Parameter Estimation of Epidemic Models

Ian Michael Adamson
St. Mary's College
Durham University

A report submitted for the degree of
MMath

Easter 2025

Abstract

There exists a symbiotic relationship between the realms of applied mathematics and statistical inference when trying to understand the intricacies of infectious diseases and their impacts on society. Often represented via dynamical systems, the models used to describe epidemics can vary greatly in complexity to explore a range of features that could appear under different scenarios and with different pathogens. Many of these differences heavily rely on the parameters of the system, which change depending on the characteristics of the disease to external factors like population movement and environmental conditions. Making informative predictions about a disease from these models requires implementing statistical methods to incorporate both data and areas of uncertainty to then predict the potential behaviors that may occur as a result of different conditions and levels of knowledge. While the potential features of a system can be heuristically explored through dynamical systems and bifurcation analysis, ideas in Bayesian statistics help link the real world to the theoretical, providing an outline of how to turn fabricated computer models into real-world estimation and prediction tools. Even with their setbacks due to ongoing developments, using statistics to explore beliefs within dynamical systems still remains essential to understanding the progression of infectious diseases, allowing for more informative measures to be instituted to mitigate the severity of an emerging epidemic.

Contents

Statement of Originality

This piece of work is a result of my own work and I have complied with the Department's guidance on multiple submission and on the use of AI tools. Material from the work of others not involved in the project has been acknowledged, quotations and paraphrases suitably indicated, and all uses of AI tools have been declared.

Chapter 1

Introduction

The field of mathematical epidemiology has significantly developed over the past century. What began with Daniel Bernoulli developing methods to determine the efficacy of smallpox interventions in 1760 has now transformed into complex, predictive modeling systems used to track spatial and temporal developments of a disease to inform public health policy and response [3]. Technological developments along with advancements in modeling and data analysis techniques have made epidemic models integral to understanding the progression and mitigation of infectious diseases [54]. At the heart of developments in epidemic modeling is the task of parameter estimation, which connects dynamical systems analysis and applied mathematics with statistical inference and real-world data. The advancements in both fields intertwine well, giving a framework to more accurately and effectively model epidemic outcomes and inform about adequate response.

1.1 Introduction to Epidemiological Modeling

Modern epidemic modeling finds its origins with William O. Kermack and A. G. McKendrick, who in 1927 first introduced the modern Susceptible-Infected-Recovered (SIR) model [34]. The introduction of the SIR model pioneered the use of compartment models to govern disease transmission dynamics, representing individuals as belonging to one of a finite number of distinct states, with the ability to move from one state to another at specified rates [7]. The SIR model comprises three states, represented by the compartments S for susceptible individuals, I for infected individuals, and R for recovered individuals. Individuals may move from S to I via the transmission rate β and from I to R via the recovery rate γ, as shown in Figure 1.1. Viewing diseases as compartment models allows the transmission dynamics to be governed by systems of ordinary differential equations, which for the SIR model take the form

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \quad \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I, \tag{1.1}$$

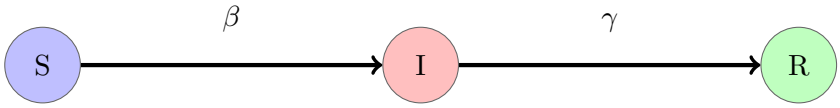


Figure 1.1: Compartment diagram for an SIR model.

which provides a mathematical foundation for understanding them [34].

Along with the implementation of the compartment-based framework came the introduction of threshold theory in epidemiology. Kermack and McKendrick solidified the idea that it takes a certain number of infected individuals in a population for a disease to spread, and hence there exist specific thresholds determining the expected progression of a disease [34]. This allowed for the explicit quantification of the transmissibility of a pathogen and the conditions required to control it and prevent widespread outbreaks, giving rise to quantities such as the basic reproduction number R_0. Defined as the average number of onward infections occurring from an infected individual, R_0 directly translate to this idea of a critical threshold for a given disease, since the barrier $R_0 = 1$ indicates the point where a disease may change from the expectation of persisting in the population to having a greater likelihood of dying out [2]. Representing the dynamics in this way also allows easy extensions beyond the SIR model, including incorporating latent classes, cyclic natures, multiple strains, co-infections, and a variety of other features [37].

As the twentieth century progressed, the downfalls of deterministic models became evident. Although they provide valuable information on transmission dynamics, they lack the ability to capture the inherent randomness of transmission due to movement of individuals and alterations in immunity as a result of spatial, dietary, or medical changes, for example. To solve this problem, deterministic compartment models were combined with probability theory to develop stochastic epidemic models in the mid-twentieth century [3]. The rise of branching models and Markov chain models helped to incorporate chance into epidemic dynamics as well as random fluctuations in the transitions between disease states. This marked a key shift in more accurately tracking transmission patterns, transitioning from simply understanding the relationships between states to also quantifying the variability and unpredictability within the models based on data.

In the late twentieth century and early twenty-first century another layer of the epidemiological problem developed as the role of contact patterns within populations became a key question. Specifically, network-based models emerged that represented individuals as nodes that interact with other nodes through edges in a network [66]. Having a key role in the research behind super-spreaders, which are highly connected individuals who disproportionately drive transmission, the use of network models incorporated the spatial element

along with the stochasticity [56]. The researchers could then target interventions based on how people live in their daily lives, such as who they interact with and the different places they go that could expose them to a pathogen [54]. Similarly, this development allowed for more structured research on the targeting of vaccines toward individuals with the highest contacts or the greatest ability to spread a disease, which further highlights the role of social structure on disease spread. Being able to integrate spatial and social data allowed for more realistic representations of disease spread in real-world populations.

Combining these advancements in modeling techniques permits the development of more complex systems that represent a wider variety of possible behaviors seen in modern epidemics. Of particular interest is the extension towards multi-strain diseases, as many prevalent pathogens in the modern world have multiple variants, each with their own unique characteristics and impact on future infections, both for the infected individual and those susceptible [53]. For example, with the COVID-19 pandemic, the evolution of the pathogen into different strains inevitably resulted in faster dissipation of immunity periods as one could be infected with a newer strain before becoming fully susceptible to the original strain after infection [41, 53]. Hence, knowledge of disease transmission, societal structure, and social behaviors became imperative to combating the epidemic, allowing for methods like contact tracing and targeted vaccination schemes to take place to help efficiently calm the severity of the pandemic [76]. As complexity of the disease structure increases, so to does the range of possible dynamics, implying the necessity to combine tools together to better understand and be able to reasonably control them.

1.2 History of Parameter Estimation

As epidemic modeling has evolved over the last century, the ability to accurately calibrate models and generate realistic predictions based on current data has become a key focus. Using the dynamical systems approach to epidemic modeling pioneered by Kermack and McKendrick and others, this calibration presents itself through parameter estimation, which introduces an interplay between mathematical theory and statistical methods that has become integral to mathematical biological studies, pushing theoretical models into reality.

With the introduction of Kermack and McKendrick's threshold theory came the determination of parameter estimates via heuristic models and analytical solutions. This meant being able to calculate values such as the basic reproduction number for the SIR model, given by $R_0 = \frac{\beta}{\gamma}$, directly from the model equations, which relied heavily on assumptions of homogeneity and equilibrium [3]. These early models assume even mixing of the population between those infected and those susceptible so that anyone can be infected at the same rate from the start. While this may occur to some extent as an epidemic progresses, in

its early stages the presence of a disease is so small that transmission occurs as an almost purely stochastic event, meaning that pulling information regarding critical thresholds directly from the equations without considering real data and disease progression may hinder the ability to give sensible output [6]. The inability of these assumptions to incorporate what occurs in the world inhibits accurate modeling of the parameter values and predicting disease trajectories.

As stochastic models came into play in the mid twentieth century, so to did the rise of statistical methods for determining parameter values directly from data. Methods like maximum likelihood estimation and least squares fitting determined the estimates as the minimal distances between the observed data and predicted values of the system [60]. While they were the primary introduction to incorporating data like case counts into the estimation process, they still remained limited in their scope, requiring many simplifying assumptions about the disease dynamics and data quality to obtain reasonable estimates [32]. These methods also do not capture well temporal dependencies when working with time-series data as they still capture the parameter estimates as fixed quantities over the whole time period of interest [75].

The advent of Bayesian inference in the late 20th century marked a significant advancement in parameter estimation. Unlike frequentist methods, which treat parameters as fixed quantities, Bayesian approaches incorporate prior knowledge and quantify uncertainty through posterior distributions [22]. This framework proved invaluable for epidemic modeling, where data are often sparse or noisy. Techniques such as Markov Chain Monte Carlo (MCMC) sampling enabled researchers to explore complex parameter spaces and generate probabilistic forecasts based on updating prior beliefs as data are seen.

Recently, developments in machine learning and data assimilation techniques have allowed for real-time updates on estimates and for more dynamic data to be used to produce them [22]. These advancements have been vital in helping make decisions in recent outbreaks, such as with the COVID-19 pandemic [18]. However, even with this progress in creating more interpretable, realistic estimates, there still remain limitations in extracting the most useful ones. Particularly with global infection data, data collection can be inconsistent across different parts of the world, which may result in increased bias and sparsity, making it harder to extract informative estimates from the data [62]. This increases the difficulty in being able to adequately quantify uncertainty with respect to the models used and the estimates made, particularly when the variations in data collection result in poor data quality [11]. Hence, even if the mathematical and statistical tools have been developed to perform the necessary calculations and provide estimates, the main hindrance with obtaining useful estimates lies heavily with the data, both in amount and quality.

1.3 Parameter Estimation of Epidemic Models

The developments in epidemic modeling coincide with those in parameter estimation: as models increase in complexity, parameter estimation methods must evolve accordingly to ensure accuracy and reliability of fitted trajectories. From the SIR model to modern-day network-based approaches, each advancement requires new techniques in parameter estimation to enable the calibration of models to real-world data, the increased understanding of the uncertainties within them, and the acquisition of reasonable insight into their dynamics. As these models increase in complexity, incorporating more potential variants and transmission pathways, they begin to exhibit richer dynamics that must be explored. This can be done theoretically at great length, with even more complex systems being created to test the potential dynamics based on a variety of initial conditions and parameters, but they only prove useful if they can be brought into the real world by fitting them to data and estimating their parameters. Only then can the true plausibility of different parameter regimes be confirmed, thus determining the realism of these theoretical explorations into the mathematical nuances of these dynamical systems.

Hence, it rests to investigate this connection between theory and reality through the context of multi-strain disease models, exploring the tools that can be used to understanding their behaviors and confirm the theoretical understanding of them. From their mathematical properties to techniques to adequately predict those behaviors as well as quantifying the uncertainties about them, the ability to comprehend the paths these diseases may take under different case scenarios will permit the justification as to the realism of these models theoretical, helping to promote the necessity to explore these hypothetical cases before driving them into reality.

Chapter 2

Introduction to Disease Parameter Estimation

Understanding the complex dynamics of multi-strain diseases requires an intuition about the models representing them and the impact of their parameter values on the model behaviors. Investigating this in the context of simpler, single-strain diseases provides a useful basis for expansion into the multi-strain realm, which often combines multiple single-strain systems together to depict more complex behaviors. Exploring the foundational concepts in the single-strain world setting allows for the introduction of key mathematical ideas that may occur in reality, as well as how well these models combine with data to give informative estimates about the real world.

2.1 Beyond the SIR Model

The original SIR model pioneered by Kermack and McKendrick forms a solid foundation for compartmental disease models, helping to form the intuition for this perspective of disease modeling. However, it merely acts as an introduction, lacking features that represent real, modern-day pathogens, including both transmission dynamics and mathematical behaviors. To effectively build the foundational mathematical understanding and the impact on disease parameters on model behaviors, the SIR model will be expanded, adding small nuances that allow for different useful features in the extension towards higher complexity.

One natural extension of the SIR model is the SEIR model, which improves on the initial framework by incorporating a latent "Exposed" state, where individuals have been infected but are not infectious. While the SIR model implies that becoming infected results in immediately being able to transmit the disease, this extra class acts as a time-delay, where someone may transmit the pathogen only after an incubation period post-initial infection [65]. For example, SEIR-type models have been used to depict overall COVID-19 transmission in Indonesia, where incorporating the incubation period of 2 to 14 days allows

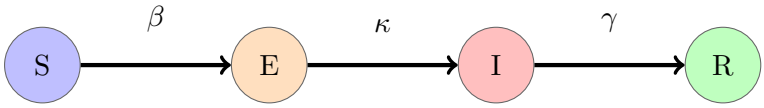


Figure 2.1: Compartment diagram for an SEIR model.

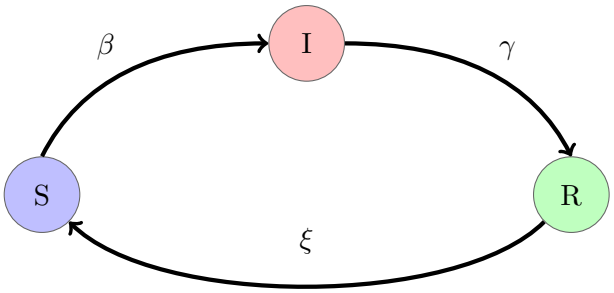


Figure 2.2: Compartment diagram for an SIRS model.

for better prediction of infection peaks [30]. In addition to the SIR model rates, the exposed to infectious rate κ describes the movement from the infected state to the infectious [25]. The compartment model describing SEIR dynamics in Figure 2.1 follows a similar structure to the SIR model, with just the addition of the "E" state and κ. Describing this model mathematically results in the set of equations given by [25]

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \quad \frac{dE}{dt} = \frac{\beta SI}{N} - \kappa E, \quad \frac{dI}{dt} = \kappa E - \gamma I, \quad \frac{dR}{dt} = \gamma I. \tag{2.1}$$

In addition to adding a time delay into the infectiousness of a disease, one may also wish to consider possible cyclic behaviors due to periods of waning immunity and the ability to be re-infected with a pathogen. Represented in its simplest form as the SIRS model, this extension retains the same three-class configuration as Kermack and McKendrick's SIR model while introducing the ability to be reinfected with a disease post-recovery. This takes form with the addition of the waning immunity rate α, depicting the rate at which fully recovered individuals become re-susceptible and thus may be infected with the disease again, described by the equations [38]

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \alpha R, \quad \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I - \alpha R. \tag{2.2}$$

Both of these models build upon the SIR model by incorporating additional features often seen in modern infectious diseases. These aid in being able to understand more complex diseases as they are systems used as building blocks for more complex, even multi-strain models, so understanding their features and possible trajectories proves important when trying to understand more intricate systems.

2.1.1 Model Analysis - Equilibria and R_0

An essential part to the mathematical understanding of systems of ordinary differential equations is identifying the points of the system where interesting behaviors may occur. In particular, identifying critical thresholds that alter the stability of the model, and hence the long-term forecast of the behavior of the system, gives significant reason to understanding the importance of knowing the parameter values given real data. Determining these points begins with analyzing the steady states of the system and then determining the conditions for their stability [47].

The SEIR model, for example, has one disease-free equilibrium (DFE), indicating that the only steady state will arise when there is no infection present, defined by

$$(S, E, I, R) = (N, 0, 0, 0).$$

Similarly, the SIRS model also has a DFE indicating no infection and is of a similar form to that of the SEIR model, as well as an endemic equilibrium, where the disease has enough presence to persist in the population, given by [16]

$$(S, I, R) = (\frac{\gamma N}{\beta}, I^*, \frac{\gamma}{\alpha} I^*),\ I^* > 0.$$

Examining the output of Figure 2.3, which provides some simulated trajectories for the the SEIR and SIRS models, the only possible long-term behavior of the SEIR model is similar to the simulated trajectory, where eventually all individuals either become infected and then recover or remain susceptible. A situation like the SEIR model is possible for the SIRS about the DFE, but the model also possesses the capability of some proportion of the population always remaining infected due to its cyclic structure. The existence of multiple equilibria of the SIRS model indicates the possibility for varying stability of the system, where under a certain parameter regime the disease will always die out and the system tends towards the DFE, and under another the disease becomes endemic to the population, with just enough of an infected population to remain everpresent [64]. Therefore, there exists some point where the stability of the system changes from tending to the DFE to tending towards the endemic equilibrium. As a result, the following derivations will focus on the SIRS model as it provides more interesting behaviors that can be extended into the more complex, multi-strain realm.

