Applications of Nonlinear Systems of Ordinary Differential Equations and Volterra Integral Equations to Infectious Disease Epidemiology

by

Emmanuel J. Morales Butler

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee:

Carlos Castillo-Chavez, Co-Chair
Juan P. Aparicio, Co-Chair
Erika T. Camacho
Yun Kang

ARIZONA STATE UNIVERSITY

May 2014

ABSTRACT

In the field of infectious disease epidemiology, the assessment of model robustness outcomes plays a significant role in the identification, reformulation, and evaluation of preparedness strategies aimed at limiting the impact of catastrophic events (pandemics or the deliberate release of biological agents) or used in the management of disease prevention strategies, or employed in the identification and evaluation of control or mitigation measures. The research work in this dissertation focuses on: The comparison and assessment of the role of exponentially distributed waiting times versus the use of generalized non-exponential parametric distributed waiting times of infectious periods on the quantitative and qualitative outcomes generated by Susceptible-Infectious-Removed (SIR) models. Specifically, Gamma distributed infectious periods are considered in the three research projects developed following the applications found in [28, 29, 31, 22, 32, 33, 37, 38]. **i)** The first project focuses on the influence of input model parameters, such as the transmission rate, mean and variance of Gamma distributed infectious periods, on disease prevalence, the peak epidemic size and its timing, final epidemic size, epidemic duration and basic reproduction number. Global uncertainty and sensitivity analyses are carried out using a deterministic Susceptible-Infectious-Recovered (SIR) model. The quantitative effect and qualitative relation between input model parameters and outcome variables are established using Latin Hypercube Sampling (LHS) and Partial rank correlation coefficient (PRCC) and Spearman rank correlation coefficient (RCC) sensitivity indices. We learnt that: For relatively low (R0 close to one) to high (mean of R0 equals 15) transmissibility, the variance of the Gamma distribution for the infectious period, input parameter of the deterministic age-of-infection SIR model, is key (statistically significant) on the predictability of the epidemiological variables such as the epidemic duration and the peak size and timing of the prevalence of infectious individuals and therefore, for the predictability these variables, it is preferable to utilize a nonlinear system of Volterra integral equations,

rather than a nonlinear system of ordinary differential equations. The predictability of epidemiological variables such as the final epidemic size and the basic reproduction number are unaffected by (or independent of) the variance of the Gamma distribution for the infectious period and therefore for the choice on which type of nonlinear system for the description of the SIR model (VIE's or ODE's) is irrelevant. Although, for practical proposes, with the aim of lowering the complexity and number operations in the numerical methods, a nonlinear system of ordinary differential equations is preferred. The main contribution lies in the development of a model based decision-tool that helps determine when SIR models given in terms of Volterra integral equations are equivalent or better suited than SIR models that only consider exponentially distributed infectious periods. **ii)** The second project addresses the question of whether or not there is sufficient evidence to conclude that two empirical distributions for a single epidemiological outcome, one generated using a stochastic SIR model under exponentially distributed infectious periods and the other under the non-exponentially distributed infectious period, are statistically dissimilar. The stochastic formulations are modeled via a continuous time Markov chain model. The statistical hypothesis test is conducted using the non-parametric Kolmogorov-Smirnov test. We found evidence that shows that for low to moderate transmissibility, all empirical distribution pairs (generated from exponential and non-exponential distributions) for each of the epidemiological quantities considered are statistically dissimilar. The research in this project helps determine whether the weakening exponential distribution assumption must be considered in the estimation of probability of events defined from the empirical distribution of specific random variables. **iii)** The third project involves the assessment of the effect of exponentially distributed infectious periods on estimates of input parameter and the associated outcome variable predictions. Quantities unaffected by the use of exponentially distributed infectious period within low transmissibility scenarios include, the prevalence peak time, final epidemic size, epidemic duration and basic reproduction number and for

high transmissibility scenarios only the prevalence peak time and final epidemic size. An application designed to determine from incidence data whether there is sufficient statistical evidence to conclude that the infectious period distribution should not be modeled by an exponential distribution is developed. A method for estimating explicitly specified non-exponential parametric probability density functions for the infectious period from epidemiological data is developed. The methodologies presented in this dissertation may be applicable to models where waiting times are used to model transitions between stages, a process that is common in the study of life-history dynamics of many ecological systems.

