 95% and treatment interval of 3γ. 'Random' treats the population at random; 'Household' treats households in order of prevalence; 'Targeted' treats children at first, then at random; 'Targeted+Household' treats children at first, then treats households in order of prevalence; 'Targeted+Household (baseline)' treats children at first, then treats households in order of their prevalence when first measured; and 'Household (baseline)' treats households in order of their prevalence when first measured. The four lines give the number infected in the different age and risk level combinations.

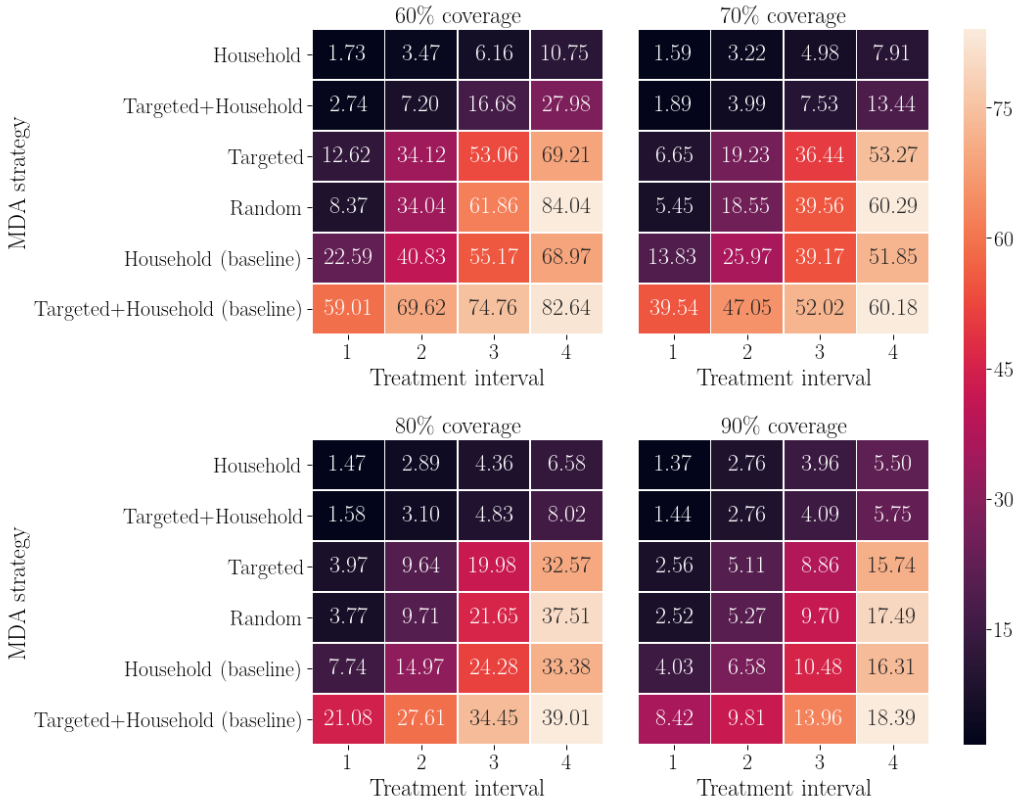

Figure 5.12: Posterior mean time to extinction for $[SI_2S]_2$ dynamics under different MDA strategies (rows), treatment intervals (columns), and treatment coverage levels (facets). 'Random' treats the population at random; 'Household' treats households in order of prevalence; 'Targeted' treats children at first, then at random; 'Targeted+Household' treats children at first, then treats households in order of prevalence; 'Targeted+Household (baseline)' treats children at first, then treats households in order of their prevalence when first measured; and 'Household (baseline)' treats households in order of their prevalence when first measured.

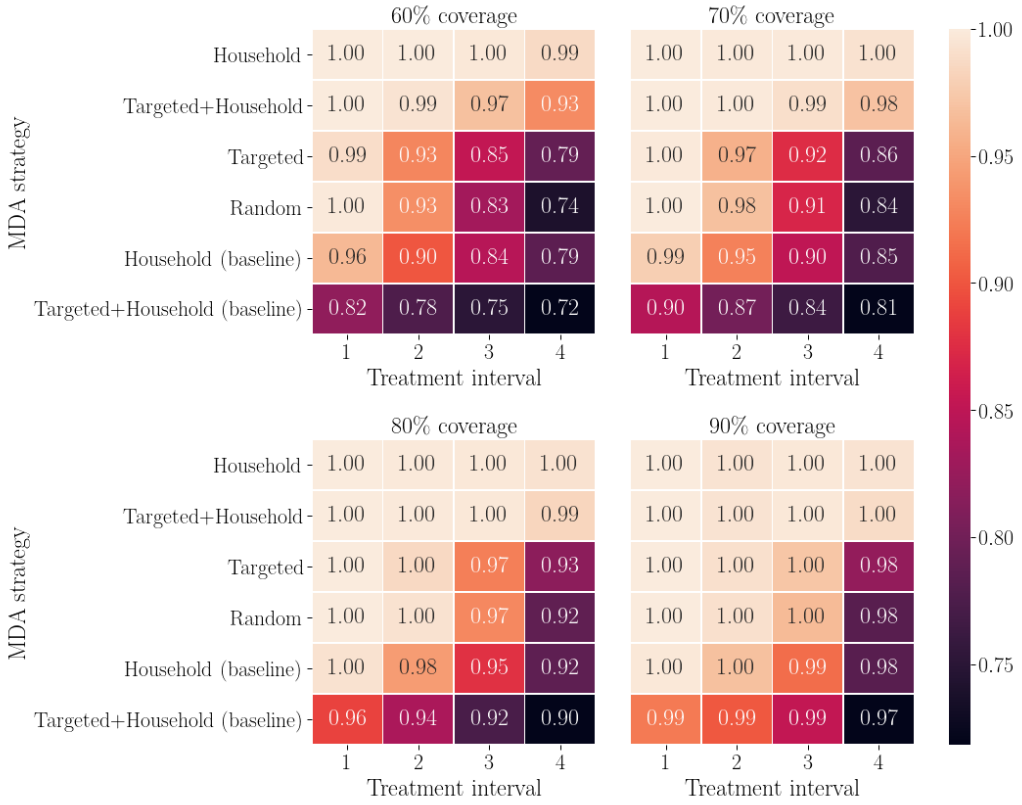

Figure 5.13: Posterior probability of extinction for $[SI_2S]_2$ dynamics under different MDA strategies (rows), treatment intervals (columns), and treatment coverage levels (facets). 'Random' treats the population at random; 'Household' treats households in order of prevalence; 'Targeted' treats children at first, then at random; 'Targeted+Household' treats children at first, then treats households in order of prevalence; 'Targeted+Household (baseline)' treats children at first, then treats households in order of their prevalence when first measured; and 'Household (baseline)' treats households in order of their prevalence when first measured.

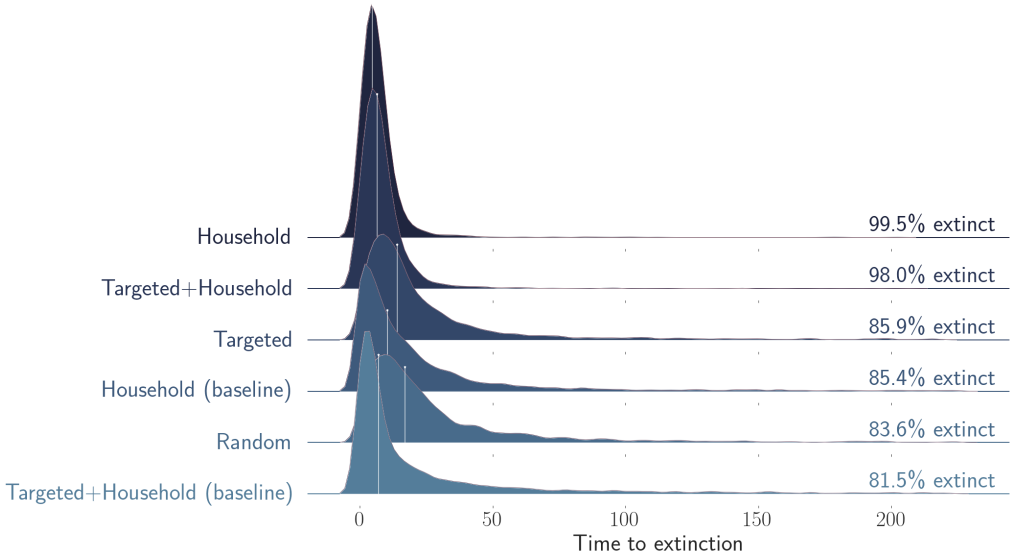

Figure 5.14: Posterior distribution of the time to extinction for each control method administered at 70% coverage and a treatment interval of 4γ. The percentage of simulations going extinct before simulation is stopped at $t = 250\gamma$ is also given. 'Random' treats the population at random; 'Household' treats households in order of prevalence; 'Targeted' treats children at first, then at random; 'Targeted+Household' treats children at first, then treats households in order of prevalence; 'Targeted+Household (baseline)' treats children at first, then treats households in order of their prevalence when first measured; and 'Household (baseline)' treats households in order of their prevalence when first measured.

infection levels and independent demographic classes. This has been simultaneously applied to our questions from the previous chapter on the role of households in the transmission and control of Ascaris for which we have estimated the ratio of within household to between household infections and simulated various control strategies to assess their effectiveness. The astute reader may question whether it is necessary to employ a dynamic model only to make an assumption that we are at equilibrium, but this approach gave us the ability to forward simulate our inferred dynamics to study the effectiveness of different control strategies - something we could not do with a GLMM.

We briefly reflect on this methodology as it relates to Ascaris, any policy implications of this analysis, and finally challenges with scalability and how scalability of this methodology may be improved.

5.7.1 Model development

Given the large household sizes in the dataset, our methodology can currently only scale to $[SI_2S]_2$ before inference starts to become computationally infeasible. This level of complexity is more than adequate for certain diseases such as Yaws for which the methods developed in this chapter have been used by Dyson et al. [52]; however for Ascaris several more infection levels and at least one more age class (to discriminate between school-aged and pre-school-aged children) would be necessary to begin to fully capture the intricacies of aggregation seen in the dataset (as evidenced by our use of Negative Binomial regression in the previous chapter) as well as other biological processes such as latent stages of infection. One possibility for reaching this level of complexity would be forgoing recovery in order to exploit the methodology Kinyanjui et al. [107] applied to Scabies by using the degree of advancement [29] model which permits a representation of the model states such that the transition matrix of the Markov process is upper triangular leading to massively decreased computational costs. Removing recovery from the model would involve throwing away even more biological realism - though it could be argued that gains in A or R could be more important - but this would likely soon fail to scale as for $A, R \geq 3$ the time taken to construct the transition matrices begins to exceed the time taken to find its null eigenvector (see section 6.2). Whilst this methodology has only been applied to a single time-point in this chapter the methodology generalises beyond one time-point by solving the full dynamical system, e.g. using the standard Runge-Kutta method for solving ODEs [174]. Matrix exponential (section 2.2.2) solutions of $\boldsymbol{Q}$ cannot be applied as if we are not at equilibrium then ϵ will be time-dependent as our independence approximation will not apply.

5.7.2 Implications for policy

The analysis in this chapter found a large amount of evidence suggesting that the household unit is a driver of the spread of Ascaris, with simulations of various MDA control strategies suggesting that those that target high prevalence households are likely to achieve elimination faster. Whilst this was a useful modelling exercise and guided our development of the methods, as mentioned above the final model does not sufficiently capture all the key dynamics and thus should not be used for making policy recommendations. An interesting observation is the tendency to hold dynamic models to a higher standard than their empirical counterparts. For example, logistic regression makes what seems like hopelessly simplistic assumptions yet remains remarkably effective. Perhaps it's easier and more meaningful to critique a dynamic

model as they are generally used as an attempt to directly model reality as closely as possible. Furthermore, beyond scaling the complexity of the household approaches it would be important to consider the effects of systematic non-compliance in our simulation of MDA strategies in order to fully assess their effectiveness [53, 55]. Whilst deterministic models cannot effectively capture the stochastic dynamics inherent in small populations, particularly those close to extinction, they can capture so many other variables and therefore relaxing the stochasticity and developing a deterministic household model for Ascaris may present an alternative opportunity for modelling.

5.7.3 Scalability challenges

The nature of our problem and dataset means that in order to perform inference we have to calculate the stationary distribution of our model many times. With each evaluation of the stationary distribution being computationally expensive we have two main options to increase the scalability of this approach: reduce the number of evaluations needed by improving the inference algorithm, or reduce the time taken to evaluate the stationary distribution each time. In the next chapter we consider various ways of improving the scalability of our methodology by trying to make improvements in both of these areas.

CHAPTER 6

Improved inference

6.1 Introduction

In the last chapter we formulated a methodology for performing inference on a stochastic household model with SIS-type dynamics ($[SI_RS]_A$) for an arbitrary number of demographic classes, A, and infection levels, R, and applied this to an analysis of the extent to which transmission of Ascaris is driven by households. Unfortunately our methodology failed to scale inference to a model specification complex enough to solve our specific problem. In this chapter we make attempts to increase the complexity of household models that are amenable to exact likelihood inference. Firstly, we examine numerous algorithms for solving large sparse linear systems and identify a choice with better scaling behaviour as our stochastic transition matrices become larger. Secondly, we investigate the use of MCMC algorithms which make use of the first-order gradients of the log posterior in order to improve the sampling efficiency. Next, we take a brief detour into variational methods to assess whether variational inference provides a feasible approach to inference for households models. Finally, we couple the inference procedure more tightly with the model solution by developing a variant of Metropolis-Hastings that rejects proposal moves before the full likelihood is calculated whilst maintaining all the usual asymptotic guarantees. We close the chapter with reflections on the household model inference methodology developed and outline approximate approaches which could be employed when a problem requires a model beyond the limit reasonably achievable with exact likelihood approaches.

6.2 Benchmarks of algorithms for solving large sparse linear systems

In section 2.2.3.2 we showed how calculation of the stationary distribution could be accelerated by calculating the null vector rather than using the Matrix exponential. In this section we briefly explore the scalability of sparse eigen-solvers compared to other methods of solving sparse linear systems.

The algorithm used to calculate the null vector in the last chapter was the implicitly restarted Arnoldi method implemented in ARPACK [114]. As A and R (the number of age and infection levels) have been increased it has become clear that the scaling with the size of the state-space (size of $\boldsymbol{Q}$) deteriorates - the 'eigs' lines in figure 6.1 show that while the computational complexity remains polynomial ($O(n^k)$), there is an increase in the order of polynomial growth, k, at matrix sizes above $n \approx 500$. We therefore appeal to several different methods for solving sparse linear systems ($\boldsymbol{Qx} = \boldsymbol{0}$). We have found that the biconjugate gradient stabilised method (BiCGSTAB) [193] gives vastly better performance than ARPACK as well as other sparse solvers we considered such as LGMRES [15]. BiCGSTAB is an iterative Krylov subspace method designed for numerical solution of non-symmetric linear systems.

Figure 6.1 shows for $[SI_3S]_2$ dynamics (notation introduced in section 5.3) the time taken to: generate $\boldsymbol{Q}$ by calculating transition rates and populating the matrix with them ('Matrix generation'); find the null eigenvector ('eigs') using the implicitly restarted Arnoldi method; solve the linear system with BiCGSTAB ('bicgstab'). The matrices are generated from the household size distribution of the first village in the dataset and samples were taken from MCMC draws with benchmarks done on the draws in both a randomly shuffled ('random') and sequential manner ('samples'). Both algorithms are iterative methods allowing the user to specify an initial guess to the solution which if close to the final solution theoretically reduces the number of iterations required to converge to the solution. This fits well with random walk MCMC methods where we are often proposing moves nearby in parameter space therefore the probability vector of the system state will often be similar from one MCMC iteration to the next (this applies even more so for gradient calculations using finite differences where step sizes in parameter-space are smaller). The difference between the 'samples' and 'random' schemes in the benchmarks illustrate the gains in efficiency iterative methods may achieve for random walk MCMC methods; however the speed-up from using the solution of the last MCMC draw ('samples' rather than

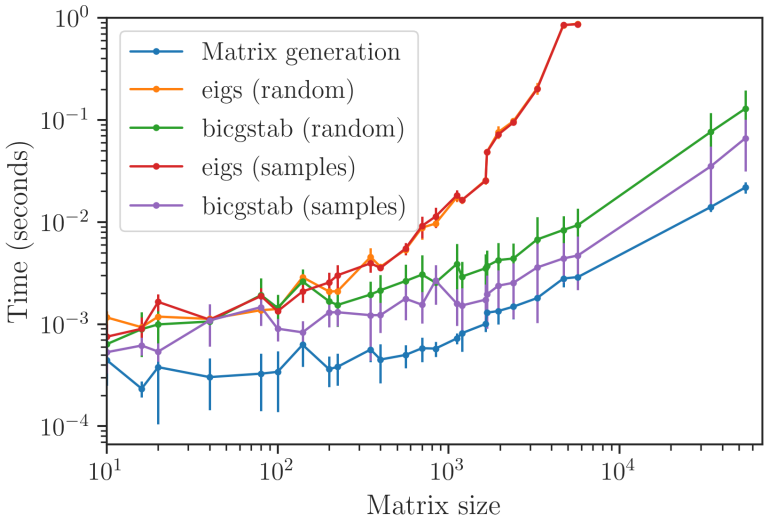

Figure 6.1: Benchmarks for the generation and solution of $\boldsymbol{Q}$ for $[SI_3S]_2$ dynamics. Solution of $\boldsymbol{Q}$ is performed with the implicitly restarted Arnoldi method ('eigs') and biconjugate gradient stabilised method ('BiCGSTAB'). Matrix sizes are determined by the household size distribution of village 1 and the benchmark parameters that parameterise $\boldsymbol{Q}$ are those of MCMC samples performed both in sequential order ('samples') and randomly shuffled ('random').

'random' scheme) as an initial guess only materialises for BiCGSTAB and gives a significant speed-up - 'bicgstab (samples)' vs. 'bicgstab (random)'. We observe that eigs and BiCGSTAB perform similarly for matrices up to $\approx 500 \times 500$; however above this BiCGSTAB is orders of magnitudes faster.

Figure 6.2 shows matrix generation and BiCGSTAB benchmarks for $[SI_3S]_3$ dynamics. Here, an uncomfortable limit is reached where the matrix generation time takes longer than its solution - for the largest matrix sizes (which dominate computation time) the time taken to construct the transition matrix ('Matrix generation') is greater than the fastest way we can calculate the stationary distribution ('bicgstab (samples)').

6.3 Gradient-based MCMC

Gradient-based MCMC methods are a class of MCMC algorithms which exploit information about the partial derivatives of the log posterior with respect to the model parameters in a way that facilitates sampling from the posterior more efficiently. For example, first order gradient information can be used to guide random walk propos-

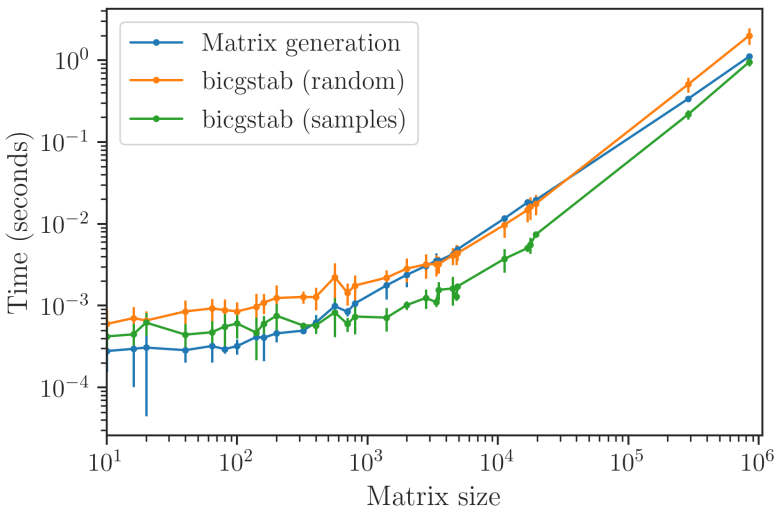

Figure 6.2: Benchmarks for the generation and solution of $\boldsymbol{Q}$ for $[SI_3S]_3$ dynamics. Solution of $\boldsymbol{Q}$ is performed with the implicitly restarted Arnoldi method ('eigs') and biconjugate gradient stabilised method ('BiCGSTAB'). Matrix sizes are determined by the household size distribution of village 1 and the benchmark parameters that parameterise $\boldsymbol{Q}$ are those of MCMC samples performed both in sequential order ('samples') and randomly shuffled ('random').

als towards regions of high posterior probability [169], and the Hessian (or other metric tensors) can be used to scale proposals to account for local curvature of the log-posterior [71]. These techniques become increasingly effective as the dimensionality of parameter space increases as they do not suffer a breakdown in efficiency to the extent of random walk methods (such as Metropolis Hastings).

For our household models we note that as we calculate our likelihood computationally we must therefore do the same for gradient information. There is no computationally efficient technique for calculating the derivative of the solution vector to a Markov chain (or its stationary distribution) with respect to some parameter set $\boldsymbol{\theta} \in \mathbb{R}^d$ that parameterise the transition matrix $\boldsymbol{Q}(\boldsymbol{\theta})$. We therefore calculate each partial derivative using finite difference methods [145]. For this d-dimensional parameter space, calculating the gradient with finite differences takes d additional evaluations of the likelihood function. It is our hope that this additional computational cost that is linear in d is offset by the curse of dimensionality where we see an exponential increase in the volume of parameter space with d, devastating the efficiency of random walk approaches to sampling.