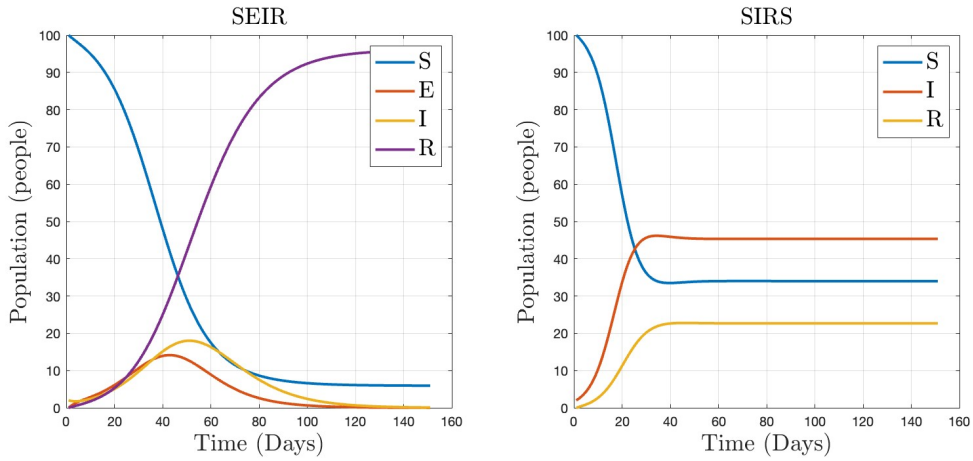


Figure 2.3: Simulated output for the SEIR model (left) and the SIRS model (right).

Following usual linear stability analysis procedures for the SIRS model, computing the Jacobian matrix for the system gives [47]

$$J = \begin{bmatrix} \frac{\partial \dot{S}}{\partial S} & \frac{\partial \dot{S}}{\partial I} & \frac{\partial \dot{S}}{\partial R} \\ \frac{\partial \dot{I}}{\partial S} & \frac{\partial \dot{I}}{\partial I} & \frac{\partial \dot{I}}{\partial R} \\ \frac{\partial \dot{R}}{\partial S} & \frac{\partial \dot{R}}{\partial I} & \frac{\partial \dot{R}}{\partial R} \end{bmatrix} = \begin{bmatrix} -\frac{\beta I}{N} & -\frac{\beta S}{N} & \alpha \\ \frac{\beta I}{N} & \frac{\beta S}{N} - \gamma & 0 \\ 0 & \gamma & -\alpha \end{bmatrix}.$$

Performing linear stability analysis allows for the computation of the expression for the basic reproduction number R_0. Intuitively, R_0 defines the threshold where if a greater proportion of individuals become infected than recover or die, then the disease is more likely to persist and become endemic to the population, described by the condition $R_0 > 1$ [9]. Conversely, $R_0 < 1$ implies the disease is more likely to die out. Mathematically, R_0 is the threshold where the largest eigenvalue of the Jacobian evaluated at the DFE transitions from negative to positive real part [40]. This is because at any steady state is stable if all eigenvalues are non-positive and unstable if at least one eigenvalue is positive, meaning this eigenvalue changing the stability of the system results in a change in the model behavior about the DFE [8]. Hence, evaluating the Jacobian at the DFE gives

$$J = \begin{bmatrix} 0 & -\beta & \alpha \\ 0 & \beta - \gamma & 0 \\ 0 & \gamma & -\alpha \end{bmatrix}.$$

Then, solving the characteristics equation $\det(J - \lambda I) = 0$ for λ, where I is the identity

matrix, gives the eigenvalues of the system [44]:

$$\begin{aligned}
\det(J-\lambda I)=0 &\implies \begin{vmatrix} -\lambda & -\beta & \alpha \\ 0 & \beta-\gamma-\lambda & 0 \\ 0 & \gamma & -\alpha-\lambda \end{vmatrix} = 0, \\
&\implies -\lambda \begin{vmatrix} \beta-\gamma-\lambda & 0 \\ \gamma & -\alpha-\lambda \end{vmatrix} = 0, \\
&\implies -\lambda[(\beta-\gamma-\lambda)(-\alpha-\lambda)-0]=0, \\
&\iff \lambda=0 \text{ or } (\beta-\gamma-\lambda)(-\alpha-\lambda)=0.
\end{aligned}$$

The system is stable when all the eigenvalues have non-positive real part, and unstable when there exists at least one eigenvalue with positive real part [47]. Hence, the critical threshold of interest arises as the eigenvalue λ that can become positive or negative depending on the values of the parameters of which it is comprised. Focusing on when $\lambda \neq 0$,

$$\begin{aligned}
&\implies (\beta-\gamma-\lambda)(-\alpha-\lambda)=0, \\
&\iff \lambda=-\alpha \text{ or } \lambda=\beta-\gamma.
\end{aligned}$$

R_0 is then determined by the condition required for the largest eigenvalue to satisfy instability. Namely, $\lambda=-\alpha$ will always be negative, but $\lambda=\beta-\gamma$ can be positive or negative depending on the relative size of β to γ. Hence, to satisfy instability,

$$\begin{aligned}
&\lambda=\beta-\gamma>0, \\
&\iff \beta>\gamma, \\
&\iff \frac{\beta}{\gamma}>1, \\
&\implies R_0=\frac{\beta}{\gamma}.
\end{aligned}$$

For simple single-strain diseases, the expression for R_0 becomes quite intuitive - if disease transmission occurs at a faster rate than recovery, the disease persists; else, the disease diminishes. While R_0 has this interpretation for diseases in general, including multi-strain models, the expression derived from these simple models provides a clear understanding as to the implications of the parameter values with regards to this critical threshold.

2.1.2 Parameter Behaviors

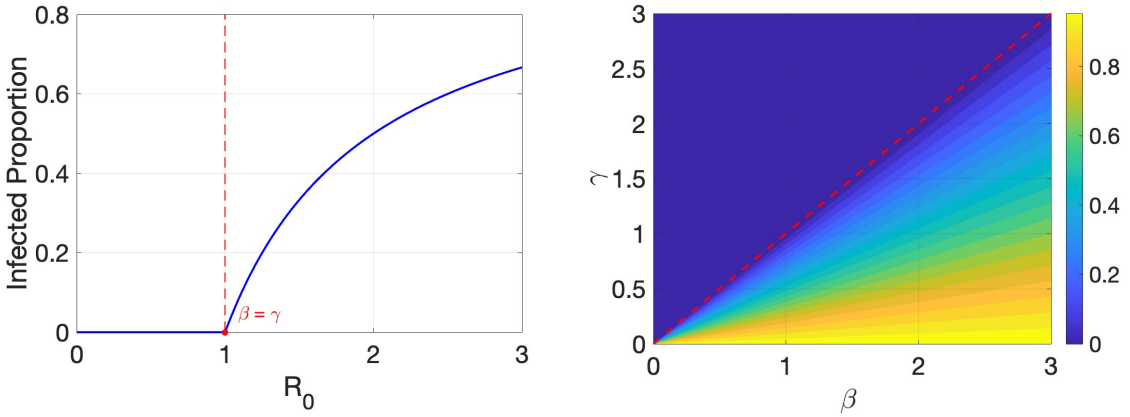
Since R_0 describes the largest eigenvalue of the system at the DFE of the SIRS model, $R_0>1$ implies the expectation that the disease will persist and the model will tend to the endemic equilibrium, whereas $R_0<1$ signifies that the disease will be more likely to die out as there are not relatively enough infections for the disease to remain present in the population. Given the stability of the system changes at $R_0=1$, corresponding to

when $\beta = \gamma$ for the SIRS model, β, and γ can be interpreted as bifurcation parameters. The $R_0 = 1$ threshold means the system will tend to the asymptotically stable DFE for $\beta < \gamma$ and towards the endemic equilibrium for $\beta > \gamma$. Varying β relative to γ over an array of β and γ values, Figure 2.4 visualizes the transcritial, or forward, bifurcation at $\beta = \gamma$, where the DFE and endemic equilibria exchange stability [8]. When $\beta < \gamma$, $R_0 < 1$ and the proportion of infected individuals in the population tends to the DFE, indicating that if there were any disease present, it becomes more likely than not that it would die out when the transmission rate lives in this region. However, as β increases relative to γ, the number of individuals being infected supersedes the number able to recover in the same time, resulting in the trajectory of the model shifting to tend towards the endemic equilibrium, indicated by the shift in stability of the two equilibria.

While the SEIR model has the same form for R_0, it is important to highlight that the interpretation is slightly different, due to the SEIR model not having an endemic equilibrium. This model does not possess any bifurcation points in the same way as in the SIRS model, as the existences of only disease-free equilibria imply that asymptotically the disease will die out either by there being no infections, or by everyone being infected and recovering. Regardless, the presence of the bifurcation point in the SIRS model emphasizes the necessity of obtaining estimates for the model parameter values. Depending on the model parameters, specifically their values relative to other parameters, the model trajectory can look completely different. A slight alteration in the parameters values can be the difference between expecting the disease to being endemic to a population and the expectation of eradication. Hence, understanding the values which the parameters take can aid in determining how to combat and control a disease, giving a quantitative, informative measure on expected behavior based on the known information.

2.2 The Time-Dependency Question

Usual introductory dynamical systems analysis often assumes fixed parameter values over time. This assumptions allows for the understanding of how different parameter values change the trajectory of the system and for what values interesting behaviors, like bifurcations, occur. However, this time-independence assumption becomes less useful when bringing models out of computer simulations and into the real world, as the parameter values may change as conditions vary, requiring a degree of time sensitivity to accurately model the disease progression. Taking the SEIR model for example, to fit the parameter values β, γ, and κ to some data would require setting each to a fixed value via a statistic over the whole time frame, such as the median or the mean. As conditions changes, be with host behaviors, weather, or anything else that alters a pathogen's ability to survive,



(a) (b)

Figure 2.4: Bifurcation plots for the SIRS model. Figure 2.4a gives how the infected equilibrium (as a proportion of the population) changes with R_0 values, and Figure 2.4b gives how the infected equilibrium changes as both β and γ changes as individual bifurcation parameters, with the gradient being the proportion of infected individuals for each parameter regime. Simulations were performed in MATLAB for an array of β and γ values.

the expected trajectory of a disease also change [42]. Treating each parameter as fixed essentially means assuming these quantities to be mathematically defined and not dependent on external forces [51]. Hence, while it may be useful to look at fixed values to see rough behaviors, it is not practical to use this assumption when working in real time as data are being seen. It becomes difficult to accurately and effectively estimate the trajectory of a disease and how to mitigate it if the temporal changes that affect the disease's transmission within the population are not considered.

Incorporating this time dependency not only requires a change in the estimation techniques, but also some changes to the model to reflect this time dependency. Any parameter being estimated must be a function of time to observe its changes due to other factors. For example, β becomes $\beta(t)$ to allow measuring of disease transmission at each time step t. For the SEIR model in Equation 2.1, adopting this assumption change results in the model, reflecting that of a stochastic differential equation, where each state variable is also a function of time to allow for real-time data collection. This makes tracking the temporal changes of the model easier, as the data may be applied for any discrete time point t, and then the parameter values can be estimated from that given the values of the model at the previous time step. For example, when estimating the transmission rate β, now $\beta(t)$, the SEIR model would become

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta(t)S(t)I(t)}{N}, & \frac{dE(t)}{dt} &= \frac{\beta(t)S(t)I(t)}{N} - \kappa E(t), \\ \frac{dI(t)}{dt} &= \kappa E(t) - \gamma I(t), & \frac{dR(t)}{dt} &= \gamma I(t). \end{aligned} \tag{2.3}$$

Disease incidence data are often recorded in regular, discrete time intervals, so adopting this change permits observing changes in the model dynamics over the entire discrete time space [67]. This in turn makes it easier to derive the quantities to assess the severity of a disease and the potential for any interesting behaviors as it progresses over time.

2.2.1 The Effective Reproduction Number, R_t

Understanding disease progression requires the ability to capture on average how many people an infected individual infects. Typically, R_0 summarizes this as the number of onward infections from an infected individual assuming all other individuals of the population are susceptible [6]. While useful in discussing the overall severity of a pathogen, it relies on the underlying assumption that every individual of the population, save for the one infected individual of interest, is susceptible [68]. To track the average number of onward infections at any given point in time, it is thus important to consider what proportion of the population could become infected at each time point. Thus, R_0 is traded in for the effective reproduction number R_t, which measures the average number of secondary infections from each primary infection at each time step t relative to the number of susceptible individuals [63]. R_0 is a dimensionless quantity, often used to characterize the average overall disease spread, whereas R_t gives a time-dependent expression for the average spread based on the proportion of the population that could be infected at time t [6]. This allows for better representation of the temporal variation in the disease trajectory, where $R_t < 1$ indicates that the disease is under control and is expected to decline and possibly die out if that trajectory follows [51]. R_t is calculated from the expression for R_0, with appropriate scaling to account for the proportion of the population an infected individual could infect at that time point, given by the susceptible proportion at that time, resulting in the expression

$$R_t = R_0 \frac{S_t}{N}, \tag{2.4}$$

where S_t is the total susceptible population at time t [51]. Including R_t permits the tracking of the impact of a disease at each time point, giving a quantifiable threshold to indicate the expected behavior of the system based on the average number of onward infections relative to the number of susceptible individuals in the population.

2.3 Estimating Time-Dependent Disease Parameters via Monte Carlo Methods

To narrow down the possible methods to estimate the disease model parameters, it is important to summarize some key considerations with regards to the data being used and the

types of estimates wished to be determined. Working with identifying temporal behaviors of model trajectories using time-series data implies a possible inherent correlation due to the time-dependency, which eliminates the efficacy of certain methods. Approaches relying on least-squares regression, for example, cannot generate usable estimates as there is often the assumption of independence of residuals as well as an inability to handle the temporal dependencies of the data [5]. Additionally, it is important to consider what data is being seen, as often incidence data only includes observed transmissions, which are those tracked from actual disease testing [23, 67]. Those who fall ill but do not get tested or report their illness would not be counted in these numbers, which adds some uncertainty to the true number of infected individuals at any given time. Hence, the ability to effectively model the disease dynamics and estimate parameter values rests on being able to specify beliefs on the parameter values and update them as data are seen, an idea that lies at the heart of Bayesian statistics and Monte Carlo methods.

2.3.1 Introduction to Markov Chain Monte Carlo

Effectively and efficiently incorporating the noise and uncertainties related to epidemic models requires methods that allow for the addition of random fluctuations and the ability to use this to modify what are believed to be plausible values for the parameters. Methods based on Markov Chain Monte Carlo achieve this by treating the uncertainties as unknown quantities and updating the beliefs about their distributions as more data are seen and more samples are taken [14]. In general, Monte Carlo methods are based on estimating distributional properties through random samples from that distribution [69]. If data are known to follow a distribution but there is uncertainty about from which distribution they come, random samples can be generated from which the parameters of interest can be derived. Sampling enough times produces a range of possible parameter values which can be used to give plausible behaviors of the disease models.

The primary idea behind Monte Carlo methods rests with Monte Carlo integration. Developed as a method for generating expectations from samples, it can be extended to generate average parameter estimates given some data [24]. Specifically, to approximately evaluate $E[f(X)]$ for some function f and random variable X, samples $\{X_t\}$ are taken from the distribution of X, $3\pi(.)$, from which an empirical expectation can be computed [24]

$$E[f(X)] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t). \tag{2.5}$$

In reality, sampling from $\pi(.)$ directly can be difficult as the true distribution is often unknown. To solve this, view the generation of samples as a Markov Chain, which is a sequence of random variables such that each sample only depends on the previous one [24].

This dependence only on the previous state of the chain is known as the Markov property and allows the samples to be continuously updated as the they are taken to better reflect the true distribution from which the data and parameters arise [68]. Producing estimates for the random variable does not require direct sampling from $\pi(.)$, but only random draws from the support of the distribution. This is because after a sufficiently long calibration period, known as the "burn-in" period, the Markov property allows later samples to be dependent on samples drawn from the true stationary distribution (i.e. the true distribution of the parameter), meaning the samples inevitably come from the required distribution [24]. A "chain" of samples parameter values, $\Theta = \{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, ...\}$, then arises, where $\theta^{(i)}$ is the vector of parameter values for the i-th sample, which allows the generation of plausible parameter estimates for the model.