This piece of work is dedicated to my parents Migdalia Butler Pérez and Jesús Morales Martínez for giving me two very important gifts: "First, the gift of my life and second, the gift of freedom to live it. This dissertation is also dedicated to my brother, Álvaro Morales Butler, for taking great care of them." Thanks, Maestro Facundo Cabral for these words.

ACKNOWLEDGMENTS

I want to thank to a subset of my friends, life mentors and role models for all your support and more: Álvaro Morales Butler, Dalvin Daniel Méndez-Hernández! and Limari Cora Huertas and Daniela and Oriol, Juan Diego Avilés Hernández, Mayteé Cruz-Aponte! Arlene Evangelista Morales, Ciera Rosario Duran, José Manuel Vega and family, Kamal Kevin Barley, Kehinde Rilwan Salau, Bichara Derdei, Oyita Udiani, Oscar Patterson Lomba, Reynaldo Castro-Estrada, Javier Gutiérrez, Mariana Chanampa, Sunmi Lee and Claire Jiwon Eum, Sang-Uk Ryu, Yunhwan Kim, Karen Ros-Soto, Anuj Mubayi, Joaqun Rivera, Xiaohong Wang! (Thanks$^{\infty}$) Fabio Sánchez, Gerardo Chowell, Michael Golinski! Marco Herrera, Sharon Harlan! Mustafa Erdem, Xiaoxia Wang, Kamuela Yong, Muntaser Safan, Zhilan Feng, Fred Brauer! Laura Serrano, Elisabeth Luquez, Margaret Murphy-Tillis, Sherry Woodley, Elsie Morales! Francisco Medina Rivera, Martin Engman! Juan Pablo Aparicio (my "pamingano") and family!!! Carlos Castillo-Chavez! Oscar López Rivera! and to all those "beautiful faces of my (black) people" (Thanks for the song Catalino "Tite" Curet Alonso) of the community of Montadero, the corner in Puerto Rico where I belong.

The financial support came from:

- Alfred P. Sloan Scholarship
- The More Graduate Education at Mountain States Alliance (MGE@MSA)
- The NSF Cooperative Agreement HRD-0602425 (Louis Stokes Alliances for Minority Participation Program WAESO Western Alliance to Expand Student Opportunities Biodesigned and Sustainable Bridges to the Doctorate Phase IV)

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Broad problem statement

Mathematical models are a critical research component in the life and social sciences. Hence the importance of addressing the *broad problem* of evaluating the impact of model assumptions and derived relationship on observed conclusions or hypotheses that emerge from qualitative and quantitative model-based studies in the life and social sciences. For example, model assessment provides useful insights that help modelers re-design or revise research protocols that may prevent help identify limitations or flaws. Systematic model assessment limits the likelihood of reporting incomplete or misguiding conclusions. The case of Hurricane Katrina (2005), the deadliest Hurricane in the past 90 years of history in the United States of America, helps exemplify our perspective. The impact of Katrina is still being felt. Katrina's death toll of about 1,500 people (see [8]) and the associated economic property loss of around 96 billions that followed its aftermath (see [9]) highlight the importance of assessing uncertainty. Models forecasted accurately Katrina's path (approximately) 56 hours ahead of time, that is, we had some of the information needed to execute emergency plans (see [10]). However, lack of preparadeness and delayed response from local and federal governments, worsened the consequences of this disaster (see [11, 12]).

1.2 Particular problem statement

Systematic model assessment naturally also plays a critical role in the field of infectious disease epidemiology, particularly within the study of a disease's transmission dynamics and control, the kind of dynamics and interventions that take place over multiple levels of organization and across highly distinct spatial and temporal scales. Hence, the systematic assessment of the robustness of model's outcomes plays a significant role in the proposition, reformulation, and evaluation of preparedness strategies