6.3.1 Metropolis adjusted Langevin algorithm (MALA)

The first such gradient based MCMC we discuss and implement is the Metropolis adjusted Langevin algorithm (MALA) [169] which uses a discrete time approximation (6.3) to a Langevin diffusion (6.1) to construct a problem-specific proposal distribution using first order gradient information.

$$\dot{X} = \frac{1}{2}\nabla \log(\pi(X)) + \dot{W} \tag{6.1}$$

$$\lim_{t\to\infty} \mathbb{P}(X(t)) = \pi(X) \tag{6.2}$$

π denotes the probability density function (posterior) to be sampled. The Langevin diffusion consists of a gradient diffusion term $\nabla \log(\pi(X))$, and a standard Brownian motion term $\dot{W}$. In the limit as $t \to \infty$ then the distribution of X approaches the invariant stationary distribution π (6.2). Discrete time approximations of the Langevin diffusion can be generated by the Euler-Maruyama [111] method with fixed time-step $\tau > 0$ (6.3) which can be expressed as a draw from a multivariate Gaussian (6.4).

$$X_{t+1} = X_t + \tau/2\nabla \log(\pi(X_t)) + \tau\xi_t \tag{6.3}$$

$$X_{t+1} \sim \mathcal{N}(X_t + \tau/2\nabla \log(\pi(X_t)), \tau I_d) \tag{6.4}$$

In addition to simulating this diffusion, MALA augments the algorithm with a Metropolis-Hastings acceptance step,

$$\alpha = \min\left(1, \frac{\pi(X')q(X_t|X')}{\pi(X_t)q(X'|X_t)}\right) \tag{6.5}$$

$$q(X'|X) \propto \exp\left(-\frac{1}{2\tau}\left\|X' - X - \tfrac{\tau}{2}\nabla \log \pi(x)\right\|_2^2\right) \tag{6.6}$$

Thus, MALA can be seen as a Metropolis-Hastings algorithm with a proposal drift proportional to the log probability density of the posterior distribution. This drift results in proposals into higher probability density regions which improves the acceptance rate of the MCMC algorithm, particularly in higher dimensions. Under certain conditions the performance of MALA can be far from optimal, for example if π has large scale heterogeneities (which is often the reason one resorts to more advanced MCMC techniques) then a very small time step τ is required to capture the

Langevin dynamics which destroys the efficient sampling properties we desire [170]. Fortunately, this can be solved by learning the covariance structure of π as with the adaptive Metropolis-Hastings algorithm and using this as a pre-conditioning matrix on the drift term.

In order to implement a practically useful adaptive version of MALA (algorithm 4) we follow Atchade [12] which extends the work of Haario et al. [77] by using a stochastic approximation algorithm that recursively and simultaneously tunes the covariance matrix of q, Λ, and the scale parameter of the proposal kernel, $\sigma = \sqrt{\tau}$ such that the covariance matrix is tuned towards that of the target distribution. In addition to this adaptive mechanism the drift function is bounded (6.7) following the T-MALA algorithm of Roberts et al. [169] which the authors show avoids the possibility of a degeneracy in the rate of convergence. This truncation gives more theoretical guarantees and additionally gives the nice property that the ordinary Metropolis-Hastings algorithm is the special case where $D = 0$.

$$D_{\text{MALA}}(x) = \frac{\delta}{\max(\delta, |\nabla \log \pi(x)|)} \nabla \log \pi(x) \tag{6.7}$$

The adaptive MCMC algorithm augments the existing Metropolis-Hastings process (X_n) with a process $(\mu_n, \Gamma_n, \sigma_n)$ that is the parameter tuning process. This requires constants $\epsilon_1, \epsilon_2, A_1$ s.t. $0 < \epsilon_1 < A_1 < \infty$ and $\epsilon_2 > 0$ in addition to projection functions p_1, p_2, p_3 (6.8-6.10) which bound the adaptation process such that Γ_n is always a semi-positive definite matrix with Frobenius norm no greater than A_1; $\mu \in \mathbb{R}^d$ has euclidean norm no greater than A_1; and $\sigma \in [\epsilon_1, A_1]$.

$$p_1(\sigma) = \begin{cases} \sigma & \text{if } \epsilon_1 \leq \sigma \leq A_1 \\ \epsilon_1 & \text{if } \sigma < \epsilon_1 \\ A_1 & \text{if } \sigma > A_1 \end{cases} \tag{6.8}$$

$$p_2(\Sigma) = \begin{cases} \Sigma & \text{if } \|\Sigma\|_F \leq A_1 \\ \frac{A_1}{\|\Sigma\|_F}\Sigma & \text{if } \|\Sigma\|_F > A_1 \end{cases} \tag{6.9}$$

$$p_3(\mu) = \begin{cases} \mu & \text{if } \|\mu\|_2 \leq A_1 \\ \frac{A_1}{\|\mu\|_F}\mu & \text{if } \|\mu\|_2 > A_1 \end{cases} \tag{6.10}$$

Let (γ_n) be a sequence of positive numbers s.t. $\gamma_n > 0, \sum \gamma_n = \infty, \gamma_n = O(n^{-\lambda}), 0.5 < \lambda \leq 1$ which will satisfy the criterion of diminishing adaptation, and let $\bar{\tau}$ be the optimal acceptance rate which has been shown to be 0.574 for MALA

[13].

Algorithm 4 ATMALA

$X_0 \sim q(x)$ ▷ Initialise
for t=1,2, ... **do**
  $X' \sim q(X_t|X_{t-1})$ ▷ Proposal
  $\alpha(X'|X_{t-1}) = \min\left(1, \frac{q(X_{t-1}|X')\pi(X')}{q(X'|X_{t-1})\pi(X_{t-1})}\right)$ ▷ Acceptance probability
  $u \sim \text{Uniform}(0,1)$
  if $u < \alpha$ **then**
    $X_t \leftarrow X'$ ▷ Accept
  else
    $X_t \leftarrow X_{t-1}$ ▷ Reject
  end if
  Set $\mu_{t+1} = p_3(\mu_t + \gamma_t(X_{t+1} - \mu_t))$
    $\Gamma_{t+1} = p_2\left(\Gamma_t + \gamma_t((X_{t+1} - \mu_t)(X_{t+1} - \mu)' - \Gamma_t)\right)$
    $\sigma_{t+1} = p_1\left(\sigma_t + \gamma_t(\alpha - \bar{\tau})\right)$
end for

μ_n not re-projected is just the empirical mean of the samples, likewise Γ_n is approximately the empirical covariances of the samples. α is a stochastic estimate of the acceptance rate in the algorithm (averaged over time it will give the acceptance rate) therefore σ_n is increased if $\alpha > \bar{\tau}$ and decreased otherwise tuning the acceptance rate towards $\bar{\tau}$. This additional adaptation process is an example of Robbins-Monro recursion (6.11) [9, 165] where a random variable $h(\theta)$ which approximates an unobservable function of interest $g(\theta)$ by satisfying the relationship $\mathbb{E}(h(\theta)|\theta) = g(\theta)$ is used to find the roots of equation $g(\theta) - \alpha = 0$. For example, in the context of our adaptation process we wish to find the value of σ (θ) which gives an acceptance rate of $\bar{\tau}$ (α) when our algorithm is run $(g(\cdot))$ so we use the acceptance probability $(h(\theta))$ to approximate it.

$$\theta_{t+1} \leftarrow \theta_t + \gamma_t(h(\theta_t) - \alpha) \tag{6.11}$$

This additional adaptation process requires the specification of multiple extra hyperparameters on top of those already required by Metropolis Hastings; however Atchade [12] claim it is relatively insensitive to the choice of $\delta, A_1, \epsilon_1, \epsilon_2$ as long as δ, A_1 are sufficiently large and ϵ_1, ϵ_2 sufficiently small.

6.3.2 Hamiltonian Monte Carlo

Duane et al. [51] united the seminal MCMC approach of Metropolis et al. [126] with the approach of Alder and Wainwright [4] who solved similar molecular dynamics

problems to Metropolis using Hamilton Dynamics[62]. This approach was termed "hybrid Monte Carlo" (Hamiltonian Monte Carlo).

Consider a particle (our MCMC sampler) moving in a d-dimensional space (our models state space) as a frictionless point-mass sliding over a surface of varying height (negative log probability density). Hamiltonian dynamics express the system state as a d-dimensional vector q giving the position (the parameters currently being 'sampled'), and a d-dimensional vector p giving the momentum. Potential energy $U(q)$ is proportional to the height of the surface at the current position (Model parameters). Kinetic energy $K(p)$ is equal to $\|p\|^2/(2m)$, where m is the mass of the point-mass. On a level part of the surface (i.e. locally a uniform distribution) the point-mass moves at constant velocity p/m. A "hill" in this space (corresponds to a less likely parameter region) causes the point mass to lose kinetic energy and gain potential energy as it rolls up it, until the point-mass loses all of its kinetic energy and slides back down.

More formally, these dynamics have a state space of $2d$ dimensions with the system being described by the Hamiltonian, $H(q,p)$:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{6.12}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \tag{6.13}$$

$$i = 1, \ldots, d.$$

$$H(q,p) = U(q) + K(p) \tag{6.14}$$

$$K(p) = \frac{1}{2} p^T M^{-1} p \tag{6.15}$$

$U(q)$ is defined to be negative the log probability distribution of q (i.e. the height of our landscape). M is a symmetric positive-definite 'mass' matrix. $K(p)$ corresponds to the negative log probability density of a Gaussian distribution with covariance matrix M and mean zero. This reduces our equations for the evolution of the Hamiltonian to:

$$\frac{dq_i}{dt} = (M^{-1}p)_i \tag{6.16}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i} \tag{6.17}$$

$$i = 1, \ldots, d.$$

Hamiltonian dynamics possess several properties necessary for constructing an MCMC algorithm such as reversibility (demonstrate the dynamics leave the distribution invariant), conservation of the Hamiltonian (conservation means that a Metropolis update gives a theoretical acceptance rate of 1), and symplecticness (avoids expensive/infeasible calculations of the determinant of the Jacobian when updating the acceptance rate). A more in-depth discussion of the theoretical properties of Hamiltonian dynamics and how these relate to MCMC can be found in [23, 36].

As with MALA, one must approximate Hamiltonian dynamics by discretising time. The Leapfrog method [59] is used as it is a symplectic integrator that conserves the Hamiltonian over time unlike more common methods such as Euler's method.

$$p_i(t+\epsilon/2) = p_i(t) - (\epsilon/2)\frac{\partial U}{\partial q_i}(q(t)) \tag{6.18}$$

$$q_i(t+\epsilon) = q_i(t) + \epsilon\frac{p_i(t+\epsilon/2)}{m_i} \tag{6.19}$$

$$p_i(t+\epsilon) = p_i(t+\epsilon/2) - (\epsilon/2)\frac{\partial U}{\partial q_i}(q(t+\epsilon)) \tag{6.20}$$

Statistical mechanics may be brought to task by using a canonical distribution [34] to relate our posterior distribution to the potential energy function. Given some energy function $E(x) = H(q,p)$ for joint state space $x = (q,p)$, then (6.21) gives the probability density function over states, where T temperature ($T = 1$ without loss of generality), and Z is the "partition function" which is the normalising constant.

$$\pi(x) = \frac{1}{Z}\exp\left(\frac{-E(x)}{T}\right) \tag{6.21}$$

$$E(x) = H(q,p) = -\log\pi(q,p) - \log Z \tag{6.22}$$

$$\pi(q,p) = \frac{1}{Z}\exp\left(\frac{-U(q)}{T}\right)\exp\left(\frac{-K(p)}{T}\right) \tag{6.23}$$

$$U(q) = -\log\left(\pi(q)L(q|D)\right) \tag{6.24}$$

With this relation between the potential energy function of the Hamiltonian and our posterior in hand, all that is left is to set the distribution of momentum as we wish by specifying $K(p)$ - typically a quadratic kinetic energy distribution is chosen leading to momentum being distributed as a zero-mean Multivariate Gaussian. Given this we can write down the Hamiltonian Monte Carlo algorithm (algorithm 5) and sample from a model posterior as long as we can evaluate the partial derivative of the log probability density (which for household models we do numerically as mentioned at

the beginning of this section).

Algorithm 5 HMC

Choose $L \in \mathbb{N}^+$, $M \in \mathbb{R}^{d \times d}$, $\epsilon \in \mathbb{R}$
$X_0 \sim \pi(q)$ ▷ Initialise
$p' \sim \text{Normal}(0, M)$ ▷ Sample momentum
$K_0 \leftarrow p'^T M^{-1} p/2$
for t=1,2, ... **do**
    $p' \sim \text{Normal}(0, M)$ ▷ Sample momentum
    $q' \leftarrow q_{t-1}$
    for l=1,..., L **do** ▷ Simulate Hamiltonian Dynamics
        $p' \leftarrow p' - \epsilon/2 \nabla U(q')$
        $q' \leftarrow q' + \epsilon p'/M$
        $p' \leftarrow p' - \epsilon/2 \nabla U(q')$
    end for
    $p' \leftarrow -p'$
    $K' \leftarrow p'^T M^{-1} p/2$
    $\alpha = \min(1, \exp(U(q_{t-1}) - U(q') + K_{t-1} - k'))$ ▷ Acceptance Probability
    $u \sim \text{Uniform}(0, 1)$
    if $u < \alpha$ **then**
        $X_t \leftarrow q'$ ▷ Accept
        $K_t \leftarrow K'$
    else
        $X_t \leftarrow X_{t-1}$ ▷ Reject
        $K_t \leftarrow K_{t-1}$
    end if
end for

HMC samples much more efficiently, particularly in high dimensions, as the trajectories are not random walks as with e.g. the Metropolis-Hastings method. Using gradient information to approximate Hamiltonian dynamics allows for a much more efficient exploration of the state space as proposals are indirectly sampled from the random sampling of the momentums. This avoids a disproportionate number of proposals being in areas of low density which becomes an increasingly large problem with random walk techniques in even modest dimensions. The end result is that HMC yields a much higher acceptance rate and a lower sample auto-correlation.

All this efficiency does not come for free however. Whilst the auto-correlation of HMC samples may be much lower, each sample requires multiple evaluations of the likelihood (potential energy) and its gradient. Furthermore, if ϵ is chosen too large then the symplecticness of the leapfrog step breaks down and the Hamiltonian will diverge. If ϵ is too small then a large number of leapfrog steps L will be required

in each iteration in order to explore the posterior. This choice of (ϵ, L) means that HMC can be incredibly hard to tune as exploration of the posterior may end up incredibly slow like a random walk if the trajectory length ϵL is too small, or that the leapfrog step may double back on itself if ϵL is too large.

This difficulty in tuning in addition to the relatively obscure formulation may explain why seeing HMC in the wild remained relatively rare for the first 20 years or so of its life. That was until the arrival of the No-U-Turn Sampler (NUTS) variant of HMC arrived on the scene [84] supplying a way of automatically tuning the number of steps L by estimating when the samplers trajectory starts to turn around and trying to avoid this. The implementation of NUTS and its subsequent extensions [22, 24, 70] are highly sophisticated and are not sensible to implement by hand as with previous samplers when two existing probabilistic programming languages, Stan and pyMC3 [38, 175], contain well tested and optimised implementations which have sky-rocketed the popularity of HMC methods. Stan however is written in C++ and models are specified in a domain specific language (DSL) much like its precursors JAGS [158], and BUGS [181]. The Stan DSL whilst flexible for many Statistical Bayesian models is not very permissive for dynamic models such as ours. On the other hand, pyMC3 models are specified directly in Python, therefore by writing a custom Theano operation for our model's likelihood and its gradient (Theano [186] is the optimising compiler for manipulating and evaluating mathematical expressions that pyMC3 is built on) then we can use NUTS for our models. The implementation of this custom Theano operation for household models may be found at `https://github.com/bishax/snepits`.

6.3.3 Results

Figures 6.3-6.5 assess the viability of pyMC3's NUTS and our adaptive MALA implementation compared to adaptive MH for several household models of varying complexity. We calculate the effective sample size per second (ESS/s) of each sampler for each model. To test both the efficiency of tuning and post-tuning regimes we measure the ESS/s for the whole sampling process ('Total') and for optimally ('Optimal') initialised samplers (the 'optimal' covariance that initialises MH and MALA is the empirical covariance from NUTS as this is the most reliable algorithm). MALA is not included in the 'Total' subplots as its adaptive properties are so poor.

NUTS consistently outperforms MH both pre-tuning and post-tuning. MALA samples with much greater efficiency when initialised 'optimally' for the simplest SIS

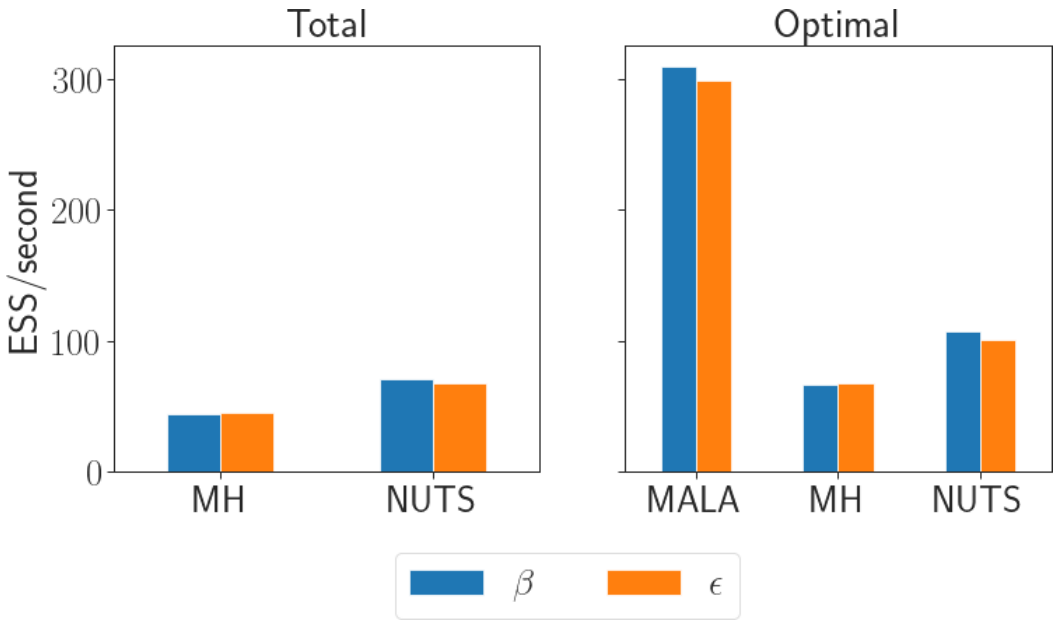

Figure 6.3: Effective Sample Size per second (ESS/s) for parameters of the household SIS model with frequency dependence parameter α fixed to unity. The left hand plot compares the ESS/s between Adaptive Metropolis-Hastings (MH) and NUTS samplers starting from randomly initialised initial conditions and no prior covariance specified. The right hand plot compares the ESS/s between Adaptive MALA, MH, and NUTS when the MALA and MH samplers are initialised with the covariance of the posterior inferred by NUTS, and the ESS/s of NUTS ignores the time taken to tune the hyper-parameters. For both plots, ESS/s are shown separately for each model parameter.

model but performs dramatically worse for all subsequent models as the chain tends to 'stick' on the same parameters for a large number of iterations - a known issue [172]. The MH implementation used is that of pyMC3 and not our own. Despite our own implementation being faster - it does not involve the overhead of fitting within a probabilistic programming language - we chose the pyMC3 implementation so that our results are closer to showing the theoretical differences between the algorithms rather than implementation specific details as this is the fundamental limit - we can always improve the performance of an implementation but we can not improve the performance of Mathematics. As pyMC3 does not implement a MALA sampler this remains our own (fast) implementation which gives MALA an edge. It is therefore all the more surprising that its performance is poor, and the fast implementation may be the sole reason that MALA outperforms NUTS and MH by a factor of more than 3 for the SIS model. For many MCMC algorithms effective sample size per iteration (ESS/iteration) is a better metric to measure their theoretical properties

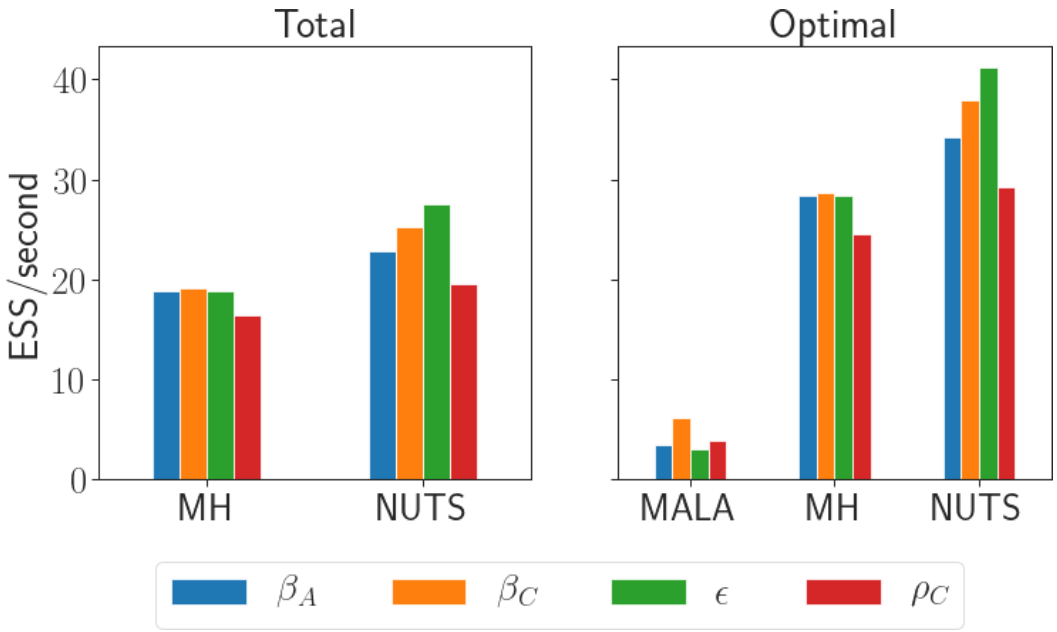

Figure 6.4: Effective Sample Size per second (ESS/s) for parameters of the $[SI_1S]_2$ household model with frequency dependence parameter α fixed to unity. The left hand plot compares the ESS/s between Adaptive Metropolis-Hastings (MH) and NUTS samplers starting from randomly initialised initial conditions and no prior covariance specified. The right hand plot compares the ESS/s between Adaptive MALA, MH, and NUTS when the MALA and MH samplers are initialised with the covariance of the posterior inferred by NUTS, and the ESS/s of NUTS ignores the time taken to tune the hyper-parameters. For both plots, ESS/s are shown separately for each model parameter.

(independent of the implementation details) as we aim to do here; however because one NUTS iteration involves a large and variable number of likelihood calculations, ESS/iteration is not a useful metric for NUTS.

6.4 Variational Inference

In recent years advances in Variational Inference (VI) [85, 106, 113] have sky-rocketed its popularity as a technique for performing scalable Bayesian inference. Rather than sampling from the posterior like MCMC, VI approximates the posterior by choosing a family of approximate densities, $\mathcal{Q}$, and finds the member of that family, $q^*(\boldsymbol{\theta})$, that minimises the Kullback-Leibler (KL) divergence to the posterior (6.25). The main benefit of VI methods are their speed, though this comes at the cost of biased inference - there are no asymptotic guarantees that one will capture the true posterior because VI aims to find a density 'close' to the target.