Combining both of these ideas results in the general setup of Markov Chain Monte Carlo (MCMC). Based on Bayesian statistical ideas, the method relies on applying Bayes' Rule to update beliefs. Suppose that the prior beliefs suggest that the data from the $(i-1)$-st sample come from a (possibly joint) prior distribution with probability density function $p(\theta)$ with (possibly joint) likelihood $f(\mathbf{x}|\theta)$. Using these, along with the proposed values of $\theta^{(i)}$ allows for the update of beliefs about the true parameter values using Bayes' Rule to approximate the posterior distribution, which follows the rule

$$p(\theta|\mathbf{x}) \propto p(\theta) f(\mathbf{x}|\theta), \tag{2.6}$$

which essentially means that the posterior distribution for the parameter θ given data $\mathbf{x}$ is some multiplicative combination of the prior and likelihood distributions [20]. To obtain the posterior distribution using this relationship, MCMC iteratively refines parameter estimates by simulating random samples from the observed data to then update beliefs based on the estimates being produced.

2.3.2 The Metropolis-Hastings MCMC Algorithm

One of the main advantages of MCMC is that generating samples and updating beliefs also determines the plausibility of sampled values compared to the previous ones, with beliefs being updated accordingly. A common method for accepting and rejecting the proposed values of the chain is through the Metropolis Hastings acceptance probability [17],

$$\alpha(\theta^{*(i)}|\theta^{(i-1)}) = \min\left\{1, \frac{\hat{f}(\mathbf{x}|\theta^{*(i)}) q(\theta^{(i-1)}|\theta^{*(i)})}{\hat{f}(\mathbf{x}|\theta^{(i-1)}) q(\theta^{*(i)})|\theta^{(i-1)})}\right\}. \tag{2.7}$$

This means that as a new value for the Markov Chain is proposed, the proposed value becomes the new value of the chain with probability α, or it remains at the same value as the previous step with probability $1-\alpha$, helping the Markov chain values stay along the

required distributions, resulting in reasonable estimates for the parameters being output [57]. With all of this in mind, Algorithm 1 briefly describes the MCMC algorithm that will be the baseline for later algorithms [69].

Algorithm 1 Metropolis-Hastings MCMC Algorithm

Input: $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)}, \xi^{(0)})$, where $\beta \sim p(\beta)$, $\gamma \sim p(\gamma)$, ...
for $i \in \{1, ..., T_{max}\}$ **do**
  Propose: $\beta^* \sim p(\beta^{(i-1)})$, $\gamma^* \sim p(\gamma^{(i-1)})$, ...
  Likelihood: $f(\beta^*, \gamma^*, ...)$
  Acceptance Probability: $\alpha(\theta^{*(i)}|\theta^{(i-1)}) = \min\left\{1, \frac{f(\beta^*,\gamma^*,...)}{f(\beta^{(i-1)},\gamma^{(i-1)},...)}\right\}$
  Generate: $u \sim \text{Uniform}(0, 1)$
  if $u < \alpha$ **then**
    $\theta^{(i)} \leftarrow (\beta^*, \gamma^*, ...)$
  else
    $\theta^{(i)} \leftarrow \theta^{(i-1)}$
  end if
end for

This gives a rough sampling framework with to determine parameter estimates. However, there still exist a few areas of improvement to effectively fit the parameters to disease incidence data. While the general MCMC algorithm does output plausible parameter values from the chain, it does not incorporate the temporal nature of disease incidence data and the hidden states given by observation error. Following in the Bayesian philosophy, there exist frameworks that allow for the incorporation of this observation error over time, which can then be used as information from which to derive the parameter estimates.

2.3.3 Hidden Markov Models

Estimating the disease parameters requires knowing not just the observed incidence at each time step, but rather the true incidence. Given the discrepancy between the two in the data, one can say there exists a dependence between the observed incidence and a hidden, or unobserved, state indicated by the true incidence. The existence of these dependent observations results in a setup known as a Hidden Markov Model (HMM), where incorporating the additional uncertainty from the hidden states requires assuming the existence of a Markov structure regarding those observations [24]. This means that the missing data from the observation error is treated as a state in the Markov chain, allowing for a more realistic view of the total transmissibility of a disease to be considered than just fitting the parameters based on observed incidence. As shown in Figure 2.5, at each time step t, the total number of infections of a disease is given by Z_t; however, due to the observation error resulting from incomplete measures of the true number of infected individuals, only

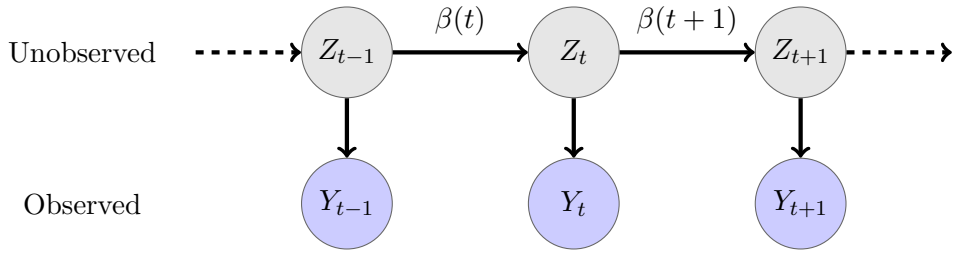


Figure 2.5: Directed graph describing a Hidden Markov Model. At each time step t, the total infected is given by Z_t, but only Y_t gets observed due to observation error, represented by some noise. The transition from time t to $t+1$ is represented by an increase in the number of infected by transmission rate $\beta(t)$ and a decrease by recovery rate γ.

Y_t is observed as data. To understand the true transmissibility, it is necessary to estimate Z_t as $\beta(t)$ comes from the change in the true infected numbers from time $t-1$ to time t, not the observed number. The addition of the time dependence allows models to adapt to real-world changes, such as the development of mitigation strategies. Hence, the data must be treated with extra care when constructing a model and sampling method to use to incorporate these details and obtain the most realistic estimates possible given the data seen.

2.3.4 Particle Filtering via Sequential Monte Carlo (SMC)

To capture estimates for the total number of infected individuals given observation error, consider more flexible sampling approaches that can be used to track and estimate these over time, and not just from random sample to random sample. Sequential Monte Carlo (SMC) is a sampling method that achieves this by performing step-wise sampling in an computationally efficient manner [4, 17]. Used as a sub-algorithm in the Particle Markov Chain Monte Carlo algorithm to follow, SMC captures more of the additional information not given from standard Monte Carlo integration simulations. Suppose state y is observed, which can be viewed as a random variable depending on a hidden state x that is unobserved but whose distribution can be sampled from. Sampling from the distribution of x means obtaining the marginal likelihood $p(y_{1:T}|\theta)$. The time-dependence of x implies a strong dependency on subsequent values of x over time, meaning sampling directly from the whole set $\{x_{1:T}\}$ would be inefficient and possibly quite inaccurate.

To overcome this issue, sample step-wise for each time step x_t, which mimics the time dependency of successive values of x_t, via importance sampling. Sequential importance resampling ensures that the samples generated follow the distribution of the parameters of interest [17]. This allows samples to be generated from the distribution from which the data come, $\pi(.)$, by taking samples from a proposal distribution, $\tilde{\pi}(.)$, and then determining

the weight of each sampled point $w_i = \frac{\pi(x_i)}{\tilde{\pi}(x_i)}$. To then ensure the samples follow the true distribution of the data as best as possible, resample these points with replacement, but using the w_i as probabilities for different points occurring to result in a higher guarantee that a sample following the required distribution is produced [14]. Once sampling is completed for each time step and $x_{1:T}$ produced, the marginal likelihood can be calculated as

$$p(y_{1:T}|\theta) = \int p(y_{1:T}|x_{1:T}, \theta)p(x_{1:T}|\theta)dx_{1:T}. \tag{2.8}$$

Using this idea, Algorithm 2 describes the steps to perform SMC [14], giving an estimation of the incidence trajectory as well as the parameter values for the system for each time step.

Algorithm 2 Sequential Monte Carlo Algorithm

Input: number of particles N, initial parameter values θ, number of time steps T, observed data y_t
for $t = 1$ **do**
  for $i \in \{1, ..., N\}$ **do**
    $X_1^{(i)} \sim p(X_1|\theta)$
    initial marginal likelihood: $g(y_1|X_1^{(i)}, \theta)$
  end for
  Resample the particles using weights $w_1^{(i)} = \frac{g(y_1|X_1^{(i)},\theta)}{\hat{p}(y_1|\theta)}$
end for
for $t \in \{2, ...T\}$ **do**
  for $i \in \{1, ..., N\}$ **do**
    Propagate: $X_t^{(i)} \sim f_\theta(X_{t-1}^{(i)}|\theta)$
    Marginal likelihood: $g(y_t|X_t^{(i)}, \theta)$
  end for
  Resample the particles using weights $w_t^{(i)} = \frac{g(y_t|X_t^{(i)},\theta)}{\hat{p}(y_t|\theta)}$
end for

2.3.5 Particle Markov Chain Monte Carlo

With the addition of SMC allowing the step-wise calculation of parameter estimates for each sample to encapsulate the uncertainties in both parameter values and the data, combine this along with the MCMC algorithm to obtain the Particle Markov Chain Monte Carlo (PMCMC) algorithm. The algorithm proves useful as it captures the sampling variability in the estimates given by MCMC as well as the temporal uncertainties obtained by the SMC algorithm, incorporating two key factors not encapsulated together earlier. Algorithm 3 combines the two algorithms together, providing a full framework with which to perform PMCMC [17].

Algorithm 3 Particle MCMC Algorithm

Input: $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)}, ...)$
Step 1: Perform the SMC algorithm with $\theta^{(0)}$ to generate an initial sample $\{X_{1:T}\}$ and initial marginal likelihood $\hat{p}(y_{1:T}|\theta^{(0)})$
Step 2: Randomly select one trajectory to be the initial sample trajectory, $x_{1:T}^{(0)}$
for $i \geq 1$ **do**
Step 3: Propose $\theta^{*(i)} \sim q(\theta^{*(i)}; \theta^{(i-1)})$
Step 4: Perform the SMC algorithm with $\theta^{*(i)}$ to generate an initial sample $\{X_{1:T}^{*}\}$ and initial marginal likelihood $\hat{p}(y_{1:T}|\theta^{*(i)})$
Step 5: Randomly select one trajectory to be the i-th sample trajectory, $x_{1:T}^{*(i)}$
Step 6: Compute the Metropolis-Hastings acceptance probability

$$\alpha(\theta^{*(i)}|\theta^{(i-1)}) = \min\left(1, \frac{\hat{f}(x|\theta^{*(i)}q(\theta^{(i-1)}|\theta^{*(i)}}{\hat{f}(x|\theta^{(i-1)})q(\theta^{*(i)}|\theta^{(i-1)})}\right)$$

if $u \sim \text{Uniform}(0,1) < \alpha$ **then**
$\theta^{(i)} = \theta^{*(i)}$
else
$\theta^{(i)} = \theta^{(i-1)}$
end if
end for

2.3.6 Implementing the PMCMC Algorithm

There are a variety of ways to efficiently implement this algorithm. One way is by using the `Rbi` and `Rbi.helpers` packages in R, which together implement the LibBi software [48]. LibBi allows for efficiently sampling from the model while incorporating the extra uncertainties due to observation error and time-depdency. In general, the process for performing PMCMC through LibBi is as follows [17]:

1. When defining the model, set the prior specifications for the parameters based on any previous knowledge available, as well as to allow for adequate scanning of the parameter space to propose a range of plausible value.

2. Calibrate the model by taking 1000 PMCMC samples.

3. In order to have low log-likelihood variance and to accept a reasonable number of particles at each step, applying the `adapt_particles` and `adapt_proposals` functions help tune the model, reducing the chance of obtaining unreasonable estimates.

4. Take 10,000 samples, using the first 5,000 as a "burn-in" period to ensure the samples are well-mixed and representative of the data. From the other 5,000 thin every 5, one sample from every five is randomly selected to keep, resulting in 1,000 posterior samples from the model.

2.3.7 Forecasting PMCMC Outputs

Once time-dependent estimates of the parameter values have been generated along with their bands of uncertainty, they can be used to predict what may occur in the future. Based on the tools previously used, the posterior samples act as the input and produce 1000 PMCMC samples to give estimates for the projection period. This provides a rough estimate of how the disease could progress in the short-term based on the information the samples from the observed data.

2.3.8 A Note on Prior Specification

At the heart of Bayesian inference lies the ability to specify beliefs on parameter distributions and update them as data are seen. All MCMC methods rely on being able to make prior assumptions about the parameter distributions, either based on knowledge from similar systems or due to the lack of such knowledge [49]. Selecting appropriate prior distributions for the model is a crucial part of Bayesian statistical analysis, as they act as the primary influence on how the data inform the parameter estimates. This can be a difficult task, particularly when knowledge of the parameter space is limited, or there is belief of variation from normal patterns seen previously. When prior information is limited, vague, or noninformative, priors are used to sweep the parameter space for a range of plausible values before being iteratively tuned to fit the data and the model more precisely [39, 50]. This then allows for generating the Markov chain based on samples from representative distributions to the beliefs about the parameters.

Packages associated with LibBi make it easy to specify vague priors and then meld them into something sensible with respect to the data provided. The R package `rbi.helpers` includes the function `adapt_proposals`, which adjusts any prior specification made to the data to ensure they are useful for sampling [48]. By taking samples from the proposal distributions and observing the Metropolis-Hastings acceptances, the function then determines the empirical standard deviation based on the acceptance of proposed values. The values that set the proposal distributions are then amended to reduce this variance. The process repeats until there is an adequate balance between having an appropriately low acceptance rate, which ensures that only reasonable values are accepted relative to the current state, and being able to explore all potential future states given the current state [43]. With this, the proposals for most of the parameters are set to uniform distributions, as once samples are compared to the data using `adopt_proposals` they will be adapted to reasonable specifications to better fit the model.

2.4 Parameter Estimation of the Dureau Model

Prior to implementing the PMCMC algorithm to complex, multi-strain disease, it is useful to examine an example on a simpler, real-world case where the algorithm has successfully fit a model to data and can be then used for further analysis on the disease trajectory. The data collected and used in previous studies depict weekly infection counts for the 2009 Pandemic Influenza outbreak in the UK, which saw a wide array of transmission behaviors surrounding school activities [15]. Specifically, the outbreak saw higher transmission rates during the academic year, with transmission decaying drastically over the holidays before starting to resurge as schools returned to session but before the start of vaccination programs [31]. Likely a result of lower interaction between children outside of the classroom, the spatial changes mark a key context required to understand the disease trajectory. While the rough behavior of the illness represents that of an SEIR model, namely the data being recorded as one strain with a latent period before individuals become infectious and then after recovering do not encounter waning immunity, the data exhibit two distinct waves of infections which is often difficult to capture via the typical SEIR model [15]. Hence, to estimate the transmissibility over time, $\beta(t)$, employing a modified version of this model to account for more variations in transmission to permits viewing these peaks and accounting for the observed data, giving the Dureau model as [17]

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta S(t)I(t)}{N}, & \frac{dZ}{dt} &= \kappa E(t) - Z(t \mod 7),\\
\frac{dE}{dt} &= \frac{\beta S(t)I(t)}{N} - \kappa E(t), & \frac{dx}{dt} &= \sigma dW,\\
\frac{dI}{dt} &= \kappa E(t) - \gamma I(t), & \beta(t) &= \exp(x(t)),\\
\frac{dR}{dt} &= \gamma I(t).
\end{aligned} \tag{2.9}$$

From the weekly data used in previous studies like [17, 15], the algorithm is run over the 37-week time span the data cover. From the 1,000 posterior samples produced, there are a few plots to describe the behaviors seen in the data, as well as estimates for the transmissibility rate $\beta(t)$. Figure 2.6 provides plots for total incidence, transmissibility, and R_0, where the mean values of the samples for each quantity are given in black, along with 50% and 95% confidence intervals. While the SEIR model forms the basis of the Dureau model, the incidence plot evidently does not match the dynamics of the output in Figure 2.3. With the addition of the additional terms, the extension on the SEIR model given by the Dureau model helps capture the natural flow of transmissions due to the societal and seasonal factors for which the general model does not account.