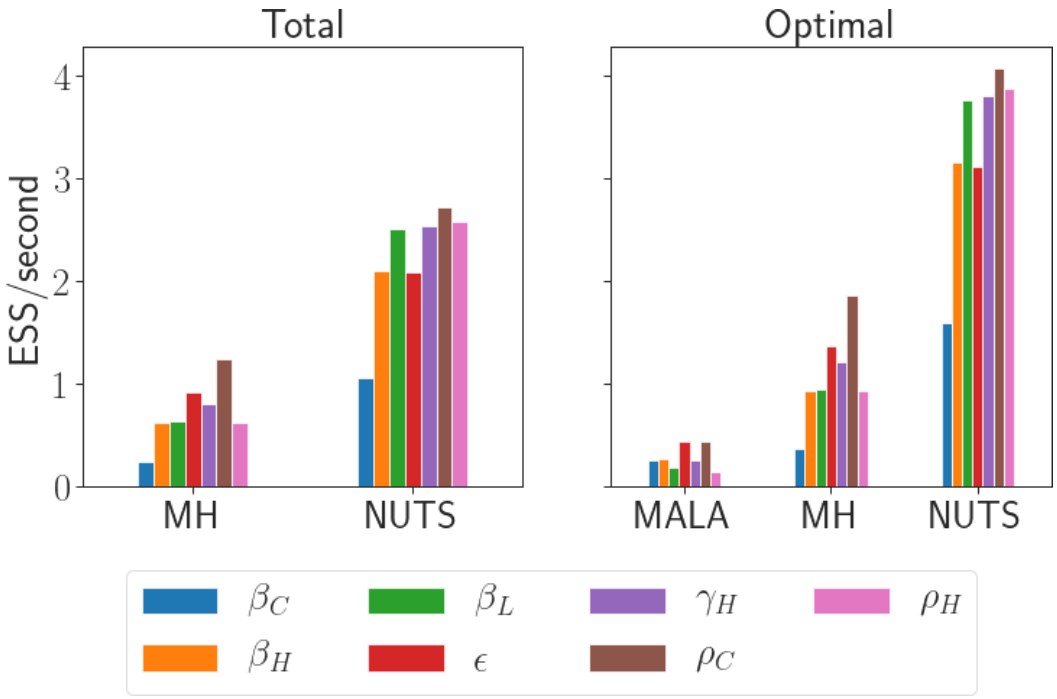

Figure 6.5: Effective Sample Size per second (ESS/s) for parameters of the $[SI_2S]_2$ household model with frequency dependence parameter α fixed to unity. The left hand plot compares the ESS/s between Adaptive Metropolis-Hastings (MH) and NUTS samplers starting from randomly initialised initial conditions and no prior covariance specified. The right hand plot compares the ESS/s between Adaptive MALA, MH, and NUTS when the MALA and MH samplers are initialised with the covariance of the posterior inferred by NUTS, and the ESS/s of NUTS ignores the time taken to tune the hyper-parameters. For both plots, ESS/s are shown separately for each model parameter.

$$q^*(\boldsymbol{\theta}) = \arg\min_{q(\boldsymbol{\theta})\in\mathcal{Q}} \mathrm{KL}(q(\boldsymbol{\theta})\|\|p(\boldsymbol{\theta}\|\mathcal{D})) \tag{6.25}$$

Minimising the KL divergence is equivalent to maximising the evidence lower bound (ELBO).

$$\mathrm{ELBO}(q) = \int_{\boldsymbol{\Theta}} \left(\log p(\boldsymbol{\theta}, \mathcal{D}) - \log q(\boldsymbol{\theta})\right) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{6.26}$$

Previously, deriving variational inference algorithms was a cumbersome process involving model-specific calculations but with the development of Automatic Differentiation Variational Inference (ADVI) [113] only a Bayesian model and dataset are required. Fortunately for us pyMC3 implements ADVI therefore we can use the custom Theano operator we wrote in the previous section to consider the feasibility

of ADVI for our household models. The VI approximation can be biased due to the inability of the approximation family to capture the true posterior; slow convergence of the optimisation problem; and KL-divergence under-penalising light tailed approximations. For this reason we do not dwell on the theory before first assessing the viability of VI methods.

We assess convergence of the variational approximation by monitoring the ELBO as well as the means and covariance of the approximations by running ADVI until the relative difference between the means and covariances now and 100 iterations earlier is less than 0.1% or until $500,000$ iterations is reached. Then samples from the posterior approximation are generated and compared to results obtained from MCMC using NUTS. More principled ways of assessing convergence using PSIS (section 2.3.4.2) have been recently proposed by Yao et al. [205]. Figures 6.6-6.7 demonstrate that the joint posterior obtained from both NUTS and ADVI are in reasonable agreement for $[SI_1S]_1$ and $[SI_1S]_2$ household models - whilst a quantitative comparison of speed is not possible as VI has no concept of ESS for these models it is 'fast' to obtain these approximations. When trying to fit ADVI to an $[SI_2S]_2$ model we found that when placing weakly informative but unbounded priors such as HalfNormal or HalfCauchy distributions then the variational approximation does not match that of the posterior with the posterior variance being several orders of magnitude too high! Furthermore, we observed that the ELBO experienced increasingly large fluctuations (figure A.11) with the means and covariances appearing to be stuck at fixed values. Choosing wide Uniform priors to bound parameter space resulted in all parameters except ϵ being close to delta functions around the upper bound of the prior with the ELBO traceplot displaying incredibly strange behaviour (figure A.10).

This short foray into variational methods confirms our prior belief that MCMC is more appropriate as it provides both theoretical guarantees in the form of asymptotic exactness and more practical methods of assessing convergence. As pointed out in the literature [31] variational inference is typically better applied to large datasets where inferences need not be exact (e.g. serving search results). Despite this exercise being of no practical use it was an interesting exercise particularly as little is known about the relative accuracy of VI and MCMC. Interestingly, VI generally underestimates the variance of the posterior density [31]; however we found that this was not the case for unbounded priors for our more complex models.

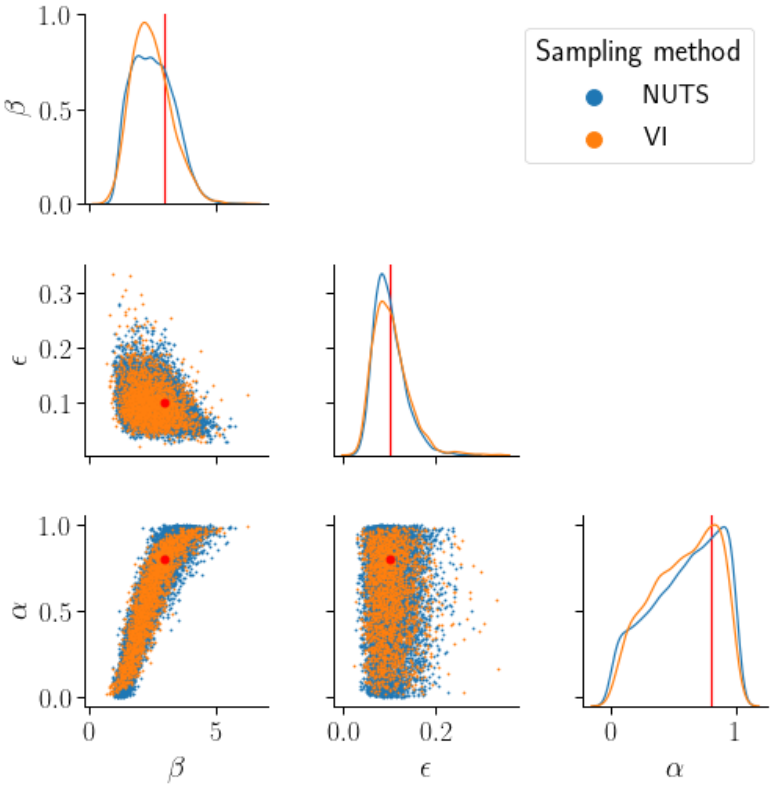

Figure 6.6: Joint posterior distributions for a household model with $[SI_1S]_1$ dynamics ($A = 1, R = 1$) using simulated data with true parameters plotted in red ($\beta = 3, \epsilon = 0.1, \alpha = 0.8$) inferred using both NUTS and ADVI.

6.5 Early rejection Metropolis Hastings

Thus far, we have been constantly plagued with problems of dimensionality - adding age and risk levels (of which we require many) increases the computational cost of solving household models exponentially; larger households (our dataset's are incredibly large) also increases the computational cost exponentially; and a larger model parameter space becomes exponentially larger in volume requiring more iterations of an MCMC algorithm to sample from. Beyond the independence approximation allowing us to treat households independently (section 5.2) we have yet to find and capitalise on any other feature of the data which we can turn to our modelling/inferential advantage. The likelihood function for our model is made up of several independent components corresponding to each household. The household size distribution of the data (Figure 4.6) comprises of mostly small and mid-sized households and a few larger households. Due to the size of the linear system to be

solved being polynomial in household size, the few larger households dominate the computational time taken to calculate a likelihood step. By computing the likelihood terms in increasing order of system (household) size it is often possible to reject a Metropolis-Hastings (algorithm 2) proposal early by drawing the uniform random number, r, for the accept/reject step before the likelihood is calculated and checking if the currently calculated likelihood (and prior) terms are enough to reject without solving the most computationally expensive households.

For a symmetric proposal kernel, $q(\cdot)$, such as the Multivariate Gaussian we have used in the previous chapter, we can define the acceptance probability after the j^{th} smallest household's likelihood has been calculated, $\alpha_j(X'|X_{i-1})$, as:

$$\alpha_j(X'|X_{i-1}) \quad = \min\left(1, \frac{\prod_{i=1}^{j} \mathbb{P}(D_i|X',N)}{\prod_{i=1}^{j} \mathbb{P}(D_i|X_{i-1},N)}\right) \tag{6.27}$$

For a given value of r there is a threshold for the difference of the log-likelihoods above which the proposal step will definitely be rejected and we can avoid calculating the likelihoods of the largest households and thus make large computational savings. Figure 6.8 illustrates the log likelihood difference necessary for each value of r to reject early (as the difference increases the likelihood of acceptance becomes vanishingly small). This leads to what we term the "quick-reject Metropolis Hastings algorithm" (QRMH) which is presented in its simplest form in algorithm 6; however we implement a version with the adaptive co-variance matrix strategy of Haario et al. [77], and a stochastic approximation that adapts the covariance scale factor σ in order to achieve the optimal acceptance rate of ≈ 0.234 for a Metropolis-Hastings algorithm.

For a $[SI_3S]_2$ household model we have observed speed-ups that average 1.55x across the four villages in our dataset. This factor will only increase with model complexity as the discrepancies in matrix generation and solution time between the smallest and largest household sizes increases.

6.6 Discussion

This chapter has explored several methods for scaling up the complexity of household models we can perform inference on.

As expected NUTS has shown itself as a robust and more efficient alternative to random-walk methods in terms of theoretical sampling efficiency, though its

performance is more ambiguous when implementation details are taken into consideration. On the other hand, employing the Quick Reject Metropolis Hastings (QRMH) algorithm that exploits the likelihood structure yields both theoretical and practical gains in performance. These improvements have allowed us to scale the inference to $[SI_3S]_2$; however as this formulation introduces multiple new parameters the posterior becomes too wide to make any meaningful inference for our specific dataset. This wide posterior appears to be due to the data and model combination being non-identifiable - we have few data-points and our model lacks sufficient biological realism for us to specify priors that constrain the model such that it is identifiable with a small number of observations. For example, correlations between the posteriors of γ_H and the β's in figure A.9 could be constrained with sufficient prior knowledge; however we cannot directly relate γ_H to any variable that could be directly measured in an epidemiological study. Computationally speaking, scaling to $[SI_3S]_3$ is also possible (though again no informative inference can be made) but we reach a regime where the time to generate the largest transition matrices exceeds the time taken to solve them. QRMH mitigates this extra slow-down somewhat but it seems that beyond this model complexity the only feasible option would be to abandon exact likelihood methods in favour of likelihood-free methods such as approximate Bayesian computation (ABC) [58, 183] or pseudo-marginal MCMC [10] unless household sizes are small.

The failure to find an efficient adaptive MALA implementation for household models is a surprising result. MALA is a much trickier beast to tame than Metropolis-Hastings, with the tendency to reject many consecutive proposals in transience - "stickiness". Some of the problems experienced are reported in the literature, for example it has been shown to have different optimal scaling properties depending on whether it is in the transient phase ($\sigma_m = O(d^{-1/3})$ [41]) or the stationary phase ($\sigma_m = O(d^{-0.5})$ [166]). This can be incredibly problematic from a tuning perspective as if optimality is achieved in transience then the stationary phase will be sub-optimal. With this in mind one strategy we have attempted to avoid this is to find the maximum a posterior (MAP) estimate of the posterior distribution and initialise the covariance matrix to be the Hessian of the log probability density at the MAP estimate (projected to the nearest positive-definite matrix if necessary) but this did not yield any significant improvements. Marshall and Roberts [122] introduce the Subsample-adapting Langevin Algorithm (SALA) which adapts the covariance at an increasing sequence of covariance adaptation times. A time-limited investigation of this approach shows that this does appear to adapt less pathologically than the approach of Atchade [12]; however the problem of stickiness remains.

Given the seeming lack of a robust adaptive approach applicable to the majority of problems, and corresponding well tested reference implementation of MALA I am of the opinion that any practitioner/mathematical modeller needing to perform Bayesian inference would be best served by using NUTS (as part of pyMC3, Stan or some other probabilistic programming language) or if their model is too niche to fit within the confines of those implementations (or if some model structure can be exploited during inference such as with QRMH) to stick to either Metropolis-Hastings or slice sampling [139] approaches which are much more robust - if you are trying to solve a specific modelling problem you want to focus your time on modelling rather than spending vast amounts of time trying to implement a marginally faster MCMC technique for your problem.

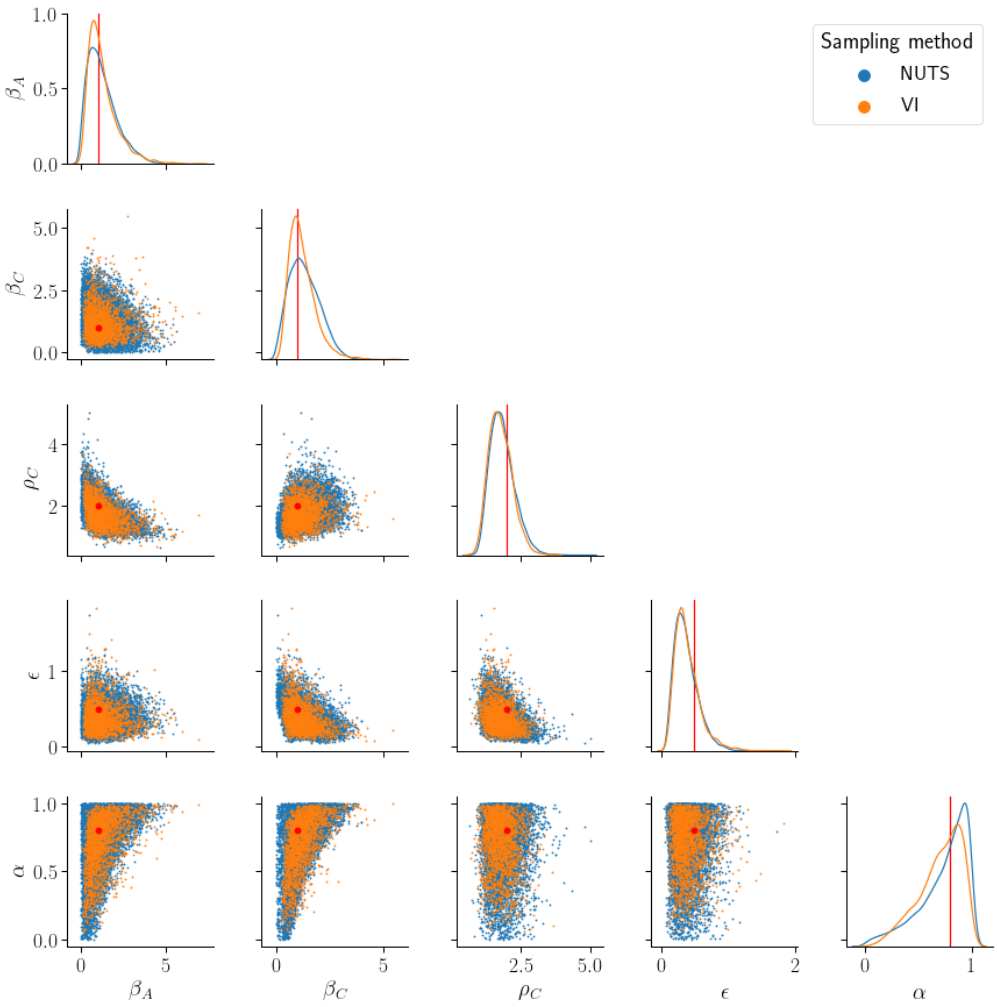

Figure 6.7: Joint posterior distributions for a household model with $[SI_1S]_2$ dynamics ($A = 2, R = 1$) using simulated data with true parameters plotted in red ($\beta_A = 1, \beta_C = 1, \rho_C = 2, \epsilon = 0.5, \alpha = 0.8$) inferred using both NUTS and ADVI.

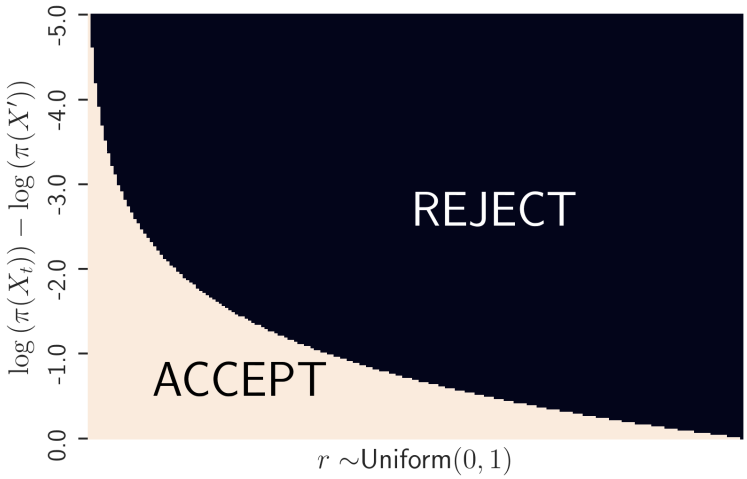

Figure 6.8: Accept-Reject regions for varying Metropolis Hastings random variable, r, and the differences in posterior density between the current iteration, X_t, and the Metropolis Hastings proposal, X', $(\log(\pi(X_t)) - \log(\pi(X')))$.

Algorithm 6 Quick reject Metropolis Hastings algorithm

$X_0 \sim \pi(x)$ ▷ Initialise
Choose $\sigma \in \mathbb{R}^+$, positive definite $\Lambda \in \mathbb{R}^+$
$q_\Lambda(y|x) \sim \mathcal{N}(x, \sigma^2\Lambda)$
for i=1,2, ... **do**
  $X' \sim q_\Lambda(\cdot|X_{i-1})$ ▷ Proposal
  $r \sim$ Uniform$(0,1)$
  for j=1,2, ..., m **do**
    $\alpha_j(X'|X_{i-1}) = \min\left(1, \frac{\prod_{i=1}^{j} \mathbb{P}(D_i|X',N)}{\prod_{i=1}^{j} \mathbb{P}(D_i|X_{i-1},N)}\right)$ ▷ Acceptance probability
    if $r \geq \alpha_j$ **then**
      $X_i \leftarrow X_{i-1}$ ▷ Reject early
      Break
    end if
  end for
  if $u < \alpha_m$ **then**
    $X_i \leftarrow X'$ ▷ Accept
  else
    $X_i \leftarrow X_{i-1}$ ▷ Reject
  end if
end for

CHAPTER 7

Forecasting drug demand for lymphatic filariasis

7.1 Introduction

The previous 3 chapters have focused on local-scale impact of mass or selective treatment on a major neglected tropical disease (NTD), soil-transmitted helminthiasis, and demonstrated the challenges in interpreting data and informing control policy. In this chapter I present a much more straightforward analysis of global surveillance data for a related NTD, and demonstrate how this analysis has informed international policy. This chapter is quite different from previous chapters in that the methods used are comparatively straightforward. However, it formed an important part of my PhD training because of the extensive exposure to working with real-life data and on the challenges of providing results in a way which can inform high-level international policy discussions. In this chapter we focus on lymphatic filariasis, a mosquito-borne helminth, which is controlled by mass drug administration (MDA).

Lymphatic filariasis (LF) is rare amongst the NTDs which use MDA, in that the global programme to eliminate lymphatic filariasis (GPELF - section 7.2.3) has already reached its targets for halting MDA in a number of areas and is now planning for how to scale back the programme. LF is able to reliably halt MDA in an area because a reliable protocol exists for measuring the disruption of transmission. The aim of this chapter was to improve the forecasts for the timeline to achieve that goal and understand where the challenges may lie in reaching the target. GPELF has a target of 2020 for all mass drug administration (MDA) programmes to be finalised [204]. A number of countries have already halted MDA and others are accelerating their programme to achieve the goals. A recent change in World Health Organisation (WHO) guidelines, introducing a new triple-drug therapy, is expected to accelerate achievement of these goals, but requires additional supply of treatments [151]. Prior

to the completion of these guidelines, there was a need for estimates of the likely time course of the international programme and the impact of alternative interventions in reaching the goal in order to assess the feasibility of the new guidelines.

The work presented in this chapter was undertaken to help fulfil this need and was shared with stakeholders and the WHO Guidelines Development Group for Lymphatic Filariasis. The model I develop in this chapter is referenced in the WHO Guidelines [151] as 'the Markov model' and was used as part of the evidence base to introduce the new guidelines. The analysis was wholly performed by myself with Dr Jonathan King, WHO, providing data and making decisions on which policy scenarios to simulate and which results were most policy relevant to present.

MDA for LF is delivered at intermediate geographic scales, such as districts, which are known as intervention units (IUs). When requesting donated drugs from WHO for the GPELF, national NTD control programmes have to report to WHO on the progress of treatment in each IU - where has treatment been delivered, what is the population at risk in those IUs, and what proportion of people were treated (treatment coverage). Therefore WHO holds a database of historic treatment at a sub-national level. In this chapter we develop a model of an MDA programme at the IU level and using historic treatment data forecast the future dynamics given the IUs current state and previous history to provide an estimate of the likely number of treatments required and the expected cost of treatments under the 'counter-factual' scenario where new guidelines are not adopted and under the 'factual' scenario where new guidelines introducing triple-drug therapy are adopted. We show that introduction of new guidelines will save on the number of treatments performed, and therefore the costs of the programme.

7.2 Lymphatic Filariasis

7.2.1 Global picture

Lymphatic Filariasis (LF) is a vector-borne neglected tropical disease (NTD) caused by infection with one or more of the parasitic nematodes: *Wuchereria bancrofti*, *Brugia Malayi*, or *Brugia timori* (less common). Transmission occurs via a variety of mosquitoes in different parts of the world. The worms live in the lymphatic vessels, and so infection can lead to damage of those vessels, potentially manifesting as lymphedema (tissue swelling), hydrocele (scrotal swelling), and elephantiasis (skin/tissue thickening) [184]. The WHO estimates that 886 million people in 52 countries worldwide are at risk of infection, with 120 million people infected in the year 2000, and about 40 million disfigured and incapacitated due to the infection

[150].