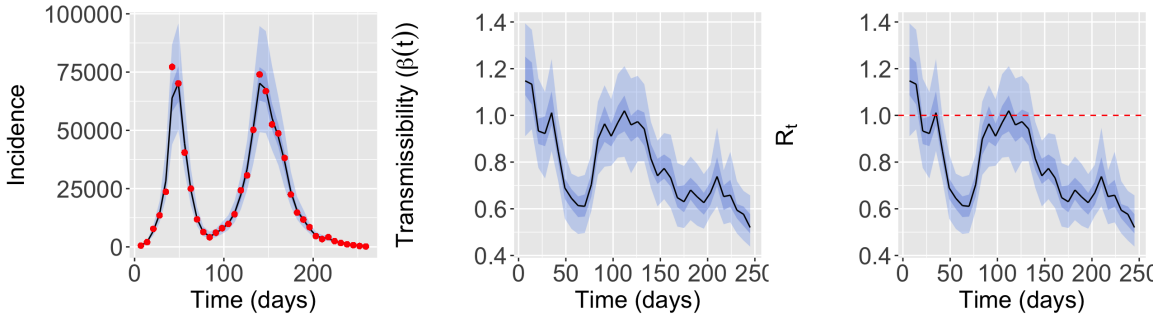


Figure 2.6: Plots for the posterior samples of PMCMC for the 2009 UK Pandemic Influenza data, including incidence vs. time (left) along with the observed incidence counts in red, $\beta(t)$ vs. time (center), and R_t vs. time (right). The left and center plots are recreations of those developed in [17].

Not only does this sampling method capture the rough dynamics seen by the data, but important to note is the behavior of R_t. At just before 50 days (i.e. the seventh time point), R_t drops below one as the transmissibility declines near the end of the first wave of infections. As has been examined previously, this lull is likely a result of the summer holiday period where contact is lower which reduces the chances of mass contact infections [15]. Around week 13, the number of cases begin to increase again as the holiday period ends, resulting in the second strain of infections. However, a significant change from the holiday lull occurs at the end of the second wave, around week 31, for one week there is a slight uptick in incidence numbers, similar to what was seen as the start of the second wave. Moreover, there did not occur a significant drop in transmissibility at the end of the second wave like at the end of the first, so it would rest to investigate the possibility of these factors combining to create a third wave. Forecasting 4 weeks into the future, Figure 2.7 shows that while the average transmission rate for the forecasting period does align with the true values, there is the possibility of increased transmission from that point, which would result in a third wave of infections. Despite this possibility, however, R_t is still expected to decline, indicating that the infections that would occur relative to the proportion of susceptible individuals becomes low enough that there never becomes a risk of uncontrollability of the virus.

This example clearly emphasizes the need for adequate parameter estimation methods that account for uncertainties within the natural world. Modeling the transmission rate as a fixed parameter misses key characteristics that drive the different waves and the controllability of the disease. Omitting consideration of the differences between the simulations and the real world, including both spatial and temporal factors that impact transmission, risks the loss of key behaviors that help guide analysis on epidemic progression, possibly providing misinformation when determining the best way to control it.

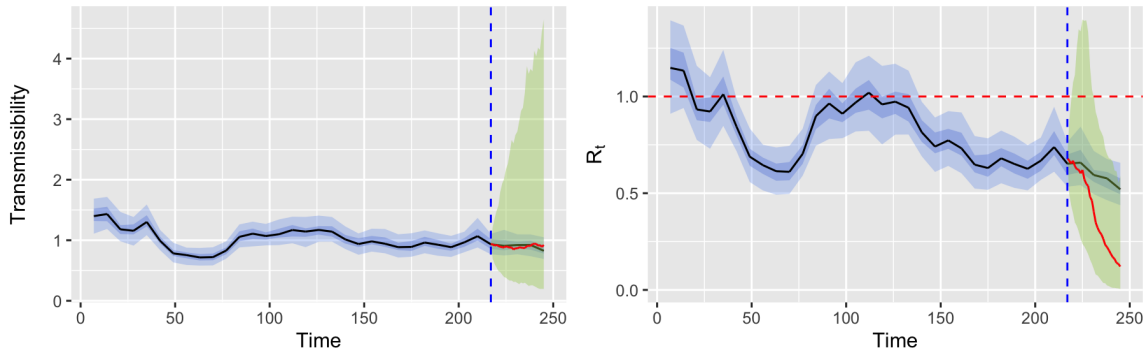


Figure 2.7: Plot for the transmissibility (left) and R_t (left), including the forecasted region from week 31 to 35, compared against the true samples taken.

2.5 A Step Towards Higher Complexity

Regardless of the model being investigated, be it a simple single-strain pathogen or a complex, multi-variant epidemic with high-level features like cross and co-infection, the basic principles remain the same. Estimating the temporal nature of a disease requires adaptations in both the model and the statistical techniques to effectively capture the changes over time. Additionally, observation error plagues the data of any disease due to lack of testing, refusal towards going to the doctor when ill, or even just asymptomatic cases making it more challenging to know for certain the true number of infected individuals. Since the challenges faced with the single strain diseases remain with more complex systems, the estimation methods used act as a foundation for which to expand upon and adapt when working with other, more complicated epidemics. By exploring the characteristics of these diseases, the potential for more complex features increases, giving rise to more uncertainties with which to incorporate into the models to then estimate the parameters of interest.

Chapter 3

Multi-Strain Epidemic Models

Extending the theory of parameter estimation into the multi-strain setting requires understanding the dynamics of multi-strain disease models, considering both similarities and differences to the single-strain case. Establishing this framework will be in the context of a model used to study the theoretical complexities of Dengue Fever transmission, but the types of dynamics can be seen in models for a number of multi-strain diseases, including Influenza and COVID-19 [1]. This model incorporates some unique features not often exhibited by single-strain models, providing a set up for the necessary considerations when trying to adapt parameter estimation techniques to the more complex setting.

3.1 The Dengue Model

As explained in [1], the model of interest is one involving two strains of which an individual could be infected. Once an individual is infected with one strain, they gain full immunity from it, but only have cross-immunity to the other strain for a limited time period. That individual may then be infected with the second strain, after which he or she is fully recovered. All individuals regardless of class are subject to demography, the natural changes in population size over time due to instances of births and deaths [1]. This is represented as a proportion of the population in each class leaving or entering the class in a balanced way.

3.1.1 Parameters and Variables

This model setup has three main classes - susceptible, infected, and recovered. Within these, there are subclasses to differentiate between strains and overall susceptible/recovered status. As with single-strain systems, the set of model parameters describe how an individual may transition between states, including transmission rates for each strain, the recovery rate, waning immunity rate, and even the demography rate. All of these are summarized in Table

Table 3.1: Multi-Strain Model Variables

Susceptible	
S	Susceptible to 1st infection
S_1	Susceptible to strain 2 after strain 1 infection
S_2	Susceptible to strain 1 after strain 2 infection
Infected	
I_1	1st infection is strain 1
I_2	1st infection is strain 2
I_{12}	2nd infection is strain 2
I_{21}	2nd infection is strain 1
Recovered	
R_1	Recovered from strain 1 primary infection
R_2	Recovered from strain 2 primary infection
R	Recovered from secondary infection of either strain

Table 3.2: Multi-Strain Model Parameters

β_1	Transmission rate, strain 1
β_2	Transmission rate, strain 2
ϕ_1	Ratio of contribution of force of strain 1 infection
ϕ_2	Ratio of contribution of force of strain 2 infection
μ	Demography rate
γ	Recovery rate
α	Waning immunity rate

3.1 for the variables and Table 3.2 for the parameters. The total population size, N, is the sum of all the variables in Table 3.1. Comparing this to the simpler models presented earlier, important alterations are the additions to allow for individual dynamics for each variant to have an impact on overall infections for both strains. Each strain possesses its own transmission rate, and any secondary infection has an additional force of infection which helps to describe whether secondary infections are stronger or weaker than primary when infecting the susceptible population. Additionally, the addition of demography incorporates another realistic property of population dynamics, allowing the model to better represent reality, and thus be more useful when making predictions.

3.1.2 The Model

Defining the transition dynamics for each of the state variables relative to the parameters results in a 10-dimensional system whose theoretical properties have previously been explored in depth, defined by [1, 35]

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta_1}{N}S(I_1+\phi_1 I_{21}) - \frac{\beta_2}{N}S(I_2+\phi_2 I_{12}) + \mu(N-S),\\
\frac{dI_1}{dt} &= \frac{\beta_1}{N}S(I_1+\phi_1 I_{21}) - (\gamma+\mu)I_1,\\
\frac{dI_2}{dt} &= \frac{\beta_2}{N}S(I_2+\phi_2 I_{12}) - (\gamma+\mu)I_2,\\
\frac{dR_1}{dt} &= \gamma I_1 - (\alpha+\mu)R_1,\\
\frac{dR_2}{dt} &= \gamma I_2 - (\alpha+\mu)R_2,\\
\frac{dS_1}{dt} &= -\frac{\beta_2}{N}S_1(I_2+\phi_2 I_{12}) + \alpha R_1 - \mu S_1,\\
\frac{dS_2}{dt} &= -\frac{\beta_1}{N}S_2(I_1+\phi_1 I_{21}) + \alpha R_2 - \mu S_2,\\
\frac{dI_{12}}{dt} &= \frac{\beta_2}{N}S_1(I_2+\phi_2 I_{12}) - (\gamma+\mu)I_{12},\\
\frac{I_{21}}{dt} &= \frac{\beta_1}{N}S_2(I_1+\phi_1 I_{21}) - (\gamma+\mu)I_{21},\\
\frac{dR}{dt} &= \gamma(I_{21}+I_{12}) - \mu R.
\end{aligned}
\tag{3.1}
$$

A full compartment diagram for this system is given by Figure 3.1. Using the initial conditions S=80, $I_1 = I_2$=10, and all other compartments are 0, Figure 3.2 shows simulation plots similar to those in [35]. Focusing on the last 3000 years the model follows a somewhat oscillatory pattern in all compartments. However, these oscillations do not resemble that of periodic behavior as there does not appear to be an even repetition of peaks and valleys along the plots. The amplitudes of the curves change with each oscillation, indicating some potentially complex behaviors in the parameter regime.

Given the complexities of this model, computing the equilibria analytically can prove to be quite cumbersome without making significant assumptions to simplify the model. Regardless, there always exists at least one disease-free equilibrium that remains true for the whole parameter space:

$$(\hat{S}, \hat{I_1}, \hat{I_2}, \hat{R_1}, \hat{R_2}, \hat{S_1}, \hat{S_2}, \hat{I_{12}}, \hat{I_{21}}, \hat{R}) = (N, 0, 0, 0, 0, 0, 0, 0, 0, 0).$$

3.2 Model Analysis - The Basic Reproduction Number R_0

As with the single strain models, the main focus is the effect of the parameter values on the model trajectory. Different parameters have a greater effect than others when it comes to determining the extent of disease control. Hence, looking for quantities like R_0 helps summarize the expected behavior of the disease with respect to the equilibria of the system as a combination of the relevant parameters to describe that behavior. Determining the

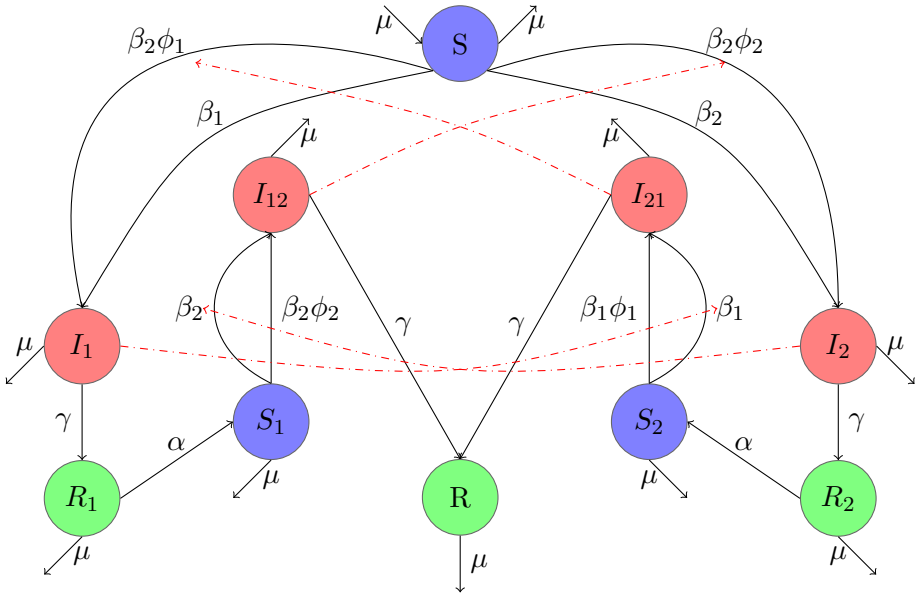


Figure 3.1: A compartment diagram of the multi-strain Dengue model. Red Dashed lines indicate those compartments that influence certain reactions but do not necessarily change as a result of them.

structure of R_0 via linear stability analysis and the eigenvalues of the Jacobian proves to be quite cumbersome, as even the two-strain Dengue model results in computing the determinant of a 10x10 matrix. Additionally, the inability to determine the other equilibria other than the DFE previously mentioned without making significant simplifying assumptions makes it more difficult to explore the changes in stability and what that looks like for a particular disease as complexity increases [7]. Therefore, it is necessary to turn to a method that only requires the DFE above and guarantees the calculation of R_0.

3.2.1 The Next Generation Matrix

In order to assess the model in an easily interpretable way, it is often useful to consider it in terms of the Basic Reproduction Number, R_0. This number describes how many individuals that one infected person infects on average, so the higher the R_0 value, the the more likely the disease is to become endemic, and the lower the R_0 value, the more likely it is for the disease to die out, or tend to the disease-free equilibrium [13]. There exist a variety of methods to determine the basic reproduction number, including a standard Jacaboian analysis. This method for multi-strain systems can prove quite cumbersome, as with the Dengue model for example, which has 10x10 Jacobian Matrix. However, the Next Generation Matrix method can be employed to obtain the same information about disease

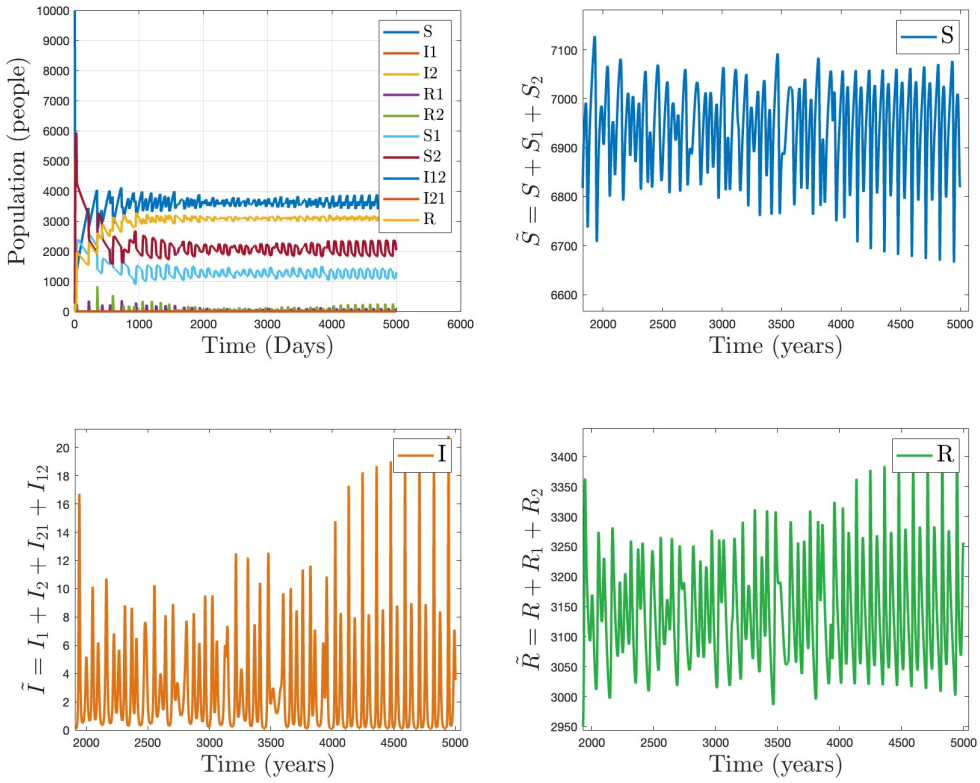


Figure 3.2: Simulated output for the Dengue model, with $S = 10,000$ and $I_1 = I_2 = I_{21} = I_{12} = 20$, with all other compartments $= 0$, and assuming parameter values $\beta_1 = 104$, $\beta_2 = 120$, $\mu = 1/65$, $\gamma = 52$, $\alpha = 2$, $\phi_1 = \phi_2 = 0.6$.

transmission and local stability of the system [7].