7.2.2 Biology

Figure 7.1 summarises the life cycle of the most common type of LF, *Wuchereria Bancrofti*. The adult worms, which can live for decades, live in the lymphatic system and, during their lifetime, produce millions of microfilariae (MF - immature larvae) that circulate in the blood. Female mosquitoes are infected with these circulating MF by ingesting blood when biting an infected human host. Microfilariae mature into infective larvae within the mosquito. It is important to note that there is no amplification of the number of MF in the mosquito, unlike malaria, where small numbers of gametocytes develop into many sporozoites in the mosquito. This is one of the characteristics of lymphatic filariasis which makes it a potentially 'eliminable' infection. Lymphatic filariasis is transmitted by different types of mosquitoes. The two most prominent are the Culex genus, widespread across urban and semi-urban areas and known for transmitting arboviruses such as West Nile virus, and the Anopheles genus, mainly found in rural areas, known for transmitting malaria, and the Aedes genus, mainly endemic in islands in the Pacific. When infected mosquitoes bite people, mature parasite larvae are not injected, but are deposited on the skin from where they can enter the body. The larvae then migrate to the lymphatic vessels where they develop into adult worms and sexually reproduce, thus continuing the cycle of transmission. Adult worms continue to live in the human lymphatic system with the parasites infecting lymph nodes and blocking the flow of lymph throughout the body.

About 10% of people develop the major symptoms of lymphatic filariasis, such as severe lymphedema of the limbs, leading to elephantiasis in the legs, or extreme swelling of the scrotum. Often these symptoms develop after the infection has cleared, as a result of the damage caused by the worms. These symptoms can be very severe, and lead to disability and stigma. This stigma and reduced productivity (due to debilitating acute attacks which may last weeks) leads to hundreds of millions of dollars in economic losses - perpetuating the cycle of poverty in the communities where it is highly prevalent.

Until recently, diagnosis of lymphatic filariasis was typically performed via. night-blood smear (Giemsa stain). For *Wuchereria Bancrofti*, these measurements must occur at night, as the parasite has evolved to be circulating in the blood at the time when the mosquito is most likely to be biting. More modern diagnostics such as the immunochromatographic test (ICT) for LF, and filarial test strip (FTS),

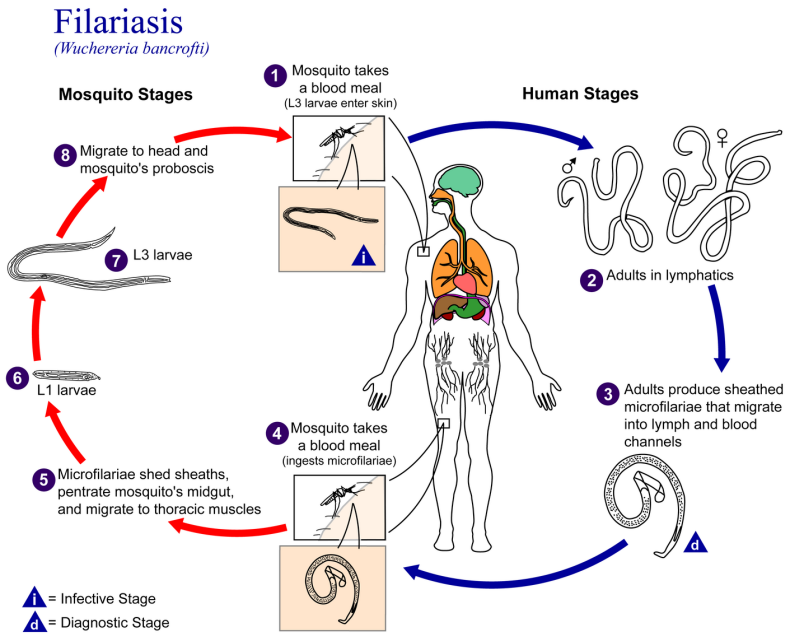

Figure 7.1: *Wuchereria bancrofti* lifecycle. Adult worms develop in the lymphatic system and give birth to transmissible stages, which are taken up by female mosquitoes taking blood meals, and then, after a period of development, are transmissible onwards in another bite. Note that there is no amplification of the number of parasites in the mosquito (unlike malaria). Obtained from the CDC Public Health Image Library. Image credit: CDC/Alexander J. da Silva, PhD/Melanie Moser., 2003.

are tests for worm antigens circulating in the blood, and as such can be done at any time of day and require less expertise to read as they are rapid tests.

Lymphatic filariasis can be treated by anthelmintic drugs with an extremely good safety profile. The most effective treatment is Diethylcarbamazine (DEC), but this cannot be used if there is any risk that the individual has onchocerciasis, an infection with another filarial worm, for fear of causing blindness or death. In most of the world, where onchocerciasis is not co-endemic with lymphatic filariasis, DEC is combined with albendazole (ALB) for mass drug administration. In areas where onchocerciasis may be co-endemic (such as most of Africa), the preferred treatment is a combination of ivermectin (IVM) and ALB. The antibiotic doxycycline has also been found to be effective in treating lymphatic filariasis, likely through killing the symbiotic bacteria, Wolbachia, but it can take weeks of treatment.

7.2.3 Global Programme to Eliminate Lymphatic Filariasis

Since lymphatic filariasis is a sexually reproducing organism, and there is a highly skewed distribution of worms between people (as shown for STH in the previous chapters), it is expected that when prevalence becomes very low, transmission could halt because not enough people are infected with both male and female worms. Unlike other diseases, this has actually been tested for lymphatic filariasis. An extensive control programme in China in the 1980s showed that mass drug administration could bring prevalence from high levels like 30%, to less than 1%, and that at this level the infection did not resurge[179]. This reinforced the idea that lymphatic filariasis might be 'eliminable'.

In addition to these data from China, the demonstration of the effectiveness and safety of the combination treatments above (both individually and as a mass treatment), and that there is no animal reservoir for lymphatic filariasis made the evidence base from which to attempt to try to eliminate lymphatic filariasis stronger [32, 82, 159, 188, 203]. This was recognised by the World Health Assembly resolution WHA50.29 which encourages Member States to eliminate lymphatic filariasis as a public health problem. In response, WHO launched the Global Programme to Eliminate Lymphatic Filariasis (GPELF) in 2000. An important part of this programme has been the donation of drugs to WHO by three leading pharmaceutical companies. WHO recommends [153] the use of preventive chemotherapy (PCT) via. mass deworming with anthelmintics, such as albendazole (ALB) donated by GSK, ivermectin (IVM) donated by Merck & Co., and diethylcarbamazine (DEC) donated by Eisai. Other interventions such as vector control and bed-nets may also be required. Over the years since 2000, over 7.1 billion treatments have been delivered, leading to reductions in the burden of disease and the health economic burden [191].

The control programme for LF is unusual amongst the NTDs because there is a stopping plan, based on the experience in China, and it is being implemented. Knowing when to stop a control programme is a challenging problem that is the subject of ongoing research [54]. The stopping strategy is based around three transmission assessment surveys [149] (TAS), performed three years apart. A transmission assessment survey (TAS) is designed to measure whether units have lowered the prevalence of infection to a level where recrudescence is unlikely to occur, even in the absence of MDA interventions. The first one informs the stopping decision, and the second and third TAS (approximately three and six years later) are for surveillance post-MDA. There are currently some concerns that the design of the transmission assessment survey is inadequate, as there is evidence of ongoing transmission in areas that have passed all three TAS surveys [162]. Ongoing studies are evaluating the im-

portance of this effect and different methodologies for the TAS survey. It is possible that these may lead to extra rounds of treatment, an effect which we did not model in this work. The population in an IU is considered to no longer require MDA once transmission assessment surveys (TAS) have passed - the number of children testing positive for LF infection is less than the allowed critical cut-off value. A failed TAS indicates persistent transmission after MDA.

Prior to the work of this chapter WHO recommended at least five years of annual MDA with effective coverage (defined as more than 65% coverage of the total eligible population) prior to assessing impact on infection levels with a TAS. After five effective MDA rounds a pre-transmission assessment survey (pre-TAS) is performed. This is a follow-up assessment of sentinel and spot-check sites with an IU passing pre-TAS when the prevalence of infection has been reduced to less than 1% microfilaraemia or less than 2% antigenaemia. If pre-TAS is passed an IU may progress to the TAS otherwise it continues treatment. Whilst tablets are donated for free, the logistical costs of administering drugs, increasing compliance, and monitoring are significant.

With the success of these stopping decisions seeming to lead to effective control of LF, in 2012, the WHO published the neglected tropical diseases roadmap [148] which proposed the target date for achieving elimination of LF, i.e. all countries having halted MDA, by 2020.

7.2.4 Triple-drug therapy question/challenges

As of 2016, 856.4 million people in 53 countries required PCT; 495.6 million people across 40 countries reported to have been treated; and 20 countries no longer required MDA. 2016 was the last year for countries to initiate effective MDA in all endemic IUs in order to complete the 5 annual rounds by the end of 2020. 22 countries were yet to scale up their control programmes and were thus unable to achieve 5 rounds by 2020 unless a more effective and shorter strategy is used [152]. Recent findings of a clinical trial [187] indicated that combining all the anthelmintics used to treat LF (IVM, DEC, and ALB) improved efficacy of treatment with the clearing of all MF from the blood within a matter of weeks compared to years using a 2-drug regimen. Equally remarkably, the study found evidence that treatment with the triple-drug therapy may have macrofilaricidal properties. This triple-drug therapy is referred to as IDA, with each letter corresponding to the first letter of one of the three drugs. These findings brought about the possibility of new guidelines which could accelerate the programmes of countries yet to start treatment or those unlikely to stop by 2020. Irvine et al. [93] developed a transmission model to compare the number of rounds

of MDA needed to achieve MF prevalence of <1% for both the triple-drug therapy and current two-drug therapy showing that IDA has the potential to reach <1% in dramatically fewer rounds. Given the results of Thomsen et al. [187] and Irvine et al. [93] the question remained whether such an acceleration was logistically possible. A forecast of the number of treatments required under each scenario was particularly important to assess the cost-effectiveness of using IDA and whether an additional production line for IVM would be necessary to meet increased demand. Furthermore, in order to pass a guideline review committee then these projections are needed to form part of the evidence base. The analysis of this chapter fulfils this policy need.

7.3 Model of intervention

In this section we develop two models of intervention for LF within an intervention unit (IU) - one for the original guidelines and one for the triple-drug (IDA) therapy. These models describe the progress of an IU through the LF treatment guidelines discussed in sections 7.2.3-7.2.4. IUs are treated independently with no transmission dynamics modelled.

We define a number of treatment states that an IU can occupy, with the transitions between these states described by a discrete-time Markov Chain (see section 2.2.1) with a time step representing one year (treatment for LF is given annually). The initial conditions are inferred from historic treatment data and expert knowledge (section 7.3.5). Initial conditions are both the state at which the IU 'starts' in, but also the year at which we begin the model. The year at which an IU simulation started varied because the latest data available for an IU varied, or treatment was to begin at some point in the future.

The key idea beyond this modelling approach is that given the initial treatment state and year for an IU, we may solve the Markov chain such that we get the probability (under our model) that an IU is in a given treatment state in a given year. This allows us to calculate quantities such as the expected number of treatments given each year which used in combination with data about the IU (population size, GDP, population eligible for treatment, treatment regimen etc.) gives the number of tablets required and cost of treatment (section 7.3.6). These IU level estimates may then be combined at the country or WHO epidemiological region level to make statements about when a country is likely to stop MDA. Furthermore, by contrasting between the factual-scenario (introduce IDA in year t_{IDA}) and the counter-factual scenario (continue with the original guidelines) we estimate the potential savings in terms of treatments, rounds, and costs.

7.3.1 Counter-factual scenario: original guidelines

The first model is for the counter-factual scenario where the original guidelines as laid out in section 7.2.3 are continued. The model state-space and transition probabilities are summarised in figure 7.2 with a summary of each state given in table 7.1. Each square represents a treatment state, i, an IU can occupy in our Markov chain; possible state transitions and their transition probabilities are denoted by orange arrows with rates in black; and state boxes enclosed by light blue correspond to states where one MDA round is given ($\mathbb{1}_2(i) = 1$).

The model is started at some year, t_0, with some initial probability distribution across states, p_0, inferred from historic data (section 7.3.5). The 'State description' given for each state in table 7.1 describes the state that an IU is in at the start of a time-step and actions that are performed 'during' each time step. The outcomes of actions (or compositions of actions) dictate the transitions.

i	**State label**	**State description**	$\mathbb{1}_2(i)$
0	0 ER	IU has received zero effective rounds of MDA. *Action*: Round given.	1
1	1 ER	IU has received one effective round of MDA. *Action*: Round given.	1
2	2 ER	IU has received two effective rounds of MDA. *Action*: Round given.	1
3	3 ER	IU has received three effective rounds of MDA. *Action*: Round given.	1
4	4 ER	IU has received four effective rounds of MDA. *Action*: Round given. If effective, pre-TAS.	1
5	TAS	pre-TAS passed. *Action*: Round given and TAS.	1
6	MDA+1	pre-TAS/TAS failed. Need further treatment. *Action*: Round given.	1
7	MDA+2	Continuing further treatment. *Action*: Round given and pre-TAS.	1
8	TAS repeat	Secondary pre-TAS passed. *Action*: TAS.	0
9	STOP	TAS passed and MDA has stopped. *Action*: None.	0

Table 7.1: Description of Markov chain model states under original guidelines. i denotes the integer index corresponding to our chosen state-space ordering; 'State label' gives the name of the state; 'State description' gives a description of the state and actions taken whilst in that state; and $\mathbb{1}_2(i)$ is 1 if being in the state involves a round of MDA.

The earliest treatment state an IU can start in is '0 ER' (it has received zero effective rounds). Each year it receives one more round of treatment. For the initial rounds (the five effective rounds required under the guidelines) of treatment, a round is successful with probability P_{eff} which causes a transition to the next state. An ineffective round does not result in a transition.

Upon transition to the state '4 ER' the IU has received 4 effective rounds. In this state it receives a round of treatment. If the round is effective, the IU will have received the 5 effective rounds required by the treatment guidelines, therefore a pre-TAS (pre-Transmission Assessment Survey) is conducted within the same time step. If the round is not effective it will remain in state '4 ER', but if the round is effective the state transitioned to is dependent on the outcome of the pre-TAS. If pre-TAS passes (probability $(1 - P_1)$) then a transition to 'TAS' occurs, otherwise a transition to 'MDA+ 1' occurs. Therefore the probability of transitioning from '4ER' to 'TAS' is $P_{\text{eff}} \cdot (1 - P_1)$ as the round must be effective (to trigger pre-TAS) and pre-TAS must be passed.

In the 'TAS' state the TAS (Transmission Assessment Survey) action is performed. Due to the fact that medicines must be requested a year in advance (when the outcome of pre-TAS is not yet known) a further round of MDA is also given in a TAS year. As with pre-TAS the outcome of TAS determines the next treatment state, TAS is passed with probability $(1 - P_2)$ and treatment for an IU stops - a transition to the absorbing state, 'STOP', is made. If TAS is failed (probability P_2) then a transition to 'MDA+ 1' occurs. As explained in section 7.2.3, there are in fact three TAS performed three years apart. Our model makes a simplifying assumption that an IU will see no sustained transmission after the first TAS is passed. This is because there was inadequate data to estimate the small probability of recrudescence after passing a TAS. Furthermore, as TAS are spaced by 3 years any further failed TAS will occur well after the peak number of global treatments.

'MDA+ 1' and 'MDA+ 2' represent additional rounds of intensive MDA given in an IU when the control programme for the IU has failed under the treatment guidelines - i.e. five effective rounds were not enough to pass pre-TAS and TAS. These are assumed to be effective with probability 1 (there is no P_{eff} term) as problems with a control programme typically exhibit a more intensive response.

A further pre-TAS is performed during 'MDA+ 2', after which a transition to state 'TAS repeat' occurs. These repeat pre-TAS and TAS are assumed to have a 100% pass rate. Failures of subsequent pre-TAS or TAS are empirically unlikely - there was insufficient data available for WHO to provide an estimate.

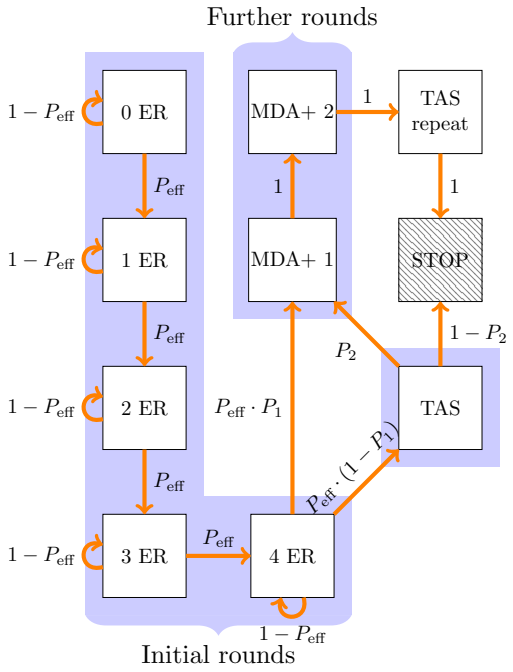

Figure 7.2: Discrete-time Markov chain model under the original guidelines. States are in black bordered rectangles; blue boxes represent states in which MDA is given ($\mathbb{1}_2$ is one); orange arrows and their adjacent black values give the probability of transitioning between states. For example, if at t_n the IU is in the state '0 ER' then the probability that it is in '1 ER' at t_{n+1} is P_{eff}.

The probability distribution at time $t \in \mathbb{N}$ may be calculated to machine-precision by calculating the following matrix multiplication: $\boldsymbol{p}_t = \boldsymbol{Q}^{(t-t_0)}\boldsymbol{p}_{t_0}$. $\boldsymbol{p}_{t_0}$ is the initial probability distribution of the intervention unit, and $\boldsymbol{Q} \in \mathbb{R}^{10\times 10}$ is the transition matrix where $\boldsymbol{Q}_{ij}$ gives the probability of transitioning from state j to state i. $\boldsymbol{Q}$ is specified by the transition rates in figure 7.2, and imposing an ordering on the states (such as that in column 'i' of table 7.1).

7.3.2 Factual scenario: Introduction of IDA

The factual scenario considers the introduction of the IDA triple-drug therapy (section 7.2.4) at some time, t_{IDA}, for any IUs that are eligible. For $t < t_{\text{IDA}}$, the model under the original guidelines applies (section 7.3.1). At $t = t_{\text{IDA}}$ an 'instantaneous' remapping from the state space of the original guidelines model to the state space of the IDA model is applied using a projection operator, φ. For $t \geq t_{\text{IDA}}$, the model state-space and transition rates of the IDA model are shown in figure 7.3 with a

i	**State label**	**State description**	$\mathbb{1}_2(i)$	$\mathbb{1}_3(i)$
0	4 ER	IU has received four effective rounds of MDA *Action:* Round given. If effective, pre-TAS.	1	0
1	TAS	pre-TAS passed. *Action:* Round given and TAS.	1	0
2	MDA+ 1	pre-TAS/TAS failed. Need further treatment. *Action:* Further round given.	1	0
3	MDA+ 2	Continuing further treatment. *Action:* Further round given and pre-TAS.	1	0
4	TAS repeat	Secondary pre-TAS passed. *Action:* TAS.	0	0
5	STOP	TAS passed and MDA has stopped. *Action:* None.	0	0
6	3 ER left	IU requires three effective rounds of IDA. *Action:* Round given.	0	1
7	2 ER left	IU requires two effective rounds of IDA. *Action:* Round given.	0	1
8	1 ER left	IU requires one effective rounds of IDA. *Action:* Round given. If effective, pre-TAS.	0	1
9	TAS (IDA)	pre-TAS passed. *Action:* Round given and TAS.	0	1
10	MDA+1 (IDA)	pre-TAS/TAS failed. Need further treatment. *Action:* Round given and TAS.	0	1

Table 7.2: Description of Markov chain model states for treatment with IDA. i denotes the integer index corresponding to our chosen state-space ordering; 'State label' gives the name of the state; 'State description' gives a description of the state and actions taken whilst in that state; $\mathbb{1}_2(i)$ is 1 if being in the state involves a round of 2-drug MDA; and $\mathbb{1}_3(i)$ is 1 if being in the state involves a round of IDA MDA.

summary of each state given in table 7.2.

If at the treatment switchover an IU has received: 0 effective rounds then it will require 3 effective rounds of IDA MDA; 1 or 2 effective rounds then it will require 2 effective rounds of IDA MDA; 3 effective rounds then it will require 1 effective round of IDA MDA; 4 effective rounds then IDA treatment is not given unless pre-TAS/TAS are failed; 5 effective rounds then IDA treatment is not given unless a pre-TAS or TAS are failed. If either pre-TAS or TAS are failed after the introduction of IDA, then 1 further round of IDA MDA is given (rather than the further 2 rounds under the original guidelines). This mapping between state-spaces is performed by the projection operator, φ, and is illustrated in figure 7.3 with dashed blue lines pointing from the model states under the old guidelines (states outside the solid blue line) to model states under the introduction of IDA (states inside the

solid blue line). The fact that the mapping under φ is non-injective requires the IDA states to count down until the requisite number of effective rounds has been received (compared to the counting up to 5 effective rounds for the counter-factual model).

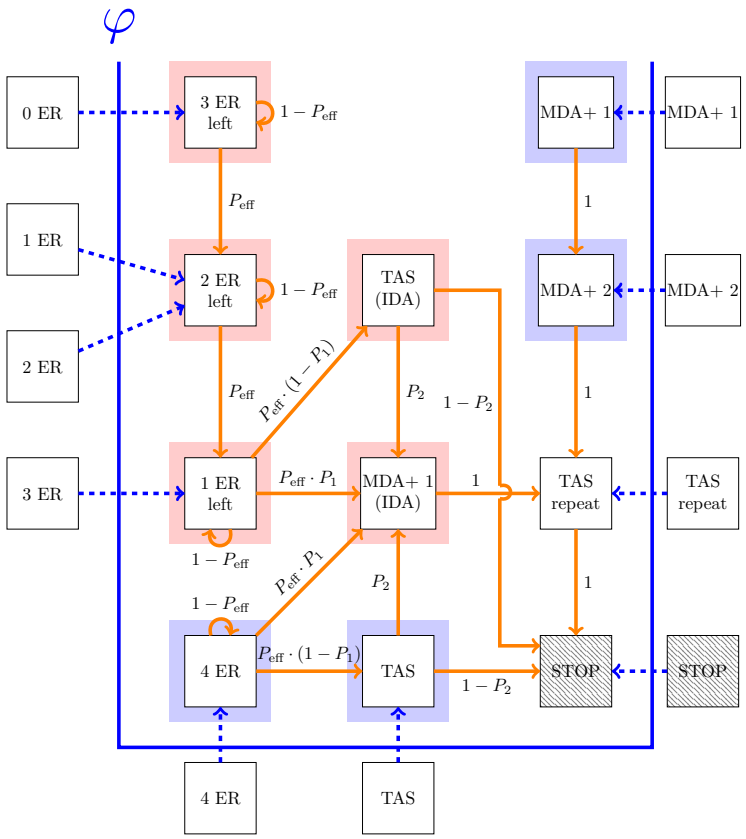

Figure 7.3: Discrete-time Markov chain model after the switch-over to IDA treatment. States in blue boxes represent treatments under the old guidelines ($\mathbb{1}_2$ is one); states in red boxes represent IDA treatments ($\mathbb{1}_3$ is one); orange arrows and their adjacent black values give the probability of state transitions; the states bordering the edge correspond to the states of the model prior to the introduction of IDA, with the dashed blue lines representing the mapping from the old model state-space to the new model state-space under application of the projection operator φ. For example, if at $t_n = t_{\text{IDA}}$ the IU is in the state '0 ER' then φ is applied which maps the IU to state '3 ER left', and the probability that it is in '2 ER left' at t_{n+1} is P_{eff}.

The system can be solved similarly to the previous subsection,

$$\boldsymbol{p}_t = \boldsymbol{Q}_{\text{IDA}}^{(t-t_{\text{IDA}})} \cdot \boldsymbol{\varphi} \cdot \boldsymbol{Q}^{(t_{\text{IDA}}-t_0)} \cdot \boldsymbol{p}_{t_0} \tag{7.1}$$

Where $\boldsymbol{Q} \in \mathbb{R}^{10\times10}$ and $\boldsymbol{Q}_{\text{IDA}} \in \mathbb{R}^{11\times11}$ are the transition matrices before and after the introduction of IDA respectively. $\boldsymbol{\varphi} \in \{0,1\}^{10\times11}$ is specified s.t. if $\boldsymbol{\varphi}_{i,j} = 1$ then state i under the original guideline model maps to state j under the IDA model, with the blue dashed lines in figure 7.3 specifying this mapping.

7.3.3 Simulated policy scenario

Due to the need to respond to changing policy scenarios desired by stakeholders, the models were implemented in a way that 'generated' Markov chains for different scenarios as differing scenarios change both the state-space and the rates. For example, a scenario where two further rounds of IDA are given after failing pre-TAS or TAS would require an extra state in the model 'MDA+ 2 (IDA)', with the transition rates for 'MDA+ 1 (IDA)' changing such that transitions are to 'MDA+ 2 (IDA)'. The models in sections 7.3.1-7.3.2 correspond to the treatment scenarios simulated to inform the guideline review process, where the required number of rounds under the introduction of IDA was informed based on previous modelling work by Irvine et al. [93].

7.3.4 Model parameters

With the policy scenario set it remains to parameterise the model. The probability of an MDA round being effective was taken to be $P_{\text{eff}} = 0.737$ (according to treatment data reported to WHO). The probability of failing pre-TAS was taken to be $P_1 = 10\%$; and the probability of failing TAS, P_2, was estimated based on the dominant parasite type or by Country where sufficient data existed: Wucheria Bancrofti $P_2 = 6\%$; Brugia Malayi $P_2 = 36\%$; India $P_2 = 15\%$. Pre-TAS and TAS failure rates were estimated by WHO from epidemiological survey data. For the factual scenario, IDA was introduced into 19 IDA eligible countries at $t_{\text{IDA}} = 2018$. Not all Countries had scaled up to 100% geographic coverage therefore assumptions were made about when this is likely to happen. Where available 2017 JRSM's (joint request for selected medicines) were used to assess whether an intervention unit would receive MDA in 2017. If no request for medicine was made for 2017 (or if no data was available) then MDA was assumed to start in 2018. Therefore, this analysis assumes that the latest year every country will scale-up to 100% geographic coverage is 2018. An informal sensitivity analysis of the model to the probability that a round is effective (P_{eff}), and the pre-TAS & TAS failure probabilities was performed; however in light of the simplistic nature of the model and the fact that the model yields a probability distribution across model states, we did not perform a formal sensitivity analysis.

7.3.5 Inferring initial conditions

Historic data was obtained from WHO detailing treatments at the level of an IU: population estimates; LF and onchocerciasis/loaisis (where appropriate) endemicity statuses; and treatment coverage by year. To begin modelling the intervention strategy for each IU, we need to extract and encode their treatment history to set the initial conditions for our Markov model. This includes whether LF is endemic, how many effective rounds have been given, and the 'current' year (some countries provide timely updates while some have a large delay in reporting and some countries have not started).

For the vast majority of IUs this simply involves counting the number of rounds with greater than or equal to 65% coverage and the year in which the last treatment data was available. There are a few scenarios in which this is not the case.

Firstly, if an IU did not receive treatment for three consecutive years then their running count of effective rounds was reset to zero.

Countries are required to submit Joint Reporting Forms (JRF) and Joint Request for Selected PCT Medicines (JRSM) to quantify the number of anthelmintics given and the number required for the next year respectively. Over the course of this project new JRF & JRSM forms were merged into the dataset, where available, to provide updates on population, endemicity, whether treatments were performed (JRF), and whether treatments are planned (JRSM). This process was crucial to provide updates to the forecasts; however it was one of the most challenging aspects of this project. Data quality suffered due to the fact that the information these forms contain must be: collected at the lowest levels within an IU; aggregated at an IU level; sent to the country's ministry of health; and finally reported to WHO. A prime example of difficulties encountered are that IUs are not afforded a unique identifier and are instead identified by name which is often misspelled or does not have a unique English translation. Assuming a one-to-one mapping of IUs, fuzzy-matching based on the Levenshtein distance [116] easily overcomes this, although we were not so fortunate because countries merge/split/rename IUs with no explicit reference in the data to this. Development of a bespoke matching algorithm and subsequent manual validation were necessary to overcome this hurdle, and minimise the errors made during updates.

Within the data some IUs contained an extra variable that gave an 'expert' estimate of which year that IU is likely to reach (not necessarily pass) the first TAS. This typically occurred for IUs that have had a complex treatment history, where the full story is not contained in the recorded history alone. Examples of a complex treatment history which may necessitate this revised (possibly optimistic

or pessimistic) estimate are the treatment programme disrupted by Ebola or other disruptive events, or that there may be known issues with the reported effectiveness of a programme. If this variable exists for an IU then regardless of its recorded treatment history we assume that the starting year of the simulation for this IU is two years before the estimated TAS1 year and that the IU is in the 'ER4' state - these two years help introduce uncertainty into whether TAS is reached and passed to reflect that the expert judgement is just an estimate. If the TAS estimate year has already occurred and a JRF/JRSM reporting form contradicts this estimate then it is disregarded.

Finally, it was not always possible to assign an IU an initial condition in our state space. For IUs who had not yet received an effective round it was assumed that treatment would start in 2018 by default unless a JRSM form for 2017 made a request for medicines. Furthermore, we made the assumption that Eritrea, Zambia, and Timor-Leste started MDA in all IUs in 2015. For IUs who had received more than 5 effective rounds we assumed that pre-TAS had been failed such that in the model state space they are in the state where the IU has received the 5th ER and has failed pre-TAS.

The processing of this data according to the above rules allowed us to determine a treatment state and starting year for our model for each IU. Furthermore, this data also provided the treatment type, IDA eligibility, dominant parasite type, and population size.

7.3.6 Costing model

It's crucial to not only assess whether the introduction of IDA treatment will decrease time to elimination and the number of tablets required but to demonstrate that IDA will be cost effective. This is particularly important from the perspective of a country's ministry of health which has limited resources to dedicate to many competing priorities.

Using the work of Fitzpatrick et al. [60] we estimate the cost of implementing MDA for a given intervention unit. The authors conducted a literature review of previous costing studies and performed a meta-regression to estimate the cost per person per round of mass treatment against NTD's. This facilitates setting specific unit costs against which we can compare the costs of our two scenarios. We used the authors' random effects meta-regression model to calculate the economic costs without use of volunteers to administer MDA in the first year of a programme at 80% coverage for one disease in a non-school based setting. National GDP data and IU population size were used to provide country and IU level heterogeneity.

7.3.7 Computing treatments and costs

We now have all the constituent parts to forecast costs and treatments at the country level: historic treatment data allows us to infer the year and model state to begin the model at for each IU; solving the model for each subsequent year yields the probability distribution across treatment states for each IU; and the costing model allows us to estimates the costs due to each round of treatment for an IU.

The probability that a treatment occurs in year t for treatment $d \in \{2, 3\}$ (2-drug MDA and triple-drug MDA), is the sum of the probability of being in each state multiplied by whether that state corresponds to treatment being given. This is trivially combined with data on the population eligible for treatment, and cost of performing treatment in the IU to estimate the number of millions of people treated and the cost of treatment. Aggregating these figures at the country level facilitates comparison between our scenarios in terms of how much faster, how much cheaper, and how many fewer treatments are required before MDA can be stopped in a country. When calculating the number of people treated by one round of MDA we multiply the IU population by 0.8 - the target coverage for a round of MDA.

7.4 Results

This section presents the results obtained by applying the data processing and modelling the policy scenarios documented in the previous section. In brief, we use the epidemiological data documenting the treatment histories of intervention units (IUs) to infer the stage at which every IUs programme is at in the state space of our model. This sets the initial conditions for each IUs Markov chain. Running our model forward from these initial conditions yields the probability distribution across treatment states and thus the probability of treatment happening in each year for each IU (which are treated independently). Treatment probabilities are weighted by IU population data and merged with the costing model to translate expected number of treatments into expected costs of delivery. This is performed for both the factual (introduce IDA) and counter-factual (do not introduce IDA) scenarios.

Of the 19 IDA eligible countries, our forecasts predict 356 million fewer total treatments will be required if IDA is introduced leading to $277M of savings in undiscounted economic costs (not accounting for any IDA related benefits to patients/communities). Of the countries receiving IVM+ALB only Sudan was considered eligible for IDA due to the limited foci of onchocerciasis there - use of DEC is contraindicated by the presence of onchocerciasis. The number of treatments (in millions), number of rounds given to IUs, and cost of treatment are given in table 7.3

	Region	AFRO	AMRO	EMR	SEARO	WPRO	TOTAL
Factual	Treatments	122	14	41	825	39	1040
	Rounds	2959	646	885	3664	884	9038
	Cost ($M)	100	17	52	473	48	691
Counter-factual	Treatments	184	16	65	1073	58	1397
	Rounds	3135	1680	885	6360	1005	13065
	Cost ($M)	151	22	83	641	71	969
Saved	Treatments	63	2	24	248	19	356
	Rounds	176	1034	0	2696	121	4027
	Cost ($M)	51	5	31	168	23	278

Table 7.3: The expected number of individuals treated in millions ('Treatments'); economic costs of treatments ('Cost'); and total number of rounds given to IUs ('Rounds') for both the factual (introduce IDA) and counter-factual scenarios (do not introduce IDA). Numbers are given by WHO region, and the difference between factual and counter-factual numbers are also given ('Saved').

for the factual and counter-factual scenarios along with the number of treatments, rounds, and costs saved by introducing IDA. These are given at the WHO regional level as we do not have permission to share results at the country level.

Figure 7.5 shows the number of IDA eligible countries obtaining a certain percentage of IUs finishing MDA by year, and demonstrates the rapid acceleration that IDA can provide to the programmes of these countries - this is shown for all countries under the counter-factual scenario in figure 7.4. In addition to giving figures corresponding to the assumed effective coverage of 73.7% ($P_{\text{eff}} = 0.737$), these figures show the significant effect on finishing times if effective coverage could be increased to 100%.

With the introduction of new guidelines [151] that this analysis informed, in countries using DEC and ALB for the elimination of LF, WHO now recommends annual IDA rather than annual DA where:

- IUs have not started treatment, or have fewer than 4 effective rounds of DA.
- IUs have not met the appropriate epidemiological targets in sentinel and spot-check site surveys or in TAS despite meeting coverage targets.
- post-MDA/post-validation surveillance communities within an IU observe infections suggesting local transmission.

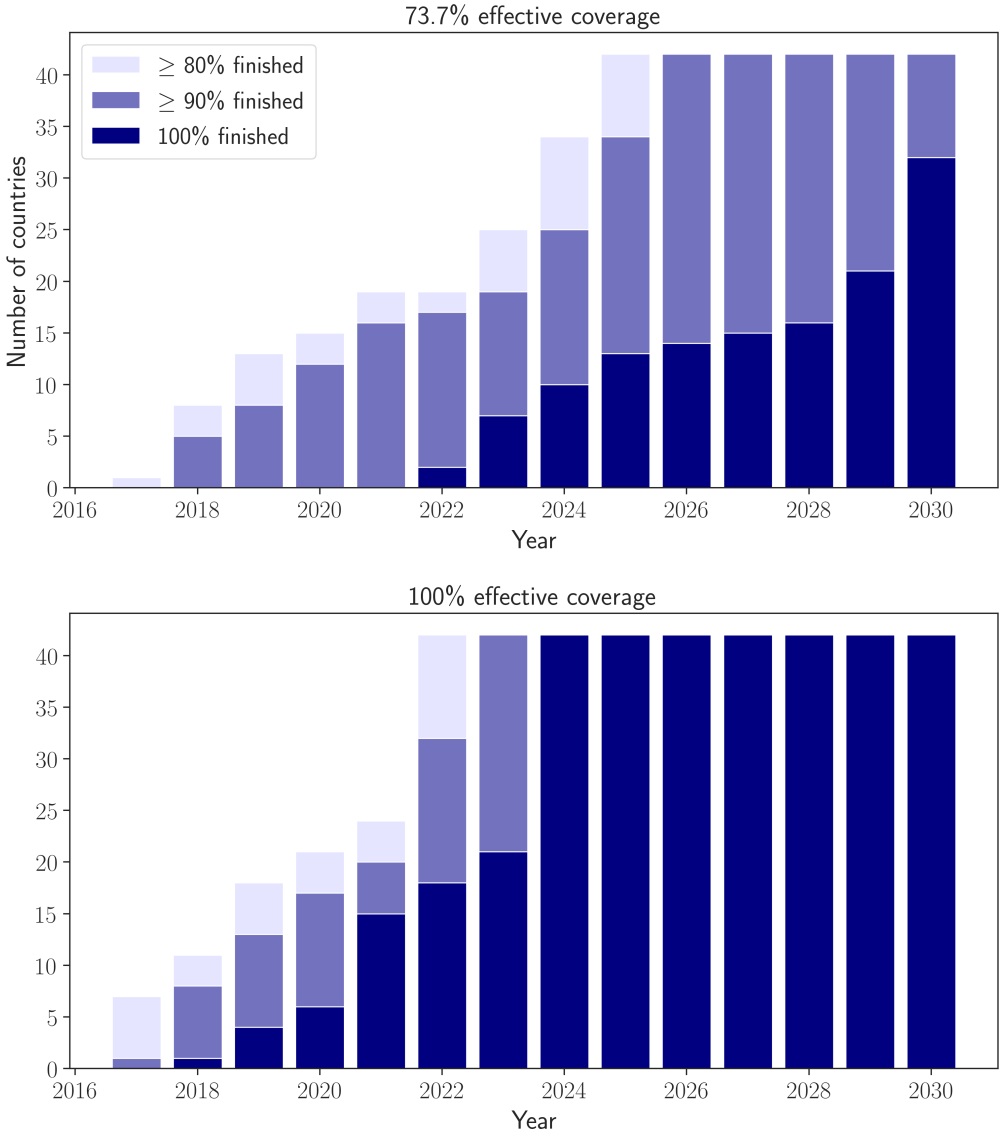

Figure 7.4: Number of countries with a certain percentage of IUs that have stopped MDA each year under the counter-factual scenario. This is plotted for effective coverages of 73.7% (top) and 100% (bottom).

7.5 Discussion

It is important for modellers to communicate to stakeholders and policy-makers that Mathematical models cannot account for all the complexities of the real world whilst simultaneously making the case that they can provide invaluable insights in the form of evidence-based policy decisions. In the context of this analysis it is important to

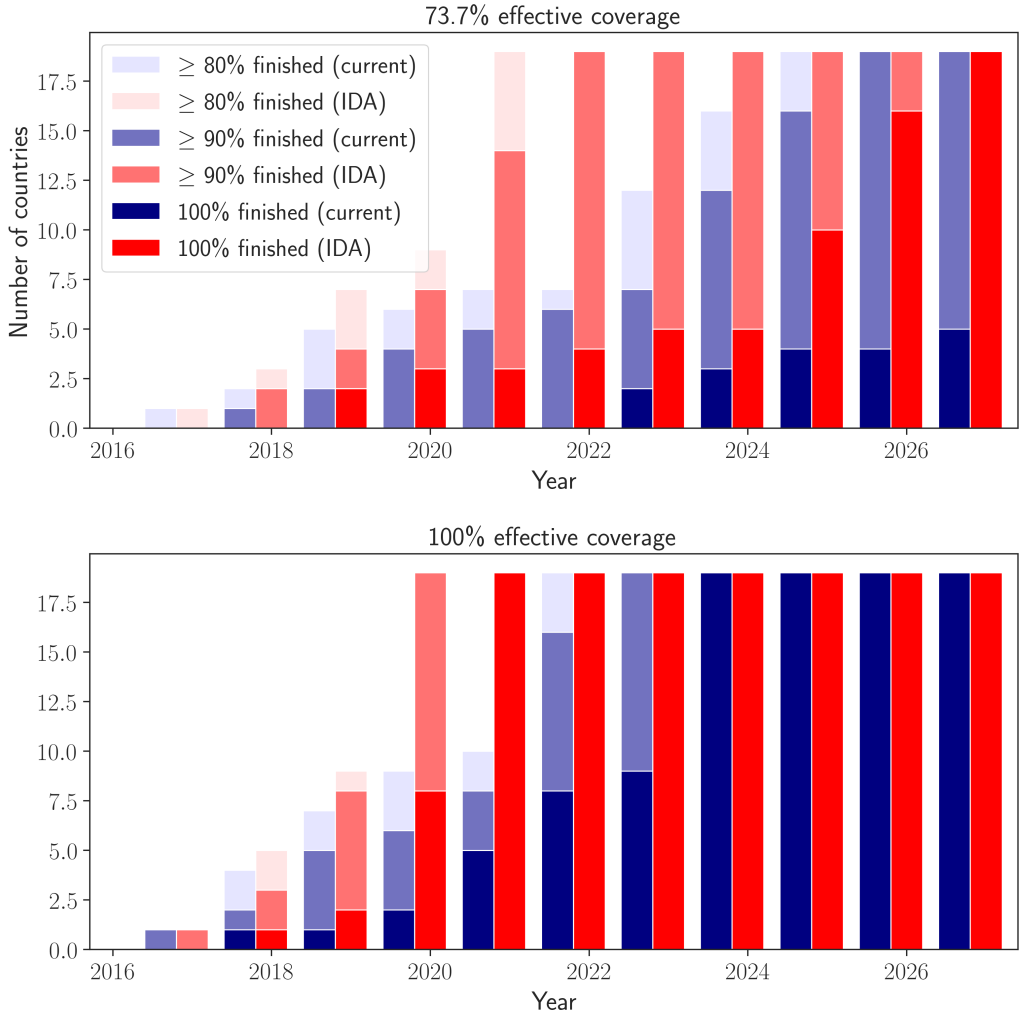

Figure 7.5: Number of IDA eligible countries with a certain percentage of IUs that have stopped MDA each year under both the factual (red) and counter-factual scenario (blue). This is plotted for effective coverages of 73.7% (top) and 100% (bottom).

see these results as forecasts and not set in stone. There are limitations, both in the model and in the data, which limit the predictive accuracy of this analysis. Firstly, there are known issues with the accuracy of reported treatment coverage, therefore estimates of the current treatment states of intervention units (before the model is applied) will not be 100% accurate. Moreover, not all Countries have achieved 100% global scale-up leading us to assume when this may occur. We assumed that 100% global scale-up will be achieved in 2018 at the latest, yet it was not certain if that would happen. Furthermore, pre-TAS and TAS pass rates were estimated from a

limited amount of historic data, therefore little/no geographic heterogeneity could be incorporated. As a result, the global picture should give an accurate representation; however results on a smaller scale (for example IU, or even country level) will be subject to lower predictive accuracy. Similar to pre-TAS and TAS rates, there is limited historic data available to estimate the probability of a round being effective ($\geq$ 65% coverage) therefore we considered the pessimistic scenario where the probability of effectiveness is estimated from historic data ($P_{\text{eff}} = 0.737$) - the optimistic scenario where all rounds are effective ($P_{\text{eff}} = 1$) was also presented for figures 7.4-7.5 which demonstrated that elimination could be achieved significantly faster if effort was made to increase P_{eff}. The pessimistic scenario ($P_{\text{eff}} = 0.737$) is labelled as such because it can be seen to be "double-counting" inefficiencies in a programme as there will be correlations between P_{eff} and the likelihood of failing pre-TAS & TAS. Finally, the figures above presented to our stakeholders are based on the expected number of treatments. Our modelling approach was chosen such that uncertainty could be easily captured; however stakeholders felt that the reporting of such uncertainty was not accessible to a policy audience.

In conclusion, we have performed data analysis to assess the state of the global programme to eliminate lymphatic filariasis (GPELF) based on MDA treatment data at the intervention unit (IU) level and have extended this through mathematical modelling to forecast continued progress and economic costs. Modelling was performed for both existing guidelines and under the introduction of the triple drug therapy (IDA) into eligible countries. This analysis demonstrated that introduction of IDA was likely to result in both significant cost savings as well as accelerating the time until an IU is likely to halt MDA. The results were used as part of the evidence base that led to the introduction of new treatment guidelines [151] which recommended the introduction of treatment with the triple-drug therapy, IDA, in countries using DEC and ALB for the elimination of LF.

CHAPTER 8

Conclusion and outlook

The chapters of this thesis all required an appropriate compromise to be found between the realistic representation of population structure in epidemic models, model tractability, and identifiability from data. In addition to the substantive insights gained for the problems in hand, this thesis gives general methodological insights into how to resolve these compromises, which are ubiquitous in the study of complex systems.

We have provided novel insight on the dynamics of epidemics on clustered networks, the household transmission dynamics of STH and contributed to the evidence base that resulted in a change of treatment guidelines for lymphatic filariasis. Furthermore, we have contributed methodological developments, particularly in chapters 4 and 5 where we fill an apparent gap in the literature of household models. Looking across the chapters with their various data, methods, and models, this thesis represents a broad training and output in mathematical modelling and complexity science.