Oftentimes, the Basic Reproduction Number can be seen as a way to consider the model in terms of the generational effects of the number of infected individuals at each time step, where the infected classes directly affect the trajectories of the other model states as they are projected forward through time [13]. So, the Next Generation Matrix (NGM) is the matrix that describes the generational relationship between infected classes of the model [13]. Since both R_0 and the NGM only depend on the change in the infected classes, only the infectious subsystem of the model, which is the system of equations from the model involving changes in the infected classes, needs to be considered. Additionally, a benefit to using the NGM method is it only requires the DFE, so it proves useful with complex systems where the endemic equilibrium may be complicated to determine explicitly. R_0 is thus the dominant eigenvalue of the NGM [7].

Determining R_0 via this method follows a straightforward procedure: suppose the Jacobian of the infectious subsystem can be decomposed as $\mathbf{J}=\mathbf{F}\text{-}\mathbf{V}$, where $\mathbf{F}$ is a matrix containing the terms of each equation involving anything relating to new infections, and $\mathbf{V}$ is a matrix containing all other terms [7]. Then, the Next Generation Matrix is $\mathbf{K}=\mathbf{F}\mathbf{V}^{-1}$. To extract R_0 from this, take the dominant eigenvalue of $\mathbf{K}$, which from [13] is denoted by

$$R_0 = \rho(\mathbf{F}\mathbf{V}^{-1}). \tag{3.2}$$

Performing a full Jacobian analysis on the 10-dimensional system with respect to the disease-free equilibrium would obtain the same value of R_0 so both methods provide the same information, but the NGM method streamlines the process [68]. Moreover, unlike Jacobian analysis the NGM method guarantees that the value found about the DFE is R_0 [74]. Hence, the NGM method simplifies the process of determining this critical threshold, reducing the complexity of the problem and eliminating any uncertainties with potential multiple solutions.

3.2.2 Computing R_0 from the Disease-Free Equilibrium

Working with the Dengue Model, start by extracting the infectious subsystem, which is any equation in the system that involves the infected individuals [7]. For the Dengue Model, the infectious subsystem is

$$\begin{aligned}
\frac{dI_1}{dt} &= \frac{\beta_1}{N} S(I_1 + \phi_1 I_{21}) - (\gamma + \mu) I_1 \\
\frac{dI_2}{dt} &= \frac{\beta_2}{N} S(I_2 + \phi I_{12}) - (\gamma + \mu) I_2, \\
\frac{dI_{12}}{dt} &= \frac{\beta_2}{N} S_1(I_2 + \phi_2 I_{12}) - (\gamma + \mu) I_{12}, \\
\frac{dI_{21}}{dt} &= \frac{\beta_1}{N} S_2(I_1 + \phi_1 I_{21}) - (\gamma + \mu) I_{21}.
\end{aligned} \tag{3.3}$$

Next, linearize by computing the Jacobian of the system around the disease-free equilibrium in the same way as in standard linear stability analysis via the Jacobian, resulting in [7, 19]

$$J = \begin{bmatrix} \frac{\beta_1}{N}S - (\gamma+\mu) & 0 & 0 & \frac{\beta_1}{N}S\phi \\ 0 & \frac{\beta_2}{N}S - (\gamma+\mu) & \frac{\beta_2}{N}S\phi & 0 \\ 0 & \frac{\beta_2}{N}S_1 & \frac{\beta_2}{N}S_1\phi_2 - (\gamma+\mu) & 0 \\ \frac{\beta_1}{N}S_2 & 0 & 0 & \frac{\beta_1}{N}S_2\phi_1 - (\gamma+\mu) \end{bmatrix}_{(N,0,0,0,0,0,0,0,0,0)},$$

$$= \begin{bmatrix} \beta_1 - (\gamma+\mu) & 0 & 0 & \beta_1\phi \\ 0 & \beta_2 - (\gamma+\mu) & \beta_2\phi & 0 \\ 0 & 0 & -(\gamma+\mu) & 0 \\ 0 & 0 & 0 & -(\gamma+\mu) \end{bmatrix}.$$

Then, separate this matrix into two matrices, $\mathbf{F}$ and $\mathbf{V}$, where $\mathbf{F}$ is the matrix relating to anything to do with new infections in any compartment, and $\mathbf{V}$ is related to everything else, giving [7]

$$J = \underbrace{\begin{bmatrix} \beta_1 & 0 & 0 & \beta_1\phi \\ 0 & \beta_2 & \beta_2\phi & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{F}} - \underbrace{\begin{bmatrix} (\gamma+\mu) & 0 & 0 & 0 \\ 0 & (\gamma+\mu) & 0 & 0 \\ 0 & 0 & (\gamma+\mu) & 0 \\ 0 & 0 & 0 & (\gamma+\mu) \end{bmatrix}}_{\mathbf{V}},$$
$$= \mathbf{F} - \mathbf{V}.$$

The Next Generation Matrix can then be extracted as

$$\mathbf{FV}^{-1} = \begin{bmatrix} \frac{\beta_1}{\gamma+\mu} & 0 & 0 & 0 \\ 0 & \frac{\beta_2}{\gamma+\mu} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

keeping in mind that $\mathbf{V}^{-1} = \frac{1}{\gamma+\mu}\mathbf{I}_4$, where $\mathbf{I}_4$ is the 4x4 identity matrix. The eigenvalues for this matrix are 0, $\frac{\beta_1}{\gamma+\mu}$, and $\frac{\beta_2}{\gamma+\mu}$. Using this, the basic reproduction number, R_0, of the system is the eigenvalue with the largest magnitude, which results in the expression [7]

$$R_0 = \rho(\mathbf{FV}^{-1}) = \max(|\frac{\beta_1}{\gamma+\mu}|, |\frac{\beta_2}{\gamma+\mu}|). \tag{3.4}$$

Then, the disease-free equilibrium is stable for $R_0 < 1$. Assuming one singular recovery rate γ and waning immunity rate α for the whole system, the exact expression for R_0 changes depending on the value of β_1 and β_2. Therefore, the average number of onward infections from an infected individual assuming a fully susceptible population depends on whichever strain has the highest rate of transmission. Important for this system, however, is that in general the expression for R_0 is similar to that for the single-strain diseases seen previously, and thus has similar interpretability.

3.3 Model Analysis - Bifurcation Points

When determining the possible behaviors of epidemic systems, it is important to look at the effects different parameter regimes have. Of particular interest is for which values the system changes stability, as these bifurcation points indicate when a disease may significantly increase in severity. The Dengue Model includes significantly more parameters compared to the simpler models of earlier investigations, and hence the possibilities for stability changes about the equilibria of the system increases, requiring deeper investigation into the characteristics of different bifurcations and their implications.

Beginning with the expression for R_0, this value acts as a key indicator for the overall presence of a disease. In its full form, so long as the magnitude of the ratio of the highest transmission rate to the sum of the rates of recovery and demography remains below one, the disease is expected to die out. Ignoring demography, this appears very similar to the R_0 expression for the SIRS and SEIR models, only with the incorporation of comparing which strain has the largest transmission rate magnitude. Also, in the symmetric case where $\beta_1 = \beta_2$, this equates to the SEIR and SIRS basic reproduction number with the same interpretation. Hence, R_0 also acts as a forward, transcritical bifurcation with respect to the number of infected individuals in the population.

Supplementary to the parameters impacting the basic reproduction number, other parameters specific to the multi-strain model and its ability to catalyze re-infections also possess the ability to drastically change the model trajectory based on their values. Specifically, previous investigations on this model have uncovered complex behaviors due to the forces of secondary infection ϕ_1 and ϕ_2 [1, 35]. Considering the symmetric case where $\beta_1 = \beta_2$ and $\phi_1 = \phi_2$ for computational efficiency, plotting the model using the Julia BifurcationKit package identifies the different bifurcations that can occur with the secondary infection forces [70]. From Figure 3.3 the symmetric Dengue system exhibits two hopf bifurcation points and one branch point. Each of these prelude the ability for the system to change as a result of the ϕ value. More specifically, the hopf bifurcation identifies a parameter value constituting a qualitative change in the model trajectory between a stable equilibrium and

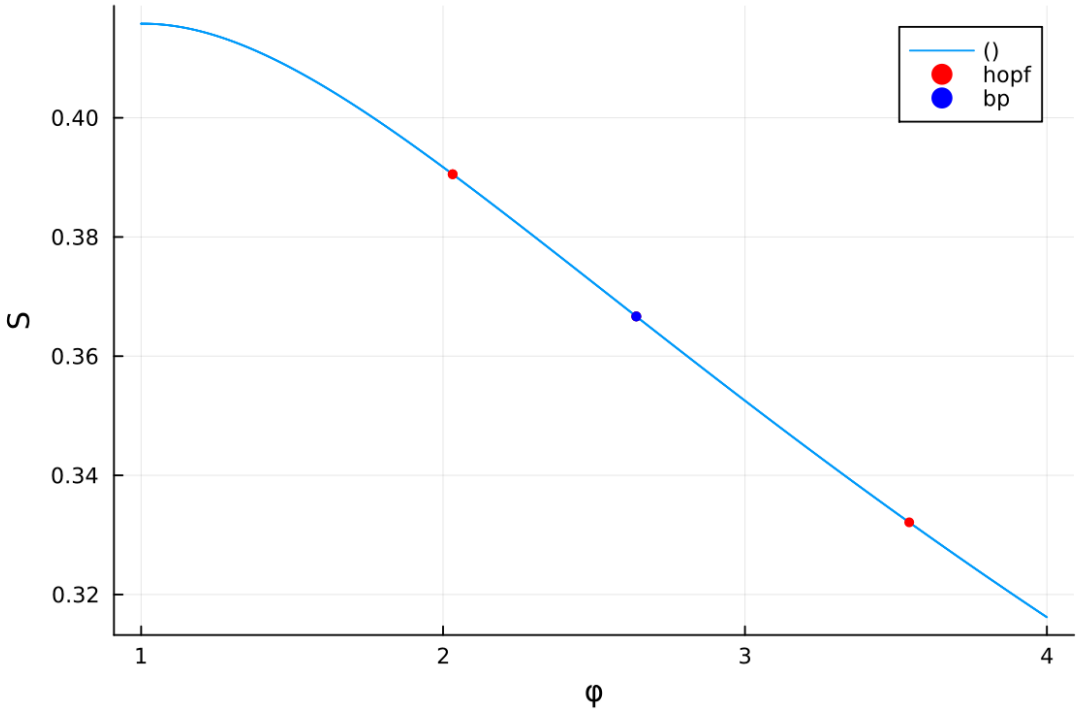


Figure 3.3: Bifurcation diagram for the $\phi = \phi_1 = \phi_2$ case of the Dengue Model using the Julia BifurcationKit package, with two hopf bifurcation points and one branch point plotted [70].

a limit cycle around the fixed point [59]. Conversely, a branch point indicates a topological change to the system, where a new branch of solutions is either created or destroyed, changing the global structure of the model [58, 59]. Both of these present the possibility for further bifurcations, which can lead to complex, possibly unpredictable behaviors. Both bifurcation types can result in further qualitative or structural changes, which eventually could lead to period-doubling behaviors that have been known to catalyze chaotic behavior [21, 46]. Entering potentially chaotic regimes increases the difficulty of controlling a disease, making it imperative to understand chaos and what occurs with the model trajectories as a result of being in a chaotic regime, as well as determining whether it is possible to identify these from estimating the parameters from real data.

3.4 Chaos

The greater complexity of the multi-strain models results in the possibility for a greater variety of behaviors to be exhibited by the system for different parameter values. In the single-strain cases, like the Dureau Model explored previously, the main concern revolved around human control over the Pandemic Influenza epidemic and the likelihood of the disease becoming endemic to the population or dying off over time. However, looking at the

simulations of the Dengue model in Figure 3.2, while the solutions in this parameter region appear somewhat periodic, they are not completely. In the time period shown from 2000 to 5000 years, the system exhibits clear oscillatory behavior, but without the repetition expected of a periodic system. When investigating the main changes of the model from the single-strain realm, the main alteration made, aside from simply adding more strains, is the introduction of the secondary infections, with primary infections being influenced by the secondary due to the forces of infections ϕ_1 and ϕ_2, which have been shown to introduce the route to more complex, possibly chaotic behavior through hopf bifurcation and branch points. Hence, it rests to investigate not simply the stability of the model under different parameter regions, but also the possibility of experiencing chaotic dynamics.

While there exist a variety of definitions, chaos typically describes the aperiodic behavior of a deterministic system that is very sensitive to initial conditions [12, 61]. Small variations in the initial values of the different state classes and for the initial parameter values can result in drastically different long-term behaviors of the model, making it more difficult to make effective predictions, as there may be other underlying patterns that are left to be understood [10, 12]. Chaos is a deterministic property, so understanding when a model may exhibit this behavior can still provide information on how to intervene, be it by better understanding what happens around critical thresholds, to learning more about appropriate timing for implementing intervention methods to help combat a disease [10]. However, chaos does not describe uncertain behavior, but rather unpredictability due to sensitivity to changes in external conditions [10].

3.4.1 Testing for Chaos

A simple method to test for chaos is the 0-1 test, developed by George A. Gottwald and Ian Melbourne to give a numerical summary as to whether a system exhibits regular or chaotic motion under a given parameter regime [26]. While there exist other methods to determine the chaos of a system, such as Lyapunov exponents, which have been used to previously study the unpredictability of multi-strain models for Dengue Fever, the 0-1 Test works directly with time-series data, which mitigates any issue arising from complex dynamical systems, including non-smooth systems, fractional derivatives, and non-chaotic strange attractors [1, 26]. The test runs as follows [26]: take some time series data $\phi(n)$ for $n = 1, 2, ...$ and for each n determine the values for the 2D system of auxiliary variables

$$\begin{aligned} p(n+1) &= p(n) + \phi(n)\cos(cn), \\ q(n+1) &= q(n) + \phi(n)\sin(cn), \end{aligned} \tag{3.5}$$

for a fixed $c \in (0, 2\pi)$. Then, calculate the time-averaged mean-square displacement

$$M(n) = \lim_{n\to\infty} \frac{1}{N} \sum_{i=1}^{N} ([p(j+n) - p(j)]^2 + [q(j+n) - q(j)]^2, \ n \in \{1, 2, ...\}. \tag{3.6}$$

From this, calculate the asymptotic growth rate $K \in [0, 1]$, which quantifies the regularity or chaos of the system

$$K = \lim_{n\to\infty} \frac{\log M(n)}{\log n}. \tag{3.7}$$

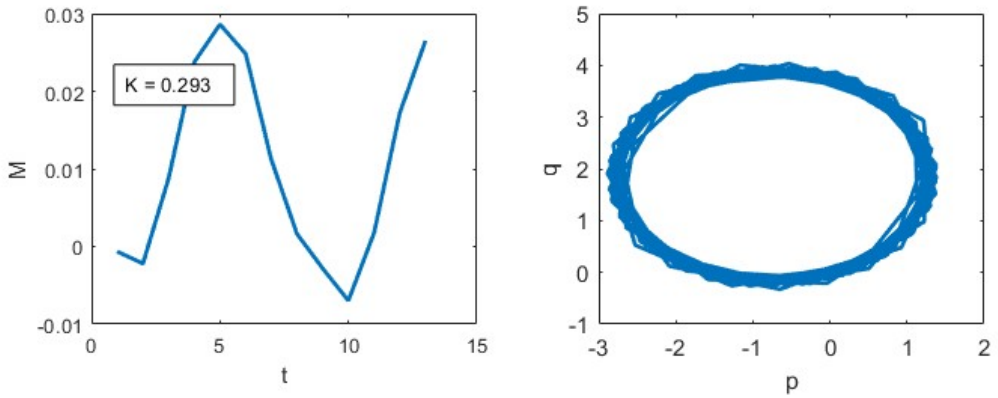
$K \approx 0$ indicates completely regular motion of the system with periodic or quasi-period trajectories, whereas $K \approx 1$ implies chaotic motion. Graphically, K near zero means that p and q have a fairly orderly relationship and $M(n)$ is periodic with respect to n. In contrast, K near one means that p and q have a disorderly relationship, appearing chaotic and random, and $M(n)$ grows linearly with respect to n [26].