The STH work specifically demonstrates that even though the more complex model is the best representation of the system of variable worm burdens and population structure, it generates such challenges for inference that ultimately a more simple model is likely to provide the current limit to our understanding of this system.

In chapter 3 we showed how epidemic dynamics on degree heterogeneous and clustered networks - that reflect the key structural components of our societies - can be well approximated by a set of ODEs under appropriate closure assumption which avoids having to use computationally intensive event-driven simulations. Further to this, our approach improves on the accuracy of the best known approximation in the literature.

In chapter 4 we were able to show the household clustering present in the distribution of Soil-Transmitted Helminths (STH) within a population using hierar-

chical statistical models. Fully capturing the hierarchical nature of the problem came at the expense of neglecting the dynamic nature of spreading processes. Chapter 5 laid the groundwork of a generic methodology for performing inference on stochastic household models with multiple demographic classes and infection levels and used this to study the nature of household driven transmission for STH in a manner that bridges the gap between Anderson and May type epidemiological models and more classical empirical approaches like those of chapter 4. Due to the nature of the data and the disease dynamics this provided a large computational hurdle to inference which ultimately limited the immediate policy relevance of this analysis. Chapter 6 focused on pushing the scalability of this class of household models. More efficient algorithms for simulating the epidemic dynamics were employed that exploited the iterative nature of MCMC inference. Next, two advanced gradient-based MCMC techniques, Metropolis Adjusted Langevin Algorithm (MALA) and the No U-turn Sampler (NUTS), were implemented in an attempt to reduce the total number of MCMC iterations required to obtain a sufficient number of samples from the posterior distribution. NUTS proved to sample more efficiently than the Metropolis Hastings approach used previously while MALA suffered from a poorly performing adaptive mechanism and a tendency for the chain to reject many subsequent proposals. The implementation of NUTS was done by writing a custom wrapper around our model that allowed it to be used within the probabilistic programming language pyMC3 which enabled us to briefly explore the possibility of using Variational Inference (VI). Our experiments with VI showed that the posterior approximations obtained became increasingly biased and poorly behaved as the complexity of our models increased. Finally, this chapter coupled the model solution and inference procedure such that the full likelihood did not always need to be calculated as it often possible to reject an MCMC proposal early before the most computationally intensive steps have been performed. The work of this chapter combined enabled us to manage large increases in the complexity of models we could perform inference on but was ultimately limited by the fact that the quantity of data available for our problem was not sufficient for so many parameters.

Chapter 7 showed that the introduction of new guidelines for the control of Lymphatic Filariasis would result in large reductions in the cost of the global control programme and would rapidly accelerate the time until mass drug administration (MDA) was completed in countries that were eligible for treatment under the new guidelines. This analysis contributed to the evidence base that led to the adoption of the new guidelines by WHO. Despite being by far the simplest modelling approach taken in this thesis it has had by far the biggest impact.

Many avenues for future work exist and have been identified. Firstly, the work of chapter 3 can be extended to explore the effects of different closures such as the closure of House and Keeling [88] or the maximum entropy method [171]. Additionally, one could consider deriving a formulation of the model that incorporates more advanced dynamics such as waning immunity. The general household model methodology developed in chapters 4 & 5 would be best served in two ways. Firstly, emulation of the stationary distribution using mixture density networks [27] could be performed in order to provide a computationally cheap approximation to the stationary distribution. Secondly, as mentioned in the discussion of the chapter, likelihood-free methods such as approximate Bayesian computation (ABC) [58, 183] or pseudo-marginal MCMC [10] could be implemented and compared to their exact-likelihood counterparts to establish the extent to which approximate inference can be performed accurately. Our modelling approach does not necessarily represent an advancement on these approaches—given sufficient data this approach would permit a higher complexity model; however this loses the guarantees of exact inference which is a highly desirable property when modelling in the context of public health. Finally, the work of chapter 7 may be extended by providing an in-depth analysis of the epidemiological data possessed by WHO so that more heterogeneity may be incorporated into the transition probabilities of the model. Interest has also been expressed by stakeholders involved in the MDA programmes for other diseases in developing a similar approach specific to their diseases.

This thesis does not provide a "manual" for how to choose an appropriate level of complexity for a given problem but provides a series of real-world examples documenting how such decisions were performed. The nuances of real-world problems and data available about them are too subtle to provide such a manual, but it is hoped that my own experiences may aid the odd lost or curious modeller.

APPENDIX A

Appendix : MCMC traceplots and diagnostics

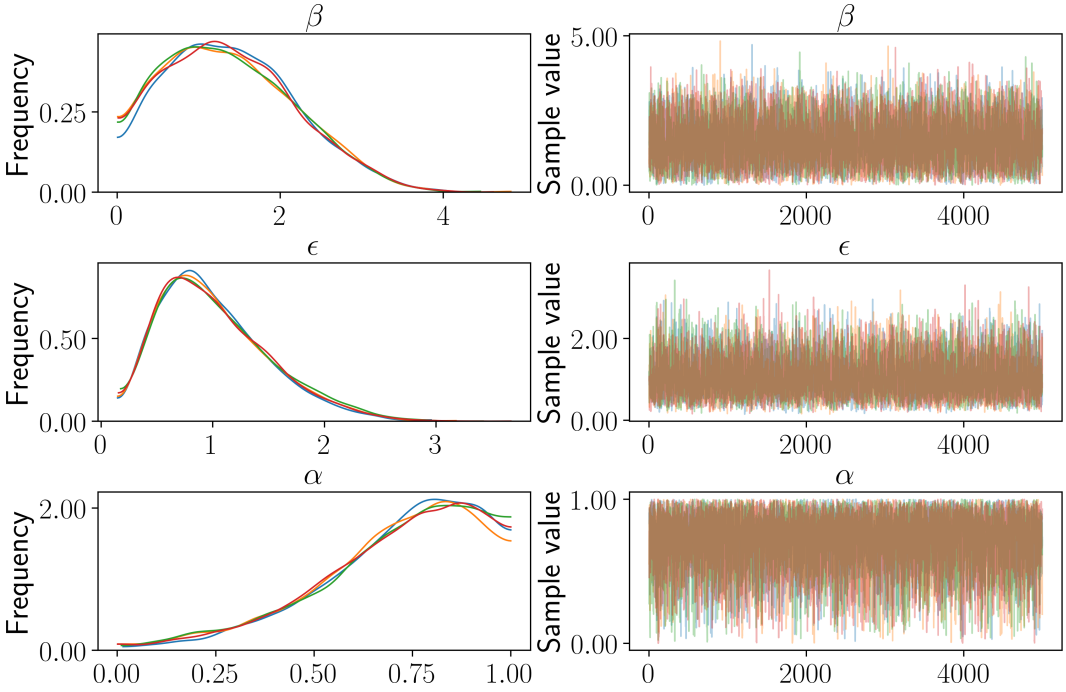

Figure A.1: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic SIS household model for Village 0. The model was run and adaptation processes tuned for $10,000$ iterations before taking the $5,000$ samples pictured.

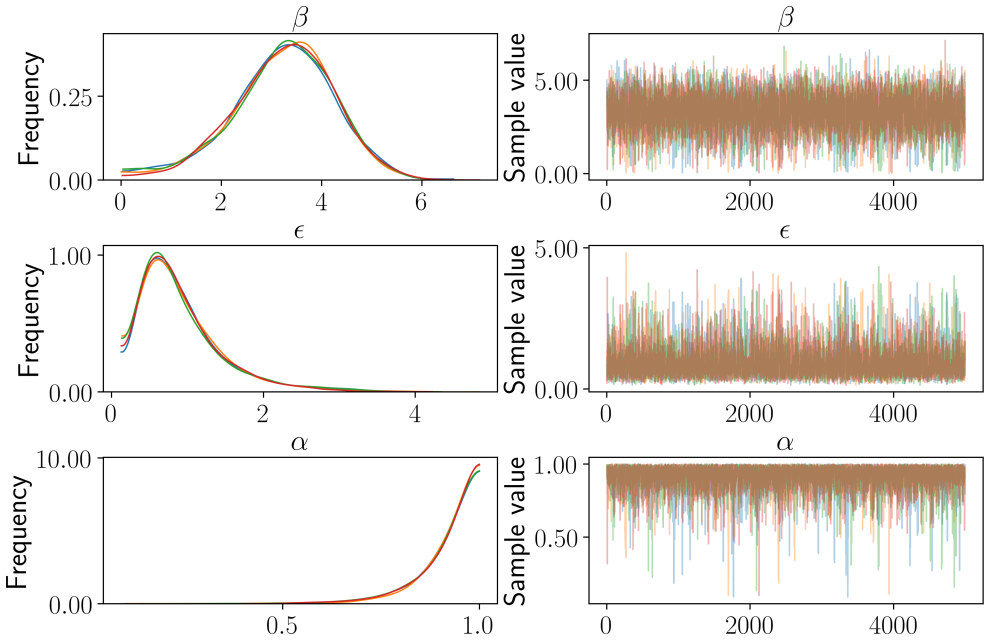

Figure A.2: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic SIS household model for Village 1. The model was run and adaptation processes tuned for 10, 000 iterations before taking the 5, 000 samples pictured.

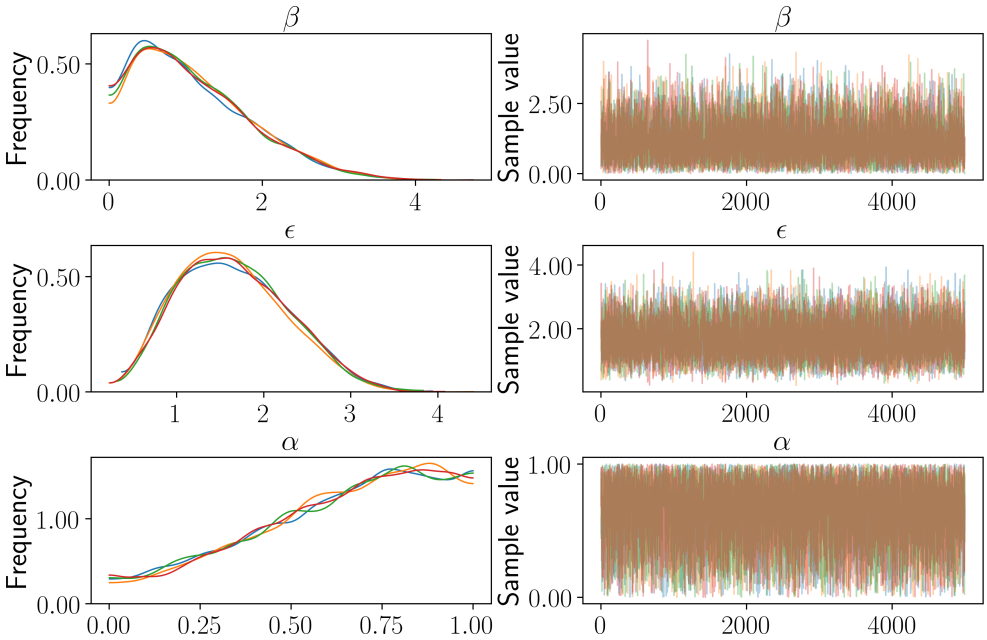

Figure A.3: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic SIS household model for Village 2. The model was run and adaptation processes tuned for 10, 000 iterations before taking the 5, 000 samples pictured.

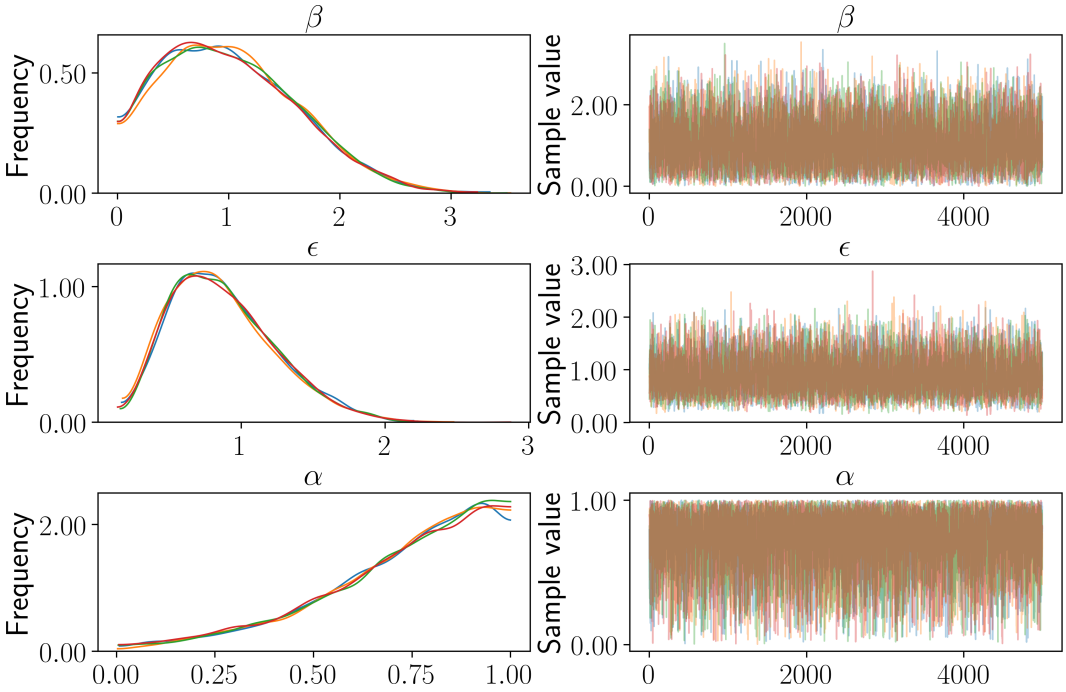

Figure A.4: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic SIS household model for Village 3. The model was run and adaptation processes tuned for $10,000$ iterations before taking the $5,000$ samples pictured.

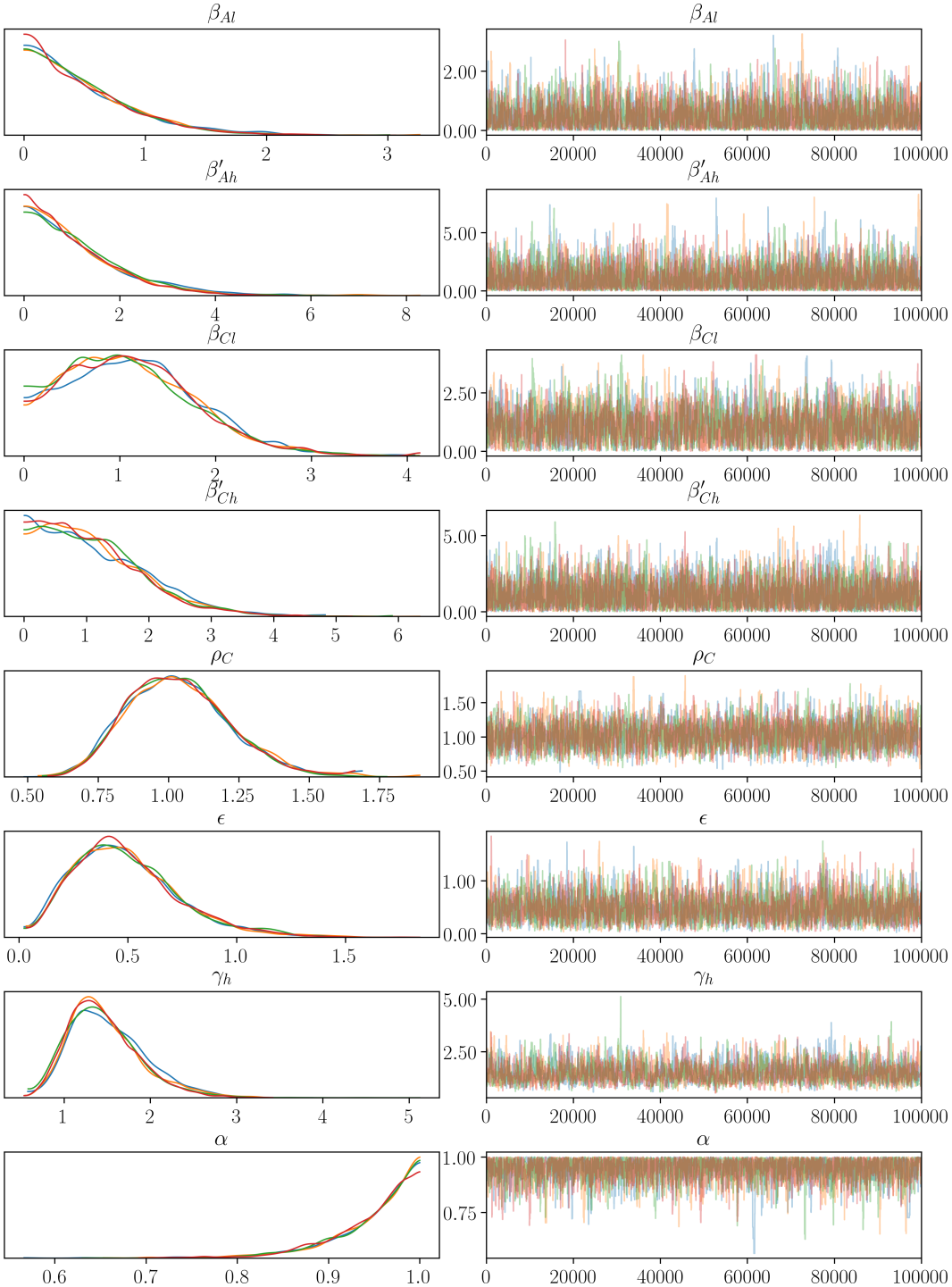

Figure A.5: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic $[SI_2S]_2$ household model for Village 0. The model was run and adaptation processes tuned for $100,000$ iterations before taking the $100,000$ samples pictured.

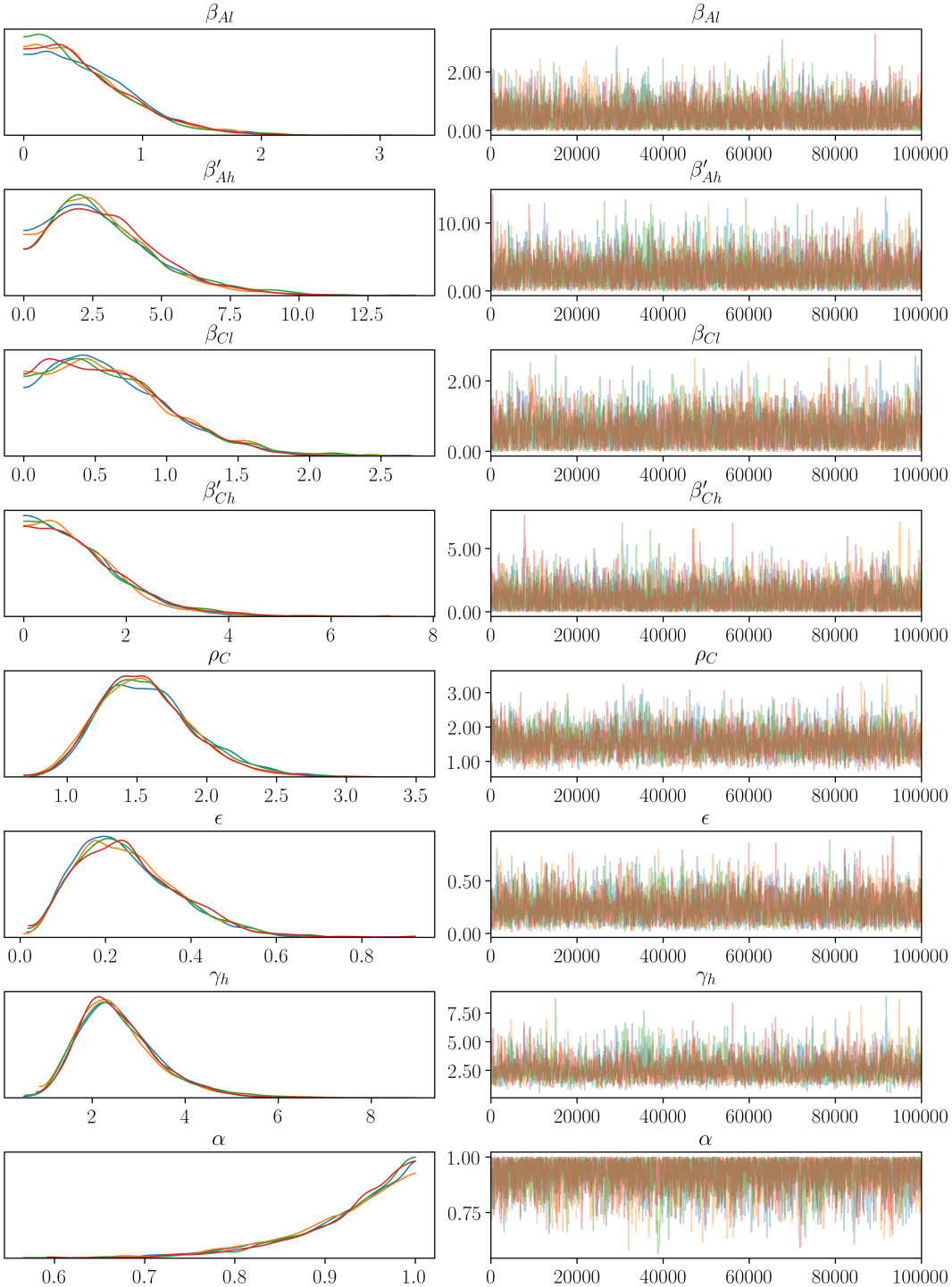

Figure A.6: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic $[SI_2S]_2$ household model for Village 1. The model was run and adaptation processes tuned for $100,000$ iterations before taking the $100,000$ samples pictured.

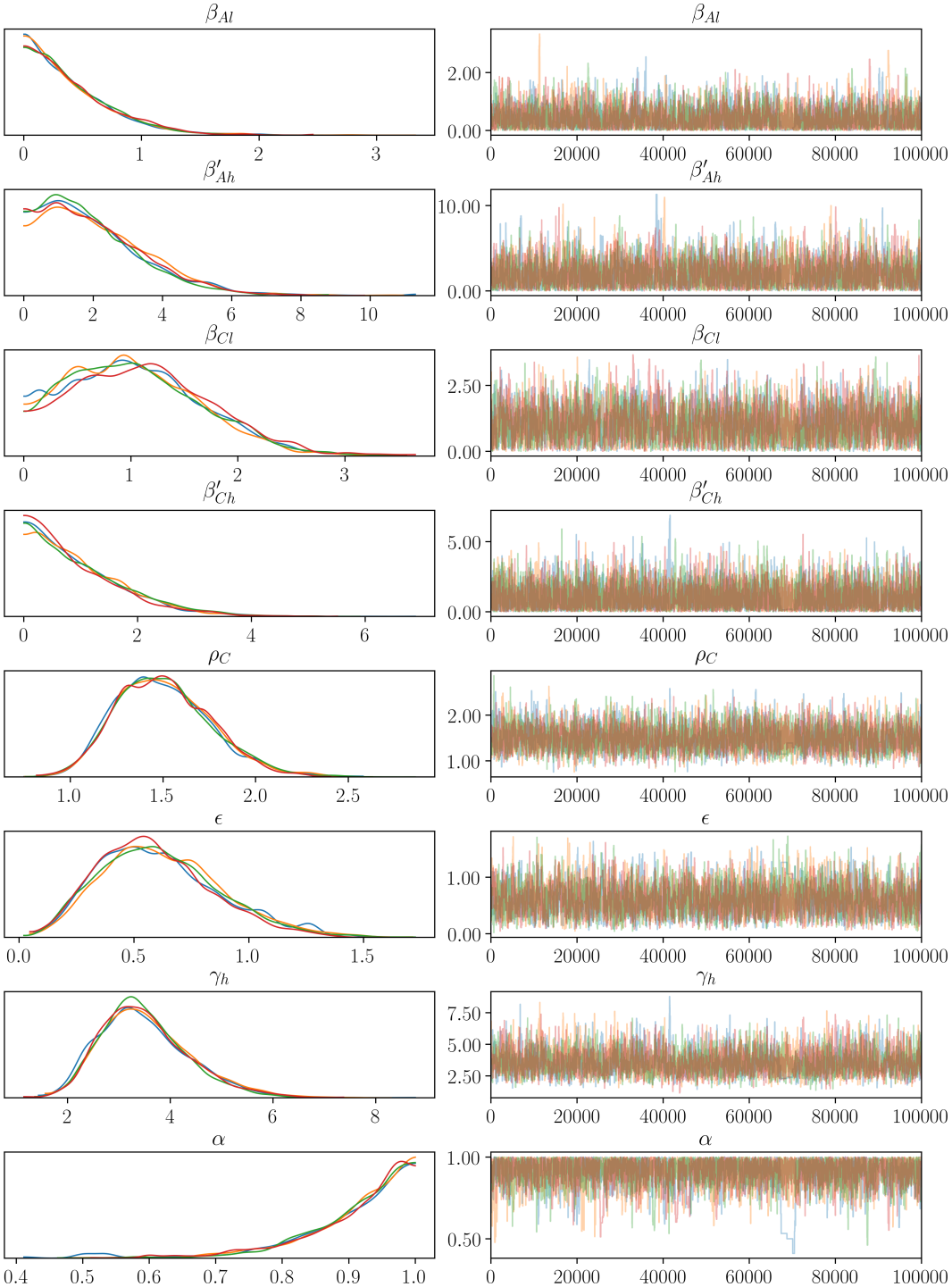

Figure A.7: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic $[SI_2S]_2$ household model for Village 2. The model was run and adaptation processes tuned for $100,000$ iterations before taking the $100,000$ samples pictured.

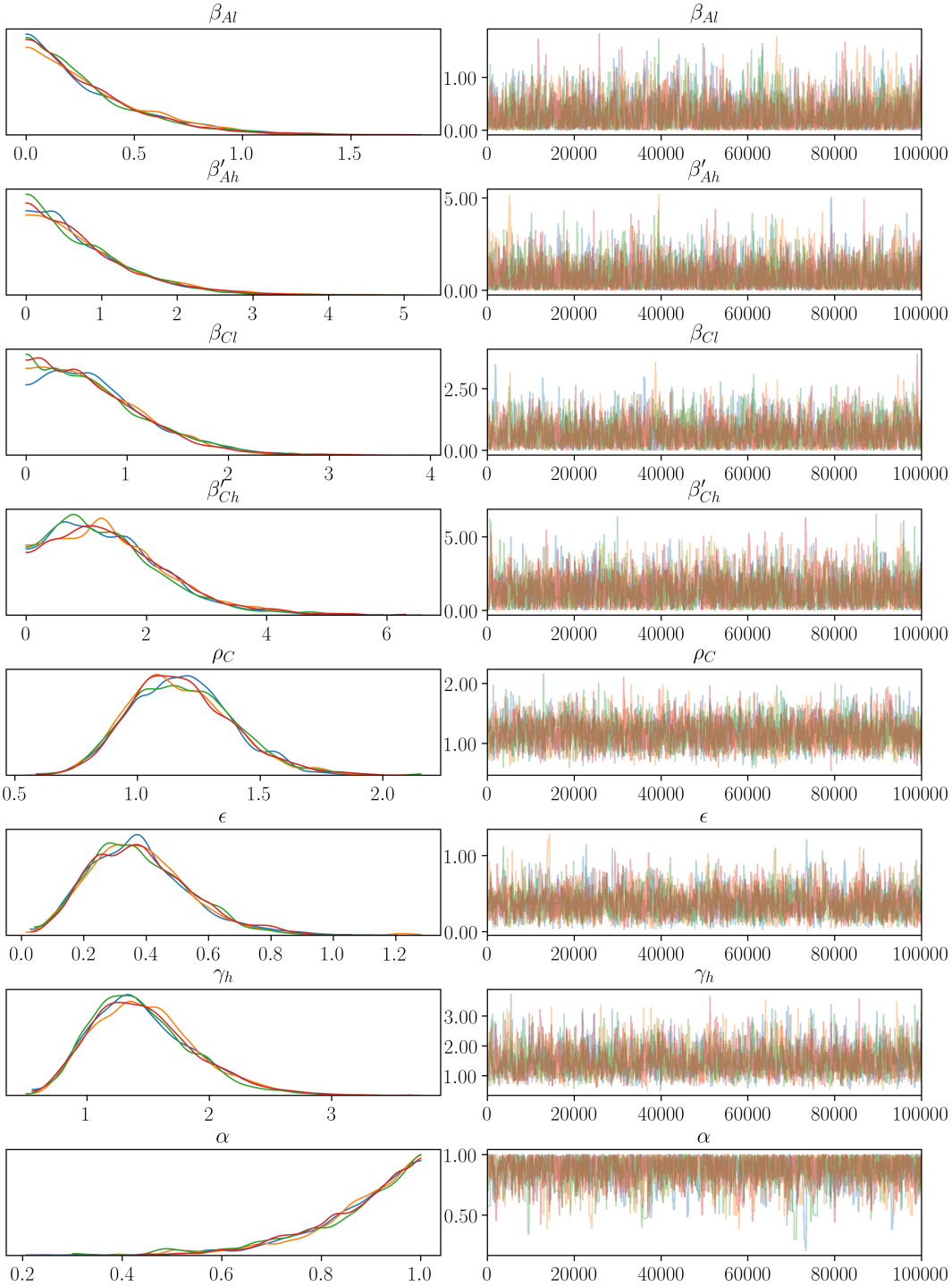

Figure A.8: Traceplots and posterior distribution obtained by running Metropolis Hastings MCMC on the stochastic $[SI_2S]_2$ household model for Village 3. The model was run and adaptation processes tuned for $100,000$ iterations before taking the $100,000$ samples pictured.

Parameter	village	Mean	s.d.	HPD 2.5	HPD 97.5	n_{eff}	$\widehat{R}$
α	0	0.95	0.05	0.85	1.00	2182.61	1.0
	1	0.92	0.07	0.79	1.00	2300.09	1.0
	2	0.91	0.09	0.74	1.00	1398.68	1.0
	3	0.87	0.12	0.63	1.00	1218.71	1.0
β'_{Ah}	0	1.68	1.06	0.12	3.73	1481.84	1.0
	1	3.51	1.90	0.60	7.40	2041.93	1.0
	2	2.48	1.49	0.25	5.40	1711.84	1.0
	3	1.10	0.71	0.06	2.45	1422.95	1.0
β_{Al}	0	0.51	0.44	0.00	1.37	2156.90	1.0
	1	0.54	0.41	0.00	1.33	2278.39	1.0
	2	0.42	0.37	0.00	1.13	1729.75	1.0
	3	0.30	0.27	0.00	0.84	1926.23	1.0
β'_{Ch}	0	2.27	0.87	0.75	4.04	1644.67	1.0
	1	1.79	0.88	0.36	3.53	1794.99	1.0
	2	2.08	0.88	0.46	3.86	1182.87	1.0
	3	2.10	0.89	0.56	3.91	1329.98	1.0
β_{Cl}	0	1.12	0.68	0.00	2.33	1656.89	1.0
	1	0.62	0.42	0.00	1.42	1791.07	1.0
	2	1.06	0.62	0.00	2.16	1513.22	1.0
	3	0.70	0.52	0.00	1.70	1404.39	1.0
ϵ	0	0.49	0.24	0.08	0.96	1982.37	1.0
	1	0.25	0.12	0.05	0.49	2266.50	1.0
	2	0.61	0.26	0.14	1.10	1725.73	1.0
	3	0.38	0.16	0.08	0.68	1520.62	1.0
γ_h	0	1.47	0.42	0.76	2.36	1375.60	1.0
	1	2.57	0.88	1.09	4.33	1715.72	1.0
	2	3.50	0.87	2.03	5.32	1588.64	1.0
	3	1.46	0.42	0.72	2.27	1596.12	1.0
ρ_C	0	1.03	0.18	0.72	1.42	1754.46	1.0
	1	1.57	0.34	0.97	2.28	2058.03	1.0
	2	1.50	0.25	1.04	2.00	2740.94	1.0
	3	1.18	0.21	0.78	1.58	1671.19	1.0

Table A.1: Summary of posterior fits (figures A.5-A.8) for the stochastic $[SI_2S]_2$ household model across villages. Mean and s.d. give the posterior mean and standard deviation; MC error gives the monte carlo error; HPD 2.5 & HPD 97.5 refer to highest posterior density intervals; n_{eff} gives the number of effective samples; and $\widehat{R}$ gives the Gelman-Rubin convergence diagnostic.

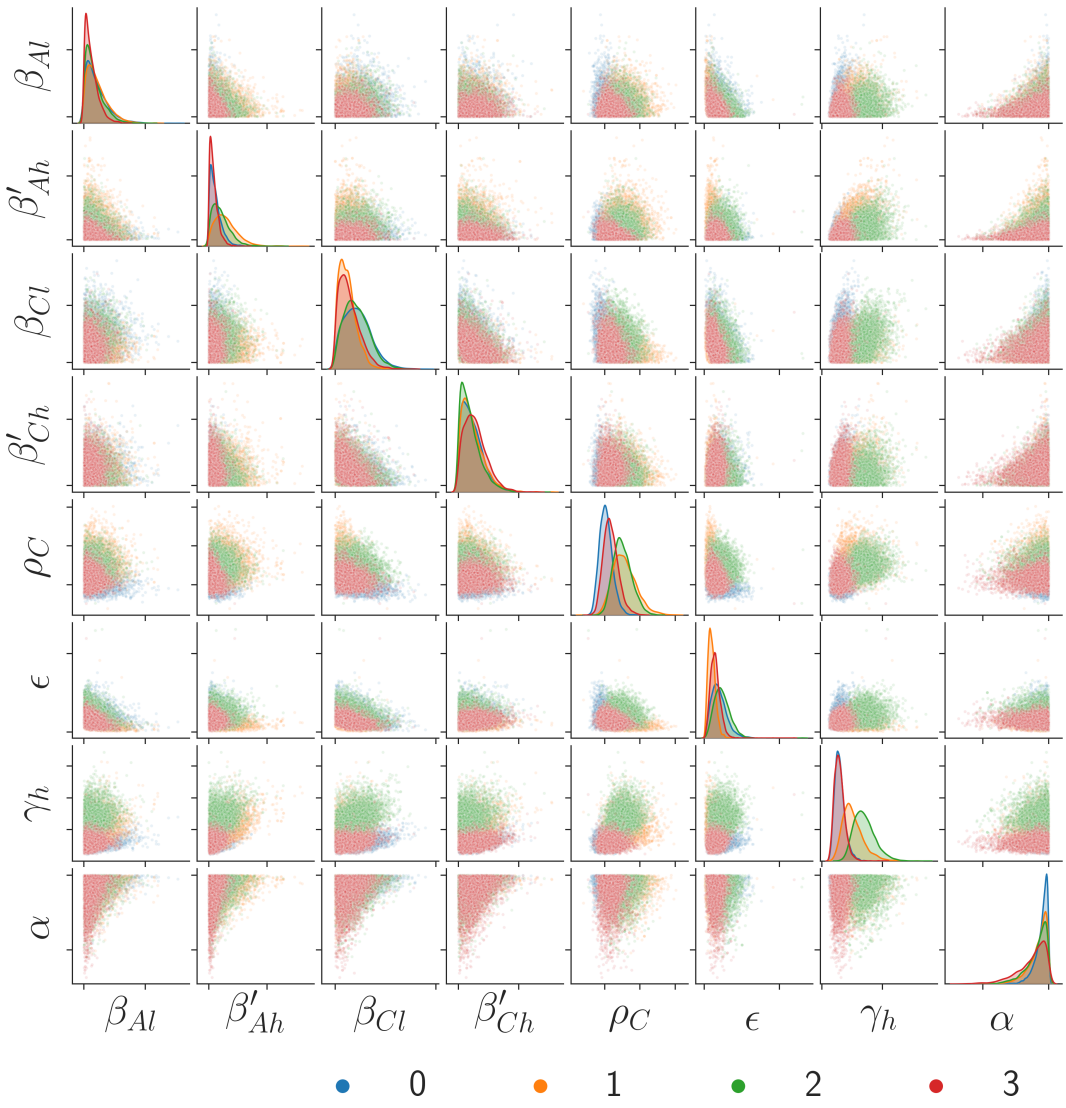

Figure A.9: Posterior joint distribution for the stochastic $[SI_2S]_2$ household model across villages. Village 0 (Blue), Village 1 (Orange), Village 2 (Green), Village 3 (Red). β_{Al} (β_{Cl}) is transmission rate of an adult (child) at the lowest infection level; β'_{Ah} (β'_{Ch}) is the transmission rate of an adult (child) at the highest infection level minus β_{Al} (β_{Cl}); ρ_C is the relative susceptibility of the child class compared to the adult class; ϵ is the external force of infection; γ_h is the recovery rate of both age classes from the highest infection level to the lowest infection level; and α is the frequency dependence parameter.

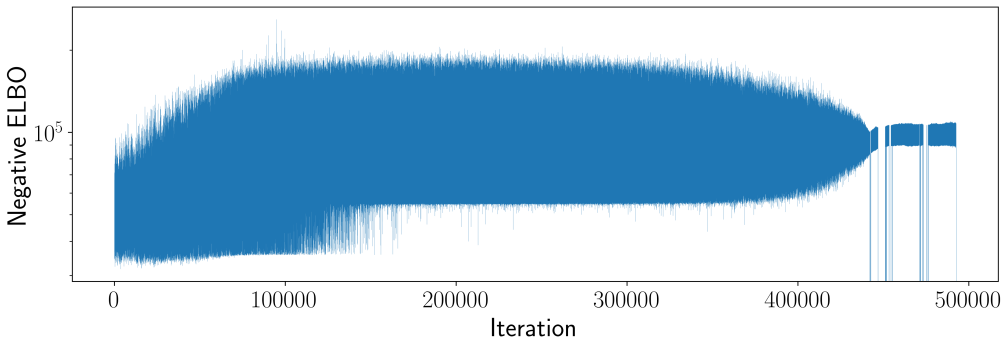

Figure A.10: Negative ELBO for each iteration of ADVI for $[SI_2S]_2$ dynamics with Uniform priors.

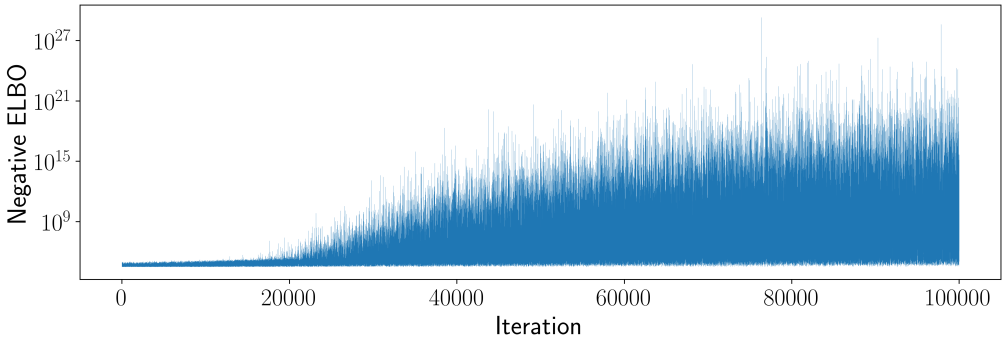

Figure A.11: Negative ELBO for each iteration of ADVI for $[SI_2S]_2$ dynamics with HalfCauchy priors.

Bibliography

[1] II Abubakar, T Tillmann, and A Banerjee. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet*, 385(9963):117–171, 2015.

[2] FR Adler and M Kretzschmar. Aggregation and stability in parasite-host models. *Parasitology*, 104(2):199–205, 1992.

[3] Hirotugu Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer, 1974.

[4] Berni J Alder and Thomas Everett Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[5] Roy M Anderson and Robert M May. Regulation and stability of host-parasite population interactions. i. regulatory processes. *Journal of animal ecology*, 47 (1):219–247, 1978.

[6] Roy M Anderson, James E Truscott, Rachel L Pullan, Simon J Brooker, and T Deirdre Hollingsworth. How effective is school-based deworming for the community-wide control of soil-transmitted helminths? *PLoS neglected tropical diseases*, 7(2):e2027, 2013.

[7] Tomohiro Ando. Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, 94(2):443–458, 2007.

[8] Tomohiro Ando. Predictive bayesian model selection. *American Journal of Mathematical and Management Sciences*, 31(1-2):13–38, 2011.

[9] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive mcmc. *Statistics and computing*, 18(4):343–373, 2008.

[10] Christophe Andrieu, Gareth O Roberts, et al. The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

[11] SO Asaolu, CV Holland, and DWT Crompton. Community control of ascaris lumbricoides in rural oyo state, nigeria: mass, targeted and selective treatment with levamisole. *Parasitology*, 103(2):291–298, 1991.

[12] Yves F Atchade. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in applied Probability*, 8(2):235–254, 2006.

[13] Yves F Atchadé, Jeffrey S Rosenthal, et al. On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[14] Norman TJ Bailey. The mathematical theory of epidemics. Technical report, 1957.

[15] Allison H Baker, Elizabeth R Jessup, and Thomas Manteuffel. A technique for accelerating the convergence of restarted gmres. *SIAM Journal on Matrix Analysis and Applications*, 26(4):962–984, 2005.

[16] Frank Ball. Stochastic and deterministic models for SIS epidemics among a population partitioned into households. *Mathematical Biosciences*, 156(1):41–67, 1999.

[17] Frank Ball and Peter Neal. Network epidemic models with two levels of mixing. *Mathematical biosciences*, 212(1):69–87, 2008.

[18] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, pages 46–89, 1997.

[19] Shweta Bansal, Shashank Khandelwal, and Lauren A Meyers. Exploring biological network structure with clustered random networks. *BMC bioinformatics*, 10(1):405, 2009.

[20] AD Barbour and MJ Luczak. A law of large numbers approximation for markov population processes with countably many types. *Probability Theory and Related Fields*, 153(3-4):727–757, 2012.

[21] Andrew Barbour and Gesine Reinert. Approximating the epidemic curve. *Electronic Journal of Probability*, 18(54):1–30, 2013.

[22] Michael Betancourt. Identifying the optimal integration time in hamiltonian monte carlo. *arXiv preprint arXiv:1601.00225*, 2016.

[23] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

[24] Michael J Betancourt. Generalizing the no-u-turn sampler to riemannian manifolds. *arXiv preprint arXiv:1304.1920*, 2013.

[25] Jeffrey Bethony, Simon Brooker, Marco Albonico, Stefan M Geiger, Alex Loukas, David Diemert, and Peter J Hotez. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet*, 367(9521):1521–1532, 2006.

[26] Alex Bishop, Thomas House, and Keeling Matt. Approximate epidemic dynamics on clustered networks: an effective degree approach. Master's thesis, University of Warwick, 2014.

[27] Christopher M Bishop. Mixture density networks. Technical report, Citeseer, 1994.

[28] Andrew J Black and Joshua V Ross. Estimating a markovian epidemic model using household serial interval data from the early phase of an epidemic. *PloS one*, 8(8):e73420, 2013.

[29] Andrew J Black and Joshua V Ross. Computation of epidemic final size distributions. *Journal of theoretical biology*, 367:159–165, 2015.

[30] Andrew J Black, Thomas House, Matthew James Keeling, and Joshua V Ross. Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza. *Journal of The Royal Society Interface*, 10(81):20121019, 2013.

[31] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112 (518):859–877, 2017.

[32] Moses J Bockarie, Daniel J Tisch, Will Kastens, Neal DE Alexander, Zachary Dimber, Florence Bockarie, Ervin Ibam, Michael P Alpers, and James W Kazura. Mass treatment to eliminate filariasis in papua new guinea. *New England Journal of Medicine*, 347(23):1841–1848, 2002.

[33] Tom Bohman and Michael Picollelli. SIR epidemics on random graphs with a fixed degree sequence. *Random Structures and Algorithms*, 41(2):179–214, 2012.

[34] Ludwig Boltzmann. Ableitung des stefan'schen gesetzes, betreffend die abhängigkeit der wärmestrahlung von der temperatur aus der electromagnetischen lichttheorie. *Annalen der Physik*, 258(6):291–294, 1884.

[35] Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[36] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[37] Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical science*, pages 101–117, 2001.

[38] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[39] Simon Cauchemez, F Carrat, C Viboud, AJ Valleron, and PY Boelle. A bayesian mcmc approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine*, 23(22):3469–3487, 2004.

[40] L Chan, DAP Bundy, and SP Kan. Aggregation and predisposition to ascaris lumbricoides and trichuris trichiura at the familial level. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88(1):46–48, 1994.

[41] Ole F Christensen, Gareth O Roberts, and Jeffrey S Rosenthal. Scaling limits for the transient phase of local metropolis–hastings algorithms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):253–268, 2005.

[42] Damian Clancy and PHILIP D O'NEILL. Exact bayesian inference and model selection for stochastic models of epidemics among a community of households. *Scandinavian Journal of Statistics*, 34(2):259–274, 2007.

[43] Charles D Criscione, Joel D Anderson, Dan Sudimack, Janardan Subedi, Ram P Upadhayay, Bharat Jha, Kimberly D Williams, Sarah Williams-Blangero, and Timothy JC Anderson. Landscape genetics reveals focal transmission of a human macroparasite. *PLoS neglected tropical diseases*, 4(4), 2010.