To understand the types of possible dynamics in the Dengue model, and by extension other similar multi-strain systems, performing the 0-1 test for different values of ϕ_i helps to observe what dynamics the system exhibits relative to its bifurcation points. Producing simulations and performing the 0-1 test for $\phi_1 = \phi_2 = 0.1,\ 0.6$, and 0.8, the plots for $M(n)$ vs. t and q vs. p help identify which parameter fall into chaotic versus non-chaotic regimes. From Figure 3.4, the small values of ϕ_i clearly lie in the normal region as the $M(n)$ vs. t curve exhibits nonlinear behavior, and the q vs. p plot has a very regular, controlled periodic pattern that appears quite orderly. In comparison, the $M(n)$ vs. t plots for $\phi_i = 0.6$ and 0.8 appear very linear and their corresponding q vs. p plots have a fairly irregular, almost random, pattern. The orderliness seen in the non-chaotic region disappears as ϕ_i increases, indicating that large values for the secondary force of infection contribute to the chaos of this epidemic model. Given that these are just simulations, knowing the system has the potential for chaos means that not only are the transmission rates important to obtain reasonable estimates for, but also for the forces of secondary infection, as they drive behaviors that make combating the disease more difficult.

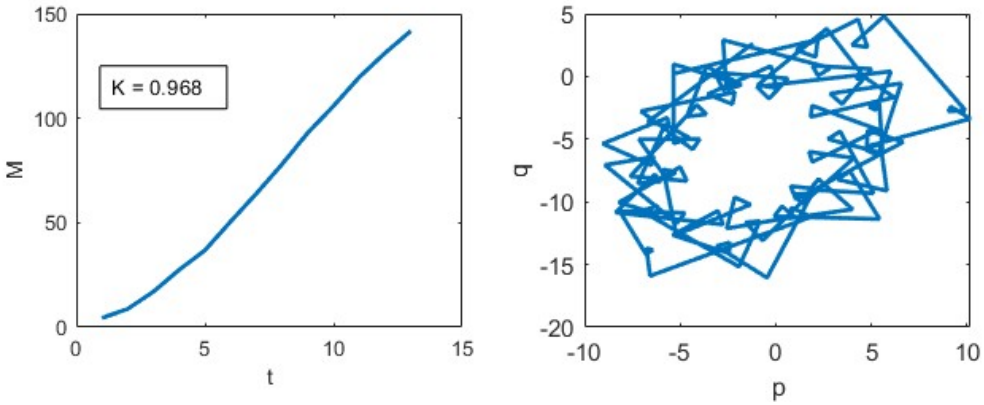
3.5 The Uncertainty of Chaos Drivers

Entering the multi-strain realm, the problem of parameter estimation becomes more complicated than that of single-strain models. Simply adding in a few extra characteristics, such as multiple strain with the possibility for re-infection, turns a straightforward stability problem into one with the potential for chaos coming from the behaviors of secondary parameters of the system. With chaotic systems being harder to predict, understanding under what conditions a disease progresses into the chaotic realm become critical knowledge, as knowing when a system will lose its predictability helps determine some of the

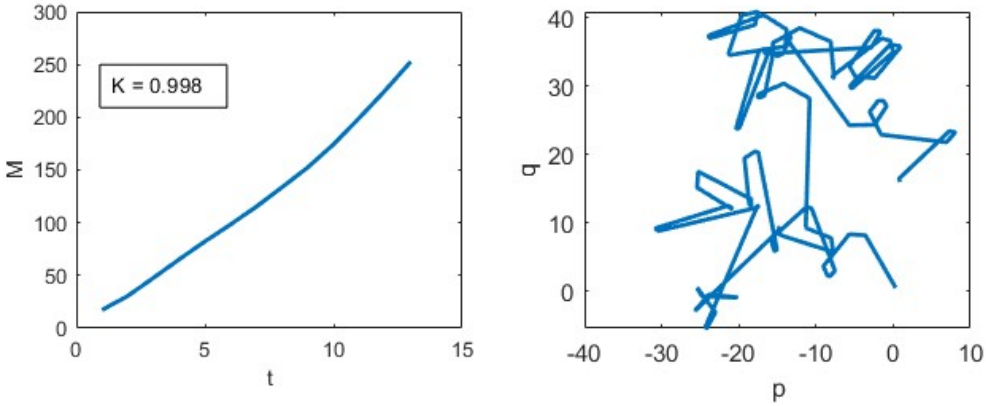
urgency needed to find solutions to mitigate its effects. Looking forward into estimating these parameters, it becomes necessary to look at more than just transmission parameters to guide the understanding of these systems, as their underlying features depend on the behaviors of these secondary parameters.



(a) $\phi_1 = \phi_2 = 0.1$



(b) $\phi_1 = \phi_2 = 0.6$



(c) $\phi_1 = \phi_2 = 0.8$

Figure 3.4: Output for the 0-1 test for chaos for $\phi_1 = \phi_2 = 0.1,\ 0.6,$ and 0.8. Plots on the left give $M(n)$ vs. t and plots on the right give q vs. p.

Chapter 4

Multi-Strain Disease Parameter Estimation

Extending the theory of parameter estimation into the multi-strain realm seems like a straightforward task of adding in some more variables and parameters to the model to match the questions to be answered. However, due to some inherent features to these complex models and issues with relevant data collection, the task of effectively estimating all the disease parameters of interest, including those promoting chaos, increases significantly. There are practicalities to consider and ideas to test, but the discussion turns from one about estimating parameters to study potential chaos to whether completing this task is even possible given the current availability of resources.

4.1 Changes to the Model Setup

Having previously explored the mathematical differences between single-strain and multi-strain models, particularly with respect to systems like the Dengue model in Equation 3.1, the additional complexities associated with these models require some additional care when preparing for parameter estimation. While incorporating the time-dependence by tracking state variables over a set of discrete time periods is straightforward as in the single-strain case, it is important to identify the differences in the observed states with respect to seen data and the full multi-strain model. The single-strain setting, for example, requires observing one reported cases class Z and fitting the rest of the states and parameters to estimate a single parameter, which previously has been the transmissibility parameter β. As model dynamics increase in complexity, accurately modeling the disease states over time requires not just the number or proportion of individuals infection within a population at each given time, but also with which strain they are infected.

Additionally, with the possibility of reinfections, such as in the Dengue model, it is important to consider the impact of secondary infections on total disease transmission. For

example, while infection with a Dengue Fever strain often causes immunity to that strain, those antibodies may aid in promoting more severe secondary infection with another strain [28]. What is known as antibody-dependence enhancement (ADE), the antibodies from the primary infection bind to the pathogen of a secondary infection, only neutralizing it enough to be able to enter immune cells, but not enough to prevent illness [29]. This allows for increased viral replication and an increased risk of severe infection. Similarly, oftentimes secondary infections of Respiratory Syncytial Virus (RSV) in children may present itself as an asymptomatic or mild infection, which also increase transmission due to unknown or undiagnosed infection [36]. Hence, given that secondary infections can result in varying transmission dynamics compared to the primary, it is just as important to consider not only with what strain an individual is infected, but also whether he or she is experiencing primary or secondary infection. Specifically, this entails estimating the force of infection from secondary infections, ϕ_i, which from previous investigation contributes to the potential for chaotic behavior in complex systems. Hence, the number of observed states needed to effectively model all the parameters related to disease transmission for the Dengue model increases to four:

$$
\begin{aligned}
Z_{11} &= \#\text{ of observed primary strain 1 infections,}\\
Z_{12} &= \#\text{ of observed primary strain 2 infections,}\\
Z_{21} &= \#\text{ of observed secondary strain 1 infections,}\\
Z_{22} &= \#\text{ of observed secondary strain 2 infections.}
\end{aligned}
$$

The changes in the number of observed infected individuals of each class can be represented similarly to the Dureau Model in Equation 2.9,

$$
\begin{aligned}
\frac{dZ_{11}}{dt} &= \frac{\beta_1 S I_1}{N} - Z_{11}(t \mod n),\\
\frac{dZ_{12}}{dt} &= \frac{\beta_2 S I_1}{N} - Z_{11}(t \mod n),\\
\frac{dZ_{21}}{dt} &= \frac{\phi_1 \beta_1 S I_1}{N} - Z_{21}(t \mod n),\\
\frac{dZ_{22}}{dt} &= \frac{\phi_2 \beta_2 S I_1}{N} - Z_{22}(t \mod n),
\end{aligned}
$$

where n is the time frame between observations. In addition to changing the number of observed states, there are two transmissibility parameters to estimate: β_1, and β_2. Each of these can be represented in a similar way to in the single-strain case, namely by assuming

each rate is distributed exponentially, where the rate is determined by some random noise:

$$\beta_1 = e^{x_1}, \qquad \frac{dx_1}{dt} = \sigma_1 dW,$$
$$\beta_2 = e^{x_2}, \qquad \frac{dx_2}{dt} = \sigma_2 dW.$$

4.2 The Identifiability Issue

Implementing the required changes to estimate the model parameters necessitates the model and data be of a workable form to extract the necessary information. However, when applying the PMCMC algorithm to real data with the Dengue Model, an identifiability issue arises. Identifiability refers to the ability to accurately estimate individual parameter values uniquely from both the model construction and the data provided [33]. Identifiability acts as a key indicator as to if adequate information can be generated from the problem setup. There exist two main types of identifiability, structural and practical, with each impacting a different aspect of the parameter estimation problem and the ability to obtain the desired results.

Structural identifiability occurs when changing a parameter value in the model alters the trajectory of the system [72]. More specifically, this refers to when a change in a parameter value cannot be compensated for elsewhere in the model without changing the model structure, resulting in the same trajectory as before. For example, take the following variation on an SIR model involving two strains with different transmission rates β_1 and β_2, with the main interest being the change in the to infected individuals I:

$$\frac{dS}{dt} = -\frac{S}{N}(\beta_1 + \beta_2), \qquad \frac{dI}{dt} = \frac{S}{N}(\beta_1 + \beta_2) - \gamma I, \qquad \frac{dR}{dt} = \gamma I.$$

If the goal is to estimate the individual β_i, this example is structurally non-identifiable, as given that β_1 and β_2 only exist in a sum together, with no terms where they act distinctly from one another, they cannot be distinguished from one another. Examining the sum as one parameter, however, resulting in the usual SIR model, would instead be identifiable, but this would require a change in the goal, as only the sum would be able to be estimated, not its individual components.

While structural identifiability refers to the construction of the model, practical identifiability takes into account the tools used to estimate the parameters as well. Practical identifiability means that not only does the model permit the parameters to be estimated,

but there also exists enough information from the data to perform the necessary calculations with enough accuracy and precision to be useful [71]. This is often an issue with sparse data, where the data do not allow the extraction of detailed enough information to perform the desired calculations. Estimating the parameters from sparse data often result in quite high bias and uncertainty, making the estimates not practically useful [27, 33]. In the Dengue model, this presents itself in that publicly available incidence data may separate between strains, but do not differentiate between primary and secondary infections, making it unreasonable to attempt to extract the forces of infections from the data directly. The inaccessibility of appropriate data not only inhibits effective parameter estimation, but also the understanding of the system's qualitative behavior, meaning it becomes difficult to analyze the potential chaotic behavior of the model.

4.3 Parameter Estimation Amidst Non-Identifiability

Amidst the issues of practical identifiability preventing the direct estimation of the forces of secondary infection, it is still possible to try to extract the plausibility of the chaotic regime given the information the data provide. While an exact estimate is not feasible, exploring a reduced system that still may have chaotic potential can still give an idea as to if chaos is a truly feasible properly to explore within the model.

4.3.1 A Truncated Model

Considering the practical concerns with respect to identifiability, it is infeasible to estimate any of the model parameters from the full Dengue Model given that the data do not support many of the state variables needed to provide useful estimations. Therefore, the data result in only being able to work with a truncated version of the model, given by

$$\begin{aligned}
\frac{d\tilde{S}}{dt} &= -\frac{\tilde{\beta}_1 \tilde{S}\tilde{I}_1}{N} - \frac{\tilde{\beta}_2 \tilde{S}\tilde{I}_2}{N} + \alpha(\tilde{R}_1 + \tilde{R}_2), \\
\frac{d\tilde{I}_1}{dt} &= \frac{\tilde{\beta}_1 \tilde{S}\tilde{I}_1}{N} - \gamma \tilde{I}_1, \\
\frac{d\tilde{I}_2}{dt} &= \frac{\tilde{\beta}_2 \tilde{S}\tilde{I}_2}{N} - \gamma \tilde{I}_2, \\
\frac{d\tilde{R}_1}{dt} &= \gamma \tilde{I}_1 - \alpha \tilde{R}_1, \\
\frac{d\tilde{R}_2}{dt} &= \gamma \tilde{I}_2 - \alpha \tilde{R}_2,
\end{aligned} \tag{4.1}$$

where $\tilde{S} = S + S_1 + S_2$, $\tilde{I}_1 = I_1 + I_{21}$, $\tilde{I}_2 = I_2 + I_{12}$, and $\tilde{R}_1$ and $\tilde{R}_2$ are similar to before, just modified to account for their value changes occurring due to different input variables.

Using the next generation matrix methods, R_0 is also computed for this system, where the infectious subsystem is

$$\dot{\tilde{I}}_1 = \frac{\tilde{\beta}_1 \tilde{S} \tilde{I}_1}{N} - \gamma \tilde{I}_1,$$
$$\dot{\tilde{I}}_2 = \frac{\tilde{\beta}_2 \tilde{S} \tilde{I}_2}{N} - \gamma \tilde{I}_2.$$

Then, for the DFE of $(\tilde{S}, \tilde{I}_1, \tilde{I}_2, \tilde{R}_1, \tilde{R}_2) = (N, 0, 0, 0, 0)$, the Jacobian matrix becomes

$$\begin{aligned} J &= \begin{bmatrix} \frac{\tilde{\beta}_1 \tilde{S}}{N} & 0 \\ 0 & \frac{\tilde{\beta}_2 \tilde{S}}{N} \end{bmatrix}_{(N,0,0,0,0)}, \\ &= \begin{bmatrix} \tilde{\beta}_1 - \gamma & 0 \\ 0 & \tilde{\beta}_2 - \gamma \end{bmatrix}, \\ &= \underbrace{\begin{bmatrix} \tilde{\beta}_1 & 0 \\ 0 & \tilde{\beta}_2 \end{bmatrix}}_{\mathbf{F}} - \underbrace{\begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}}_{\mathbf{V}}. \end{aligned}$$

The expression for R_0 comes from the spectral radius of $\mathbf{F}\mathbf{V}^{-1}$, calculated as

$$\begin{aligned} \mathbf{F}\mathbf{V}^{-1} &= \begin{bmatrix} \tilde{\beta}_1 & 0 \\ 0 & \tilde{\beta}_2 \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}^{-1}, \\ &= \begin{bmatrix} \tilde{\beta}_1 & 0 \\ 0 & \tilde{\beta}_2 \end{bmatrix} \frac{1}{\gamma^2} \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}, \\ &= \gamma^{-2} \begin{bmatrix} \tilde{\beta}_1 \gamma & 0 \\ 0 & \tilde{\beta}_2 \gamma \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}, \\ &= \begin{bmatrix} \frac{\tilde{\beta}_1}{\gamma} & 0 \\ 0 & \frac{\tilde{\beta}_2}{\gamma} \end{bmatrix}, \end{aligned}$$

from which the spectral radius is the largest eigenvalue of

$$\begin{aligned} \det\left(\mathbf{F}\mathbf{V}^{-1} - \lambda I\right) &= \begin{vmatrix} \frac{\tilde{\beta}_1}{\gamma} - \lambda & 0 \\ 0 & \frac{\tilde{\beta}_2}{\gamma} - \lambda \end{vmatrix} = 0, \\ &\iff (\frac{\tilde{\beta}_1}{\gamma} - \lambda)(\frac{\tilde{\beta}_2}{\gamma} - \lambda) = 0. \end{aligned}$$

Thus, the largest eigenvalue of the system, and by extension the expression for R_0, is given by

$$R_0 = \max(|\frac{\tilde{\beta}_1}{\gamma}|, |\frac{\tilde{\beta}_2}{\gamma}|). \tag{4.2}$$

Ignoring demography, this expression is essentially the same as for the full Dengue Model in Equation 3.4.

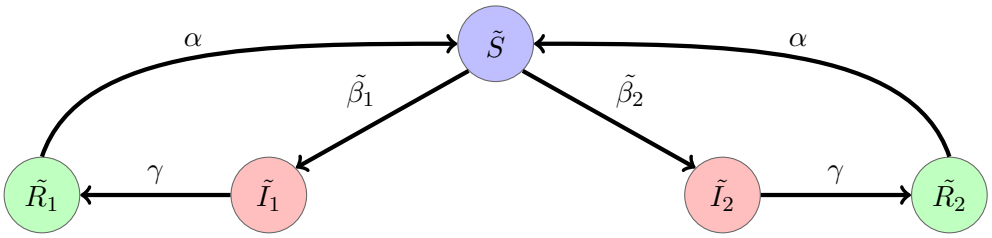


Figure 4.1: Compartment diagram of the truncated Dengue Model in Equation 4.1

Even though practical identifiability prevents the estimation of ϕ_1 and ϕ_2 from incidence data, other key quantities like R_0 which help summarize the impact of a disease can still be estimated with reasonable precision for the disease at hand [33]. The truncated model also provides a reasonable estimate for the total susceptible population, so the effective reproduction number R_t may also be found by estimating the transmission rates for each strain of the truncated model. With the only missing feature due to non-identifiability with respect to the full model being the chaotic potential from the secondary infection forces, it rests to determine if estimating these quantities alone as a representation of the full model is sufficient to determine chaotic potential of the full model.

4.3.2 Estimating β_1 and β_2 via PMCMC

To investigate the impact of different values of ϕ_1 and ϕ_2 on the model given the data, it is necessary to begin by estimating parameters from the truncated model of Equation 4.1. Proceeding as in the single-strain case, taking γ and α as fixed for simplicity and considering any demography as negligible, the PMCMC algorithm is run on general priors given the two observed states Z_1 and Z_2 for the observed infections of each strain. Data are pulled from the Singapore Ministry of Health, taking weekly recordings of Dengue and Dengue Hemorrhagic Fever numbers from 2014 until 2018 [52]. Singapore, being a nation with high urban population density and warm, humid climate, fosters an environment for high mosquito fertilization, which in turn contributes to transmission of mosquito-borne illnesses like Dengue Fever [55]. Following a Dengue Fever outbreak in 2013, infections remained higher than average, contributing to the need to estimate true infections and the likelihood of future spikes in transmission [55]. Performing the PMCMC algorithm on the first 100 weeks, given a population of about 5.47 million in 2014, produces 1000 PMCMC posterior samples, from which β_1 and β_2 samples can be extracted [73]. Afterwards, the posterior values of R_t can also be calculated, giving an idea of the overall control of the infection given the data.

Figure 4.2 gives plots for the posterior samples of total incidence, $\beta_1(t)$, $\beta_2(t)$, and R_t for the 100 weeks of observed cases, along with their 50% and 95% confidence intervals

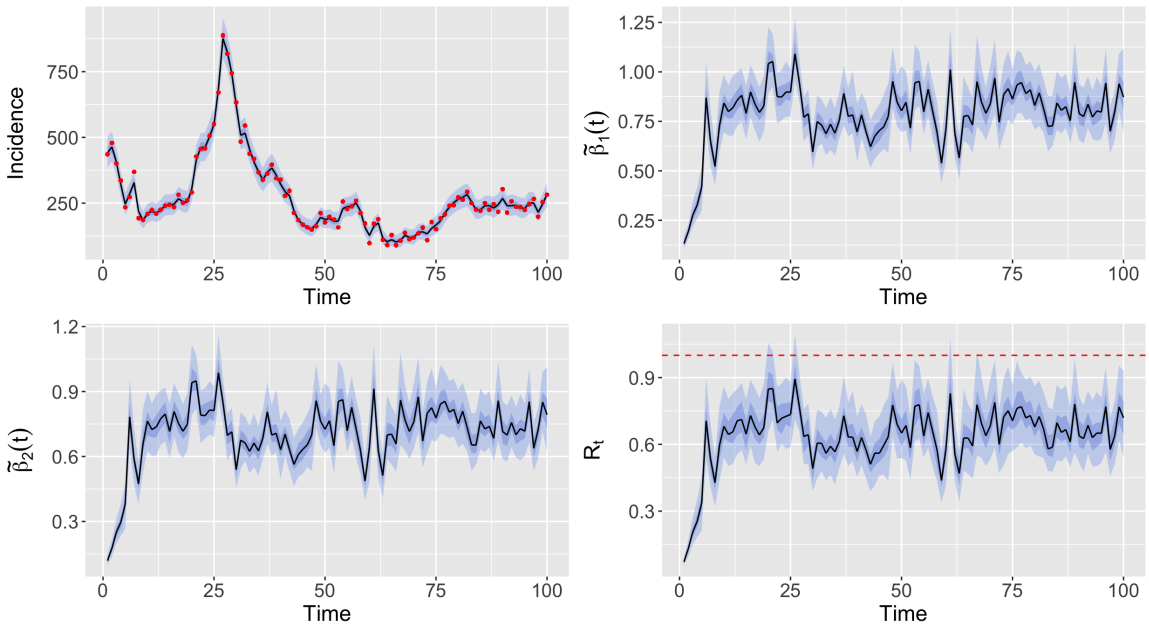


Figure 4.2: Output for 1000 PMCMC samples taken on the Singapore Ministry of Health dataset.

associated with the samples. As with the Dureau example, the incidence plot gives the data points, in which case is the sum of total infections for both Dengue and DHF at each time point, showing that given the adaptive priors in the algorithm, the model fits the data points well on average, so is useful to track the total number of cases. Additionally, there does not appear to be severe deviation from the mean case count in black, indicating that it is possible there is not too much uncertainty with respect to the number of infections at each time point. Similarly, the plots for both transmission rates appear relatively controlled, with both remaining in the 0.6 to 1.0 range. Observing the output, estimates for γ are above 1.0, and the samples are centered around 1.23, indicating that R_0 will almost always remain close to or below 1. When considering this with respect to the total susceptible population, the effective reproduction number implies that despite even the peak of infection at around 27 weeks for this time period, total Dengue Fever transmissions remained relatively low, which means the infection is still considered well-controlled.

4.3.3 A Chaotic Potential

Given that the forces of secondary infection cannot be estimated due to practical identifiability, it rests to observe whether the data-backed parameter estimates of the truncated model are sufficient to observe chaos. In order to proceed, an expression for β_1 and β_2 in the full model must be found, given that the truncated model computes the transmissions based on the total susceptible population, whereas the full model takes into account those

susceptible only to a particular strain as well. Noting that $\tilde{S} = S + S_1 + S_2$, $\tilde{I_1} = I_1 + I_{21}$, and $\tilde{I_2} = I_2 + I_{12}$, for $\tilde{I_1}$,

$$\begin{aligned}
\frac{d\tilde{I}}{dt} &= \frac{dI_1}{dt} + \frac{dI_2}{dt}, \\
\implies \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{N} - \gamma\tilde{I_1} &= \frac{\beta_1}{N}S(I_1 + \phi_1 I_{21}) - \gamma I_1 + \frac{\beta_2}{N}S_2(I_1 + \phi_1 I_{21}) - \gamma I_{21}, \\
\implies \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{N} - \gamma\tilde{I_1} &= \frac{\beta_1}{N}(S + S_2)(I_1 + \phi_1 I_{21}) - \gamma(I_1 + I_{21}), \\
\implies \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{N} - \gamma\tilde{I_1} &= \frac{\beta_1}{N}(S + S_2)(I_1 + \phi_1 I_{21}) - \gamma\tilde{I_1}, \\
\implies \tilde{\beta_1}\tilde{S}\tilde{I_1} &= \beta_1(S + S_2)(I_1 + \phi_1 I_{21}), \\
\implies \beta_1 &= \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{(S + S_2)(I_1 + \phi_1 I_{21})}.
\end{aligned}$$

Similarly, for $\tilde{I_2}$,

$$\beta_2 = \frac{\tilde{\beta_2}\tilde{S}\tilde{I_2}}{(S + S_1)(I_2 + \phi_2 I_{12})}.$$

This then gives expression that can be substituted into the Dengue model in Equation 3.1, incorporating the data for the susceptible and infected compartments and the transmission rates supported by the PMCMC samples. Substituting the expressions for β_1 and β_2 gives produces the model

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{(S + S_2)N}S - \frac{\tilde{\beta_2}\tilde{S}\tilde{I_2}}{(S + S_1)N}S, \\
\frac{dI_1}{dt} &= \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{(S + S_2)N}S - \gamma I_1, \\
\frac{dI_2}{dt} &= \frac{\tilde{\beta_2}\tilde{S}\tilde{I_2}}{(S + S_1)N}S - \gamma I_2, \\
\frac{dR_1}{dt} &= \gamma I_1 - \alpha R_1, \\
\frac{dR_2}{dt} &= \gamma I_2 - \alpha R_2, \\
\frac{dS_1}{dt} &= -\frac{\tilde{\beta_2}\tilde{S}\tilde{I_2}}{(S + S_1)N}S_1 + \alpha R_1, \\
\frac{dS_2}{dt} &= -\frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{(S + S_2)N}S_2 + \alpha R_2, \\
\frac{dI_{12}}{dt} &= \frac{\tilde{\beta_2}\tilde{S}\tilde{I_2}}{(S + S_1)N}S_1 - \gamma I_{12}, \\
\frac{dI_{21}}{dt} &= \frac{\tilde{\beta_1}\tilde{S}\tilde{I_1}}{(S + S_2)N}S_2 - \gamma I_{21}, \\
\frac{dR}{dt} &= \gamma(I_{21} + I_{12}).
\end{aligned} \tag{4.3}$$

This version of the Dengue Model no longer depends on the forces of secondary infection ϕ_i, but still allows for the simulation of the transitions between the ten compartments of the full model. Eliminating parameters that cannot be directly estimated due to the lack of identifiability from the data allows for the use of the sampled state and parameter values as guides for giving plausible simulations for the other non-identifiable compartments, to at least observe the possibility for chaos of the full model. However, while the model eliminates the need to estimate the secondary infection fores, the computational complexity of feeding the data into the model at each time step as the model is being fitted cannot be efficiently calculated in MATLAB. Despite the updated model being such that it should retain the same dynamics as the Dengue Model, the inability to efficiently simulate restricts the ability to investigate the chaotic behavior of Dengue Fever in the real world. The problem with data limitations extends beyond the Dengue Model, as data quality has presented a major hindrance to epidemiological modeling and response in numerous modern-day epidemics. Particularly with the COVID-19 outbreak, issues such as inconsistent data acquisition, delayed reporting, and missing data hindered the ability to estimate the constantly evolving dynamics of the disease, significantly increasing the efforts needs to predict its progression and inform public response [45]. The crux of the identifiability, therefore, remains not a model-specific one but rather one around epidemiological data collection in general. Bringing these complex, theoretical models into reality successfully requires data that can explore all the state variables. Else, the risk of estimating with insufficient observations results in highly variable, often extrapolated estimates that would have at most minuscule biological interpretability and utility.

4.4 The Reality of Complex Parameter Estimation

Performing time-dependent parameter estimation on complex epidemic models proves to be a difficult task, particularly with respect to data limitations that prevent the whole system from being identifiable. While intricate systems like the Dengue model have predominantly been used for theoretical examination of disease dynamics and chaotic potential through controlled simulation, there still exists a gap when trying to fit these models to real data as methods for exploring the more hidden features of disease transmission, with the hopes of generating better analysis and providing more informative insight into the progression of an epidemic. There exist approximations to deal with the practical identifiability issues by working with truncated models that better represent the information the data can provide, but the cost of the inability to effectively re-simulate the original model dynamics with all of its intricacies make it difficult to bring these highly complex, theoretical investigations into the reality.

Chapter 5

Conclsion

The past century has seen significant developments in both the fields of epidemiological modeling and parameter estimation. What started with Kermack and McCormick's mass action-based SIR model and estimates via analytical solutions has progressed into complex problems requiring a mix of dynamical systems ideologies along with cunning statistical tools to better capture how pathogens react and spread in the natural world. There would not exist these developments, however, if there was not a starting point from which to improve upon, and then questions that drove the need for further improvement in the methods used to analyze these diseases. No method is perfect, and with each iteration of a model or a technique, there is always something left unanswered to propel into the next investigation, and this study is no different.

The ability to mathematically define the progression of multi-strain diseases proves a useful tool when understanding the relationships between different strains, cross and waning immunity, and other factors that affect the transmission of a pathogen. Looking at these models purely analytically, one can begin to see many ways in which more elements can be added to make the model better match the intricacies, including how different types of infections impact transmission. Seen in the Dengue model through the forces of secondary infection ϕ_1 and ϕ_2, these additional factors allow for rich mathematical behavior, as they add enough additional nonlinearity to the multi-strain system where chaotic behavior can occur. A feature not seen as well in single strain or less complex systems, chaos makes it harder to predict exactly when transmission may spike or lull, which then requires extra care when determining the best time to implement measures to counter disease spread. However, as data are seen a hindrance in being able to identify the parameter regime of the force of secondary infection terms comes in the lack of data availability, meaning that with current methods it is unreasonable to try and estimate them from the available data, as doing so would be based solely on extrapolation, reducing the usefulness of any results. Hence, adaptations have to be made to match the data availability, which allow precise fitting of

the data, even in the presence of hidden features like observation error, but potentially losing the interpretability of mathematical features like chaos. The inability to estimate such features due to practical identifiability marks a key area of improvement needed in the field of epidemiological modeling, but those improvements go beyond epidemiology and instead turn back to how to best work with and collect data so that all features of these models can be represented in the estimations.

One of the main goals of mathematical and statistical research is to continuously learn from the work that has been done and make improvements to fix the problems that exist. While this reigns true, such that novel approaches and improvements on the relationships between data and mathematical models will have to be made to better capture the intricacies in the complex systems that cannot be done so currently, it also raises a question that lies at the heart of applied mathematics - is it just in the theoretics? Particularly in the multi-strain setting, the Dengue model used as the basis for investigation has previously been used in studies that look at the dynamics in a purely theoretical, simulation-based approach. Meant to explore more the mathematical ideas of chaos and stability theory, these previous studies have focused on setting initial conditions and controlled parameter values to explore what could happen in the real world if the data supported the models. However, it is possible that as more questions arise and more features are added to the systems that there becomes a point where they diverge from realism, becoming too specific in certain areas to truly capture what happens in the real world. It is also possible that delving too deeply into mathematical intricacies and ways to explore it on a theoretical level causes deviation from the purpose of these models which are to be tools to give informative outlooks to help inform public health response.

Looking at the complex, multi-strain models theoretically, the potential for chaos seems important. Entering a chaotic parameter regime means that disease progression becomes harder to predict due to sensitivities of the system, which in a real-world setting would mean it becomes harder to effectively implement measures at the most appropriate time as small changes in external conditions can completely alter the trajectory. But at the same time, the data are not recorded in enough detail to allow for these behaviors to be tracked, so it leaves to ask if it is truly necessary to public health response to investigate long-term chaotic behavior, or if it is truly important only in the realm of computer simulations.