[44] DWT Crompton. Ascaris and ascariasis. 2001.

[45] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, and Matthew C Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.

[46] Laurent Decreusefond, Jean-Stéphane Dhersin, Pascal Moyal, and Viet Chi Tran. Large graph limit for an SIR process in random network with heterogeneous connectivity. *The Annals of Applied Probability*, 22(2):541–575, 2012.

[47] Charo I. Del Genio and Thomas House. Endemic infections are always possible on regular networks. *Physical Review E*, 88:040801, 2013.

[48] Charo I Del Genio, Hyunju Kim, Zoltán Toroczkai, and Kevin E Bassler. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PloS one*, 5(4):e10012, 2010.

[49] Klaus Dietz. Transmission and control of arbovirus diseases. *Epidemiology*, 104:121, 1975.

[50] Christina Dold and Celia V Holland. Ascaris and ascariasis. *Microbes and infection*, 13(7):632–637, 2011.

[51] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[52] Louise Dyson, Michael Marks, Oliver M Crook, Oliver Sokana, Anthony W Solomon, Alex Bishop, David CW Mabey, and T Déirdre Hollingsworth. Targeted treatment of yaws with household contact tracing: How much do we miss? *American journal of epidemiology*, 187(4):837–844, 2017.

[53] Louise Dyson, Wilma A Stolk, Sam H Farrell, and T Déirdre Hollingsworth. Measuring and modelling the effects of systematic non-adherence to mass drug administration. *Epidemics*, 18:56–66, 2017.

[54] Darin S Evans, Kal Alphonsus, Jon Umaru, Abel Eigege, Emmanuel Miri, Hayward Mafuyai, Carlos Gonzales-Peralta, William Adamani, Elias Pede, Christopher Umbugadu, et al. Status of onchocerciasis transmission after more than a decade of mass drug administration for onchocerciasis and lymphatic filariasis elimination in central nigeria: challenges in coordinating the stop mda decision. *PLoS neglected tropical diseases*, 8(9):e3113, 2014.

[55] Sam H Farrell, James E Truscott, and Roy M Anderson. The importance of patient compliance in repeated rounds of mass drug administration (mda) for the elimination of intestinal helminth transmission. *Parasites & vectors*, 10(1): 291, 2017.

[56] Massimiliano Fasi and Nicholas J Higham. An arbitrary precision scaling and squaring algorithm for the matrix exponential. 2018.

[57] Paul Fearnhead and Dennis Prangle. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.

[58] Paul Fearnhead and Dennis Prangle. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.

[59] David Fincham. Leapfrog rotational algorithms. *Molecular Simulation*, 8(3-5): 165–178, 1992.

[60] Christopher Fitzpatrick, Fiona M Fleming, Matthew Madin-Warburton, Timm Schneider, Filip Meheus, Kingsley Asiedu, Anthony W Solomon, Antonio Montresor, and Gautam Biswas. Benchmarking the cost per person of mass treatment for selected neglected tropical diseases: an approach based on literature review and meta-regression with web-based software application. *PLoS neglected tropical diseases*, 10(12):e0005037, 2016.

[61] JE Forrester, ME Scott, DAP Bundy, and MHN Golden. Clustering of ascaris lumbricoides and trichuris trichiura infections within households. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 82(2):282–288, 1988.

[62] Massimo Galuzzi. *Analytical Mechanics: Translated from the" Mecanique analytique," nouvelle edition of 1811. JL Lagrange, Auguste Boissonnade, Victor N. Vagliente*. Springer, 1998.

[63] Andrew Gelman, Donald B Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[64] Andrew Gelman, Gareth O Roberts, Walter R Gilks, et al. Efficient metropolis jumping rules. *Bayesian statistics*, 5(599-608):42, 1996.

[65] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[66] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

[67] John Geweke, José M Bernardo, James O Berger, A Philip Dawid, and A Smith. Bayesian statistics. *Bayesian statistics*, 1992.

[68] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144, 12 1959.

[69] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[70] Mark Girolami and Michael Betancourt. Hamiltonian monte carlo for hierarchical models. In *Current Trends in Bayesian Methodology with Applications*, pages 116–138. Chapman and Hall/CRC, 2015.

[71] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[72] James P Gleeson. Bond percolation on a class of clustered random networks. *Physical Review E*, 80(3):036107, 2009.

[73] James P Gleeson. Binary-state dynamics on complex networks: pair approximation and beyond. *Physical Review X*, 3(2):021004, 2013.

[74] D.M. Green and I.Z. Kiss. Large-scale properties of clustered networks: Implications for disease dynamics. *Journal of Biological Dynamics*, 4(5):431–445, 2010.

[75] Bryan T Grenfell, K Wilson, VS Isham, HEG Boyd, and K Dietz. Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode–ruminant interactions as a case study. *Parasitology*, 111(S1):S135–S151, 1995.

[76] HL Guyatt, S Brooker, and CA Donnelly. Can prevalence of infection in school-aged children be used as an index for assessing community prevalence? *Parasitology*, 118(3):257–268, 1999.

[77] Heikki Haario, Eero Saksman, Johanna Tamminen, et al. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[78] Isabel Hagel and Tatiana Giusti. Ascaris lumbricoides: an overview of therapeutic targets. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)*, 10(5):349–367, 2010.

[79] A Hall and C Holland. Geographical variation in ascaris lumbricoides fecundity and its implications for helminth control. *Parasitology Today*, 16(12):540–544, 2000.

[80] Andrew Hall. Intestinal helminths of man: the interpretation of egg counts. *Parasitology*, 85(3):605–613, 1982.

[81] Andrew Hall, Kazi Selim Anwar, Andrew Tomkins, and Lutfar Rahman. The distribution of ascaris lumbricoides in human hosts: a study of 1765 people in bangladesh. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(5):503–510, 1999.

[82] Hanan Helmy, Gary J Weil, Abou Sree T Ellethy, Ehab S Ahmed, Maged El Setouhy, and Reda MR Ramzy. Bancroftian filariasis: effect of repeated treatment with diethylcarbamazine and albendazole on microfilaraemia, antigenaemia and antifilarial antibodies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(7):656–662, 2006.

[83] Julian Herbert and Valerie Isham. Stochastic host-parasite interaction models. *Journal of mathematical biology*, 40(4):343–371, 2000.

[84] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[85] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[86] Celia V Holland, SO Asaolu, DWT Crompton, RC Stoddart, R Macdonald, and SEA Torimiro. The epidemiology of ascaris lumbricoides and other soil-transmitted helminths in primary school children from ile-ife, nigeria. *Parasitology*, 99(2):275–285, 1989.

[87] T. House. Generalised network clustering and its dynamical implications. *Advances in Complex Systems*, Vol.13(No.3):281–291, 2010.

[88] T. House and M. J. Keeling. The impact of contact tracing in clustered populations. *PLoS Comput Biol*, 6(3):e1000721, 2010.

[89] Thomas House and Matt J Keeling. Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface*, 8 (54):67–73, 2011.

[90] Thomas House and Matt J Keeling. Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface*, 8 (54):67–73, 2011.

[91] Thomas House, Geoffrey Davies, Leon Danon, and Matt J Keeling. A motif-based approach to network epidemics. *Bulletin of Mathematical Biology*, 71 (7):1693–1706, 2009.

[92] Thomas House, Joshua V Ross, and David Sirl. How big is an outbreak likely to be? methods for epidemic final-size calculation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2150):20120436, 2013.

[93] Michael A Irvine, Wilma A Stolk, Morgan E Smith, Swaminathan Subramanian, Brajendra K Singh, Gary J Weil, Edwin Michael, and T Deirdre Hollingsworth. Effectiveness of a triple-drug regimen for global elimination of lymphatic filariasis: a modelling study. *The Lancet Infectious Diseases*, 17(4): 451–458, 2017.

[94] Svante Janson, Malwina Luczak, and Peter Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures and Algorithms*, 45(4):726–763, 2014.

[95] B. Karrer and M.E.J. Newman. Random graphs containing arbitrary distributions of subgraphs. *Phys. Rev. E*, 82:066118, 2010.

[96] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[97] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867, 1999.

[98] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

[99] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[100] Matt J Keeling, Mark EJ Woolhouse, Darren J Shaw, Louise Matthews, Margo Chase-Topping, Dan T Haydon, Stephen J Cornell, Jens Kappey, John Wilesmith, and Bryan T Grenfell. Dynamics of the 2001 uk foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543): 813–817, 2001.

[101] Matt J. Keeling, Thomas House, Alison J. Cooper, and Lorenzo Pellis. Systematic approximations to susceptible-infectious-susceptible dynamics on networks. *PLOS Computational Biology*, 12(12):e1005296, 2016.

[102] Matthew J Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867, 1999.

[103] Matthew James Keeling and Joshua V Ross. On methods for studying stochastic disease dynamics. *Journal of the Royal Society Interface*, 5(19):171–181, 2007.

[104] Jennifer Keiser and Jürg Utzinger. Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. *Jama*, 299(16):1937–1948, 2008.

[105] WO Kermack and AG McKendrick. Wo kermack and ag mckendrick, proc. r. soc. london, ser. a 115, 700 (1927). *Proc. R. Soc. London, Ser. A*, 115:700, 1927.

[106] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[107] Timothy Kinyanjui, Jo Middleton, Stefan Güttel, Jackie Cassell, Joshua Ross, and Thomas House. Scabies in residential care homes: Modelling, inference and interventions for well-connected population sub-units. *PLoS computational biology*, 14(3):e1006046, 2018.

[108] John G Kirkwood and Elizabeth Monroe Boggs. The radial distribution function in liquids. *The Journal of Chemical Physics*, 10(6):394–402, 1942.

[109] Istvan Z Kiss and Darren M Green. Comment on 'properties of highly clustered networks'. *Physical Review E*, 78(4):048101, 2008.

[110] István Z Kiss, Joel C Miller, and Péter L Simon. *Mathematics of Epidemics on Networks*. Springer, 2017.

[111] Peter E Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

[112] Andrei Kolmogoroff. Über die analytischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(1):415–458, 1931.

[113] Alp Kucukelbir, Rajesh Ranganath, Andrew Gelman, and David Blei. Automatic variational inference in stan. In *Advances in neural information processing systems*, pages 568–576, 2015.

[114] RB Lehoucq, DC Sorensen, and C Yang. Arpack users' guide: Solution of large scale eigenvalue problems with implicitly restarted arnoldi methods. *Software Environ. Tools*, 6, 1997.

[115] Ka Yin Leung, MEE Kretzschmar, and Odo Diekmann. Dynamic concurrent partnership networks incorporating demography. *Theoretical population biology*, 82(3):229–239, 2012.

[116] V. I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707, February 1966.

[117] J. Lindquist, J. Ma, P. van den Driessche, and F.H. Willeboordse. Effective degree network models. *J. Math. Biol.*, 62:143, 2010.

[118] Jennifer Lindquist, Junling Ma, P van den Driessche, and Frederick H Willeboordse. Effective degree network disease models. *Journal of mathematical biology*, 62(2):143–164, 2011.

[119] William A Link and Mitchell J Eaton. On thinning of chains in mcmc. *Methods in ecology and evolution*, 3(1):112–115, 2012.

[120] Ira M Longini Jr, James S Koopman, Arnold S Monto, and John P Fox. Estimating household and community transmission parameters for influenza. *American journal of epidemiology*, 115(5):736–751, 1982.

[121] George Macdonald et al. Symposium on insecticides. ii. the objectives of residual insecticide campaigns. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 46(3):227–34, 1952.

[122] Tristan Marshall and Gareth Roberts. An adaptive approach to langevin mcmc. *Statistics and Computing*, 22(5):1041–1057, 2012.

[123] Robert M May and Roy M Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111, 1984.

[124] Don L McLeish et al. A maximal inequality and dependent strong laws. *The Annals of probability*, 3(5):829–839, 1975.

[125] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[126] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[127] Lauren Ancel Meyers, Babak Pourbohloul, Mark EJ Newman, Danuta M Skowronski, and Robert C Brunham. Network theory and sars: predicting outbreak diversity. *Journal of theoretical biology*, 232(1):71–81, 2005.

[128] E Michael, Bryan T Grenfell, VS Isham, DA Denham, and DAP Bundy. Modelling variability in lymphatic filariasis: macrofilarial dynamics in the brugia pahangi–cat model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1391):155–165, 1998.

[129] J.C. Miller, A.C. Slim, and E.M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*, 9(70): 890–906, 2012.

[130] Joel C Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.

[131] Joel C. Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology*, 62(3):349–358, 2011.

[132] Joel C Miller, Anja C Slim, and Erik M Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, page rsif20110403, 2011.

[133] Joel C. Miller, Anja C. Slim, and Erik M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, 9(70):890–906, 2012.

[134] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.

[135] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[136] A Montresor. Prevention and control of intestinal parasitic infections. *WHO Technical Report Series*, (749), 1987.

[137] A Montresor. Eliminating soil-transmitted helminthiases as a public health problem in children. *World Health Organization*, 2012.

[138] Peter J Neal and Gareth O Roberts. Statistical inference and model selection for the 1861 hagelloch measles epidemic. *Biostatistics*, 5(2):249–261, 2004.

[139] Radford M Neal et al. Slice sampling. *The annals of statistics*, 31(3):705–767, 2003.

[140] Marcel F Neuts and Jian Min Li. An algorithmic study of sir stochastic epidemic models. In *Athens Conference on Applied Probability and Time Series Analysis*, pages 295–306. Springer, 1996.

[141] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

[142] Mark EJ Newman. Properties of highly clustered networks. *Physical Review E*, 68(2):026121, 2003.

[143] M.E.J. Newman. *Networks: an introduction*. Oxford University Press, 2009.

[144] M.E.J. Newman. Random graphs with clustering. *Physical review letters*, 103 (5):058701, 2009.

[145] Peter J Olver. *Introduction to partial differential equations*. Springer, 2014.

[146] World Health Organization. Diagnostic techniques for intestinal parasitic infections (ipi) applicable to primary health care (phc) services. *WHO/P.D.P./85.2*, 1985.

[147] World Health Organization. Fact sheet: Soil-transmitted helminth infections. *See https://www.who.int/en/news-room/fact-sheets/detail/soil-transmitted-helminth-infections (accessed 27 March 2020)*, 2020.

[148] World Health Organization et al. Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation: executive summary. Technical report, Geneva: World Health Organization, 2012.

[149] World Health Organization et al. Transmission assessment surveys in the global programme to eliminate lymphatic filariasis: Who position statement. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, 87(48): 478–482, 2012.

[150] World Health Organization et al. Lymphatic filariasis: Fact sheet. *See http://www. who. int/mediacentre/factsheets/fs102/en/(accessed 13 March 2017)*, 2017.

[151] World Health Organization et al. Guideline: alternative mass drug administration regimens to eliminate lymphatic filariasis. Technical report, World Health Organization, 2017.

[152] World Health Organization et al. Summary of global update on preventive chemotherapy implementation in 2016: crossing the billion. *Wkly Epidemiol Rec*, 92(40):589–608, 2017.

[153] World Health Organization et al. Validation of elimination of lymphatic filariasis as a public health problem. 2017.

[154] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[155] Omiros Papaspiliopoulos, Gareth O Roberts, and Martin Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, pages 59–73, 2007.

[156] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *arXiv preprint arXiv:1408.2701*, 2014.

[157] Lorenzo Pellis, Thomas House, and Matt J. Keeling. Exact and approximate moment closures for non-Markovian network epidemics. *Journal of Theoretical Biology*, 382:160–177, 2015.

[158] Martyn Plummer et al. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124. Vienna, Austria., 2003.

[159] Reda MR Ramzy, Maged el Setouhy, Hanan Helmy, Amr M Kandil, Ehab S Ahmed, Hoda A Farid, Rifky Faris, and Gary J Weil. The impact of single-dose diethylcarbamazine treatment of bancroftian filariasis in a low-endemicity setting in egypt. *The American journal of tropical medicine and hygiene*, 67 (2):196–200, 2002.

[160] DA Rand. Correlation equations and pair approximations for spatial ecologies. *Advanced ecological theory: principles and applications*, 100, 1999.

[161] David Rand. *Advanced ecological theory: principles and applications*, chapter Correlation equations and pair approximations for spatial ecologies, pages 100–142. John Wiley & Sons, 2009.

[162] Ramakrishna U Rao, Kumara C Nagodavithana, Sandhya D Samarasekera, Asha D Wijegunawardana, Welmillage DY Premakumara, Samudrika N Perera, Sunil Settinayake, J Phillip Miller, and Gary J Weil. A comprehensive assessment of lymphatic filariasis in sri lanka six years after cessation of mass drug administration. *PLoS neglected tropical diseases*, 8(11):e3281, 2014.

[163] Martin Ritchie, Luc Berthouze, Thomas House, and Istvan Z Kiss. Higher-order structure and epidemic dynamics in clustered networks. *Journal of theoretical biology*, 348:21–32, 2014.

[164] Martin Ritchie, Luc Berthouze, and Istvan Z Kiss. Beyond clustering: Mean-field dynamics on networks with arbitrary subgraph composition. *arXiv preprint arXiv:1405.6234*, 2014.

[165] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[166] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[167] Gareth O Roberts and Jeffrey S Rosenthal. Coupling and ergodicity of adaptive markov chain monte carlo algorithms. *Journal of applied probability*, 44(2): 458–475, 2007.

[168] Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[169] Gareth O Roberts, Richard L Tweedie, et al. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[170] Gareth O Roberts, Jeffrey S Rosenthal, et al. Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367, 2001.

[171] Tim Rogers. Maximum-entropy moment-closure for stochastic systems on networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(05): P05007, 2011.

[172] Jeffrey S Rosenthal et al. Optimal proposal distributions and adaptive mcmc. *Handbook of Markov Chain Monte Carlo*, 4(10.1201), 2011.

[173] Joshua V Ross, Thomas House, and Matt J Keeling. Calculation of disease dynamics in a population of households. *PLoS One*, 5(3):e9666, 2010.

[174] Carl Runge. Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, 1895.

[175] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.

[176] Fred Luciano Neves Santos, Elúzio José Lima Cerqueira, and Neci Matos Soares. Comparison of the thick smear and kato-katz techniques for diagnosis of intestinal helminth infections. *Revista da Sociedade Brasileira de Medicina Tropical*, 38(2):196–198, 2005.

[177] Thomas Sellke. On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20(2):390–394, 1983.

[178] M. Ángeles Serrano and Marián Boguñá. Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.*, 97:088701, Aug 2006.

[179] Zhang Shaoqing, Cheng Feng, and Roger Webber. A successful control programme for lymphatic filariasis in hubei, china. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88(5):510–512, 1994.

[180] P.L. Simon, M. Taylor, and I.Z. Kiss. Exact epidemic models on graphs using graph automorphism driven lumping. *J. Math. Biol.*, 62(479-508), 2010.

[181] David Spiegelhalter, Andrew Thomas, Nicky Best, and Wally Gilks. Bugs 0.5: Bayesian inference using gibbs sampling manual (version ii). *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, pages 1–59, 1996.

[182] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

[183] Simon Tavaré, David J Balding, Robert C Griffiths, and Peter Donnelly. Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518, 1997.

[184] Mark J Taylor, Achim Hoerauf, and Moses Bockarie. Lymphatic filariasis and onchocerciasis. *The Lancet*, 376(9747):1175–1185, 2010.

[185] Michael Taylor, Péter L Simon, Darren M Green, Thomas House, and Istvan Z Kiss. From markovian to pairwise epidemic models and the performance of moment closure approximations. *Journal of mathematical biology*, 64(6):1021–1042, 2012.

[186] The Theano Development Team, Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas, Frédéric Bastien, Justin Bayer, Anatoly Belikov, et al. Theano: A python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, 2016.

[187] Edward K Thomsen, Nelly Sanuku, Manasseh Baea, Samson Satofan, Elit Maki, Bart Lombore, Mark S Schmidt, Peter M Siba, Gary J Weil, James W Kazura, et al. Efficacy, safety, and pharmacokinetics of coadministered diethylcarbamazine, albendazole, and ivermectin for treatment of bancroftian filariasis. *Clinical Infectious Diseases*, 62(3):334–341, 2015.

[188] Daniel J Tisch, Edwin Michael, and James W Kazura. Mass chemotherapy options to control lymphatic filariasis: a systematic review. *The Lancet infectious diseases*, 5(8):514–523, 2005.

[189] James Truscott, T Déirdre Hollingsworth, and Roy Anderson. Modeling the interruption of the transmission of soil-transmitted helminths by repeated mass chemotherapy of school-age children. *PLoS neglected tropical diseases*, 8(12): e3323, 2014.

[190] James E Truscott, T Déirdre Hollingsworth, Simon J Brooker, and Roy M Anderson. Can chemotherapy alone eliminate the transmission of soil transmitted helminths? *Parasites & vectors*, 7(1):266, 2014.

[191] Hugo C Turner, Alison A Bettis, Brian K Chu, Deborah A McFarland, Pamela J Hooper, Eric A Ottesen, and Mark H Bradley. The health and economic benefits of the global programme to eliminate lymphatic filariasis (2000–2014). *Infectious diseases of poverty*, 5(1):54, 2016.

[192] Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza. Analytical computation of the epidemic threshold on temporal networks. *Physical Review X*, 5(2):021005, 2015.

[193] Henk A Van der Vorst. Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems. *SIAM Journal on scientific and Statistical Computing*, 13(2):631–644, 1992.

[194] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[195] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2015.

[196] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432, 2017.

[197] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc. *arXiv preprint arXiv:1903.08008*, 2019.

[198] E.M. Volz, J.C. Miller, A. Galvani, and L. Ancel-Meyers. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput Biol*, 7(6):e1002042, 2011.

[199] Erik M Volz. SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56(3):293–310, 2008.

[200] Martin Walker, Andrew Hall, and María-Gloria Basáñez. Individual predisposition, household clustering and risk factors for human infection with ascaris lumbricoides: new epidemiological insights. *PLoS neglected tropical diseases*, 5 (4):e1047, 2011.

[201] Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25. Cambridge University Press, 2009.

[202] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.

[203] Gary J Weil, Will Kastens, Melinda Susapu, Sandra J Laney, Steven A Williams, Christopher L King, James W Kazura, and Moses J Bockarie. The impact of repeated rounds of mass drug administration with diethylcarbamazine plus albendazole on bancroftian filariasis in papua new guinea. *PLoS Neglected Tropical Diseases*, 2(12):e344, 2008.

[204] "World Health Organisation". Global programme to eliminate lymphatic filariasis: progress report, 2017. *Weekly epidemiological record*, 91(44):589–604, 2018.

[205] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Yes, but did it work?: Evaluating variational inference. *arXiv preprint arXiv:1802.02538*, 2018.

[206] Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman, et al. Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3):917–1003, 2018.