This is not, however, meant to detract from the usefulness of theory and simulations. Even if just in a theoretical exploration, simulations of different models provide insights into potential outcomes over a variety of assumptions, helping build intuition as to how pathogens are impacted by different features. Even if they do not provide precise predictions, their value lies in the ability to explore the impacts of intervention strategies over a range of

possible situations and constraints to help determine what types of intervention strategies may be needed in the future. Additionally, knowing the strength of the relationships between different theoretical features, such between the force that secondary infections and total infections, may help to plan which groups intervention strategies should be weighted towards to hopefully provide the most impactful response possible.

While multi-strain disease modeling is far from a complete story, each piece of insight will help inform future advancements in both data acquisition and analytical modeling and statistical techniques to unlock new ways of understanding and controlling these diseases. It is by acknowledging the current setbacks and embracing the uncertainties they provide, and then using those setbacks as propulsion to refine methods and improve epidemic prediction which drives further developments in the field, promoting new research into improvements in both epidemiology and data acquisition. As these developments are made, the link between theory and reality strengthens, promoting new ways of data collection and estimating disease transmission, withe hope of contributing to another century of developments towards understanding these ever-evolving pathogens.

Bibliography

[1] Maíra Aguiar, Bob Kooi, and Nico Stollenwerk. Epidemiology of Dengue Fever: A Model with Temporary Cross-Immunity and Possible Secondary Infection Shows Bifurcations and Chaotic Behaviour in Wide Parameter Regions. *Mathematical Modelling of Natural Phenomena*, 3(4):48–70, 2008.

[2] Roy M. Anderson and Robert M. May. Directly Transmitted Infections Diseases: Control by Vaccination. *Science*, 215(4536):1053–1060, February 1982.

[3] Roy M. Anderson and Robert M. May. *Infectious diseases of humans : dynamics and control*. Oxford University Press, Oxford, 1991.

[4] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342, June 2010.

[5] Richard T. Baillie, Francis X. Diebold, George Kapetanios, Kun Ho Kim, and Aaron Mora. On Robust Inference in Time Series Regression, May 2024. arXiv:2203.04080.

[6] Archana Singh Bhadauria and Hom Nath Dhungana. Epidemic theory: Studying the effective and basic reproduction numbers, epidemic thresholds and techniques for the analysis of infectious diseases with particular emphasis on tuberculosis. In *Methods of Mathematical Modelling*, pages 1–21. Elsevier, 2022.

[7] Julie Blackwood and Lauren Childs. An Introduction to Compartmental Modeling for the Budding Infectious Disease Modeler. *Letters in Biomathematics*, 5(1), 2018.

[8] N. F. Britton. *Essential mathematical biology*. Springer Undergraduate Mathematics Series. Springer, London, 2003.

[9] Andrew F. Brouwer. Why the Spectral Radius? An intuition-building introduction to the basic reproduction number. *Bulletin of Mathematical Biology*, 84(9):96, August 2022.

[10] Arianna Calistri, Pier Francesco Roggero, and Giorgio Palù. Chaos theory in the understanding of COVID-19 pandemic dynamics. *Gene*, 912:148334, June 2024.

[11] Alex Capaldi, Samuel Behrend, Benjamin Berman, Jason Smith, Justin Wright, and Alun L. Lloyd. Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical biosciences and engineering: MBE*, 9(3):553–576, July 2012.

[12] Robert L. Devaney. *An Introduction to Chaotic Dynamical Systems.* Benjamin/Cummings, Menlo Park, Calif, 1986.

[13] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47):873–885, June 2010.

[14] Arnaud Doucet, Nando de Freitas, and Neil Gordon. An Introduction to Sequential Monte Carlo Methods. In Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 3–14. Springer New York, New York, NY, 2001.

[15] Joseph Dureau, Konstantinos Kalogeropoulos, and Marc Baguelin. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, 14(3):541–555, July 2013.

[16] Leah Edelstein-Keshet. *Mathematical Models in Biology.* Random House, New York, 1 edition, 1988.

[17] Akira Endo, Edwin van Leeuwen, and Marc Baguelin. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics*, 29:100363, December 2019.

[18] Oscar Fajardo-Fontiveros, Mattia Mattei, Giulio Burgio, Clara Granell, Sergio Gómez, Alex Arenas, Marta Sales-Pardo, and Roger Guimerà. Machine learning mathematical models for incidence estimation during pandemics. *PLOS Computational Biology*, 20(12):e1012687, December 2024.

[19] Gabriel Obed Fosu, Emmanuel Akweittey, and Albert Adu-Sackey. Next-Generation Matrices and Basic Reproductive Numbers for All Phases of the Coronavirus Disease.

[20] Dani Gamerman. *Markov chain Monte Carlo : stochastic simulation for Bayesian inference.* Chapman and Hall texts in statistical science series. Chapman & Hall, London, 1997.

[21] L. Gardini, R. Lupini, and M. G. Messia. Hopf bifurcation and transition to chaos in Lotka-Volterra equation. *Journal of Mathematical Biology*, 27(3):259–272, May 1989.

[22] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. Bayesian Data Analysis Third edition (with errors fixed as of 20 February 2025). Technical report, Columbia University, 2025.

[23] Cheryl L. Gibbons, Marie-Josée J. Mangen, Dietrich Plass, Arie H. Havelaar, Russell John Brooke, Piotr Kramarz, Karen L. Peterson, Anke L. Stuurman, Alessandro Cassini, Eric M. Fèvre, and Mirjam EE Kretzschmar. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, 14(1):147, February 2014.

[24] W.R Gilks, S Richardson, and D.J. Spieglehalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1996.

[25] Paolo Girardi and Carlo Gaetan. An SEIR Model with Time-Varying Coefficients for Analyzing the SARS-CoV-2 Epidemic. *Risk Analysis*, page 10.1111/risa.13858, November 2021.

[26] Georg A. Gottwald and Ian Melbourne. On the Implementation of the 0-1 Test for Chaos. *SIAM Journal on Applied Dynamical Systems*, 8(1):129–145, January 2009. arXiv:0906.1418 [nlin].

[27] Sander Greenland, Judith A. Schwartzbaum, and William D. Finkle. Problems due to Small Samples and Sparse Data in Conditional Logistic Regression Analysis. *American Journal of Epidemiology*, 151(5), 2000.

[28] Maria G. Guzman, Mayling Alvarez, and Scott B. Halstead. Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection. *Archives of Virology*, 158(7):1445–1459, July 2013.

[29] Maria G. Guzman and Susana Vazquez. The Complexity of Antibody-Dependent Enhancement of Dengue Virus Infection. *Viruses*, 2(12):2649–2662, December 2010.

[30] M. H. Hassan, Tamer El-Azab, Ghada AlNemer, M. A. Sohaly, and H. El-Metwally. Analysis Time-Delayed SEIR Model with Survival Rate for COVID-19 Stability and Disease Control. *Mathematics*, 12(23):3697, November 2024.

[31] Dame Deirdre Hine. The 2009 Influenza Pandemic, 2009.

[32] Hyukpyo Hong, Eunjin Eom, Hyojung Lee, Sunhwa Choi, Boseung Choi, and Jae Kyoung Kim. Overcoming bias in estimating epidemiological parameters with realistic history-dependent disease spread dynamics. *Nature Communications*, 15(1):8734, October 2024.

[33] Yu-Han Kao and Marisa C. Eisenberg. Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. *Epidemics*, 25:89–100, December 2018.

[34] William O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, August 1927.

[35] Bob W. Kooi, Maíra Aguiar, and Nico Stollenwerk. Bifurcation analysis of a family of multi-strain epidemiology models. *Journal of Computational and Applied Mathematics*, 252:148–158, November 2013.

[36] A. Kutsaya, T. Teros-Jaakkola, L. Kakkola, L. Toivonen, V. Peltola, M. Waris, and I. Julkunen. Prospective clinical and serological follow-up in early childhood reveals a high rate of subclinical RSV infection and a relatively high reinfection rate within the first 3 years of life. *Epidemiology and Infection*, 144(8):1622–1633, June 2016.

[37] Teddy Lazebnik. Computational applications of extended SIR models: A review focused on airborne pandemics. *Ecological Modelling*, 483:110422, September 2023.

[38] Teddy Lazebnik and Gaddi Blumrosen. Advanced Multi-Mutation With Intervention Policies Pandemic Model. *IEEE Access*, 10:22769–22781, 2022.

[39] Nathan P. Lemoine. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7):912–928, July 2019.

[40] Jing Li, Daniel Blakeley, and Robert J. Smith? The Failure of R 0. *Computational and Mathematical Methods in Medicine*, 2011:527610, 2011.

[41] Rui Li, Yan Li, Zhuoru Zou, Yiming Liu, Xinghui Li, Guihua Zhuang, Mingwang Shen, and Lei Zhang. Evaluating the Impact of SARS-CoV-2 Variants on the COVID-19 Epidemic and Social Restoration in the United States: A Mathematical Modelling Study. *Frontiers in Public Health*, 9, January 2022.

[42] Xinzhi Liu and Peter Stechlinski. Infectious disease models with time-varying parameters and general nonlinear incidence rate. *Applied Mathematical Modelling*, 36(5):1974–1994, May 2012.

[43] Nimalan Mahendran, Ziyu Wang, Firas Hamze, and Nando de Freitas. Bayesian Optimization for Adaptive MCMC, October 2011. arXiv:1110.6497.

[44] Anthony Martin and Michele Harvey. *Linear Algebra: Concepts and Methods*. Cambridge University Press, Cambridge, 1 edition, June 2012.

[45] Kelly R. Moran, Tammie Lopez, and Sara Y. Del Valle. The future of pandemic modeling in support of decision making: lessons learned from COVID-19. *BMC Global and Public Health*, 3(1):24, March 2025.

[46] Simona Muratori. Bifurcations and Chaos in a Periodic Predator-Prey Model. *International Journal of Bifurcation and Chaos*, January 1992.

[47] J.D. Murray. *Mathematical Biology: I: An Introduction*, volume v.17. Springer, New York, London, 3 edition, 2002.

[48] Lawrence M. Murray. Bayesian State-Space Modelling on High-Performance Hardware Using LibBi, June 2013. arXiv:1306.3277 [stat].

[49] Prathiba Natesan, Ratna Nandakumar, Tom Minka, and Jonathan D. Rubright. Bayesian Prior Choice in IRT Estimation Using MCMC and Variational Bayes. *Frontiers in Psychology*, 7, September 2016.

[50] Willie Neiswanger and Eric Xing. Post-Inference Prior Swapping. In *Proceedings of the 34th International Conference on Machine Learning*, pages 2594–2602. PMLR, July 2017.

[51] Hiroshi Nishiura and Gerardo Chowell. The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In Gerardo Chowell, James M. Hyman, Luís M. A. Bettencourt, and Carlos Castillo-Chavez, editors, *Mathematical and Statistical Estimation Approaches in Epidemiology*, pages 103–121. Springer Netherlands, Dordrecht, 2009.

[52] Ministry of Health. Weekly Number of Dengue and Dengue Haemorrhagic Fever Cases, 2019.

[53] Olusegun Michael Otunuga. Analysis of multi-strain infection of vaccinated and recovered population through epidemic model: Application to COVID-19. *PLOS ONE*, 17(7):e0271446, July 2022.

[54] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, August 2015. Publisher: American Physical Society.

[55] Jayanthi Rajarethinam, Li-Wei Ang, Janet Ong, Joyce Ycasas, Hapuarachchige Chanditha Hapuarachchi, Grace Yap, Chee-Seng Chong, Yee-Ling Lai, Jeffery Cutter, Derek Ho, Vernon Lee, and Lee-Ching Ng. Dengue in Singapore from 2004 to 2016: Cyclical Epidemic Patterns Dominated by Serotypes 1 and 2. *The American Journal of Tropical Medicine and Hygiene*, 99(1):204–210, July 2018.

[56] Nicholas G. Reich, Logan C. Brooks, Spencer J. Fox, Sasikiran Kandula, Craig J. McGowan, Evan Moore, Dave Osthus, Evan L. Ray, Abhinav Tushar, Teresa K. Yamana, Matthew Biggerstaff, Michael A. Johansson, Roni Rosenfeld, and Jeffrey Shaman. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences*, 116(8):3146–3154, February 2019. Publisher: Proceedings of the National Academy of Sciences.

[57] Christian P. Robert. The Metropolis-Hastings algorithm, January 2016. arXiv:1504.01896 [stat].

[58] R. Seydel. Branch Switching in Bifurcation Problems for Ordinary Differential Equations. *Numerische Mathematik*, 41:93–116, 1983.

[59] Rüdiger Seydel. *Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos*. Interdisciplinary Applied Mathematics. Springer-Verlag, New York, 2 edition, 1994.

[60] Abdullah A. Smadi and Nour H. Abu-Afouna. On Least Squares Estimation in a Simple Linear Regression Model with Periodically Correlated Errors: A Cautionary Note. *Austrian Journal of Statistics*, 41(3):211–226, 2012.

[61] Julien Clinton Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, Oxford, 2003.

[62] Ben Swallow, Paul Birrell, Joshua Blake, Mark Burgman, Peter Challenor, Luc E. Coffeng, Philip Dawid, Daniela De Angelis, Michael Goldstein, Victoria Hemming, Glenn Marion, Trevelyan J. McKinley, Christopher E. Overton, Jasmina Panovska-Griffiths, Lorenzo Pellis, Will Probert, Katriona Shea, Daniel Villela, and Ian Vernon. Challenges in estimation, uncertainty quantification and elicitation for pandemic modelling. *Epidemics*, 38:100547, March 2022.

[63] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and A. Cori. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, December 2019.

[64] Miliyon Tilahun. Backward bifurcation in SIRS malaria model, July 2017. arXiv:1707.00924.

[65] Sirachat Tipsri and Wirawan Chinviriyasit. Stability Analysis of SEIR Model with Saturated Incidence and Time Delay. *International Journal of Applied Physics and Mathematics*, 4(1):42–45, 2014.

[66] Riitta Toivonen, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo Kaski. A comparative study of social network models: network evolution models and nodal attribute models, December 2008. arXiv:0805.0512.

[67] Imelda Trejo and Nicolas Hengartner. A modified Susceptible-Infected-Recovered model for observed under-reported incidence data, 2020.

[68] P. van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1):29–48, November 2002.

[69] Don van Ravenzwaaij, Pete Cassey, and Scott D. Brown. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, 25(1):143–154, 2018.

[70] Romain Veltz. BifurcationKit.jl, July 2020.

[71] Linda Wanika, Joseph R. Egan, Nivedhitha Swaminathan, Carlos A. Duran-Villalobos, Juergen Branke, Stephen Goldrick, and Mike Chappell. Structural and practical identifiability analysis in bioengineering: a beginner's guide. *Journal of Biological Engineering*, 18(1):20, March 2024.

[72] Franz-Georg Wieland, Adrian L. Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, March 2021.

[73] WorldData.info. Population growth in Singapore, 2025.

[74] Hyun Mo Yang. The basic reproduction number obtained from Jacobian and next generation matrices – A case study of dengue transmission modelling. *Biosystems*, 126:52–75, December 2014.

[75] Jie Yu, Huimin Wang, Miaoshuang Chen, Xinyue Han, Qiao Deng, Chen Yang, Wenhui Zhu, Yue Ma, Fei Yin, Yang Weng, Changhong Yang, and Tao Zhang. A novel method to select time-varying multivariate time series models for the surveillance of infectious diseases. *BMC Infectious Diseases*, 24(1):832, August 2024.

[76] Hsiang-Yu Yuan and Colin Blakemore. The impact of contact tracing and testing on controlling COVID-19 outbreak without lockdown in Hong Kong: An observational study. *The Lancet Regional Health: Western Pacific*, 20:100374, January 2022.

Appendix A

Code and Data

The code and datasets used throughout the paper can be accessed via the following repository: `https://github.com/i-adamson/Epidemic_Parameter_Estimation/blob/main/README.md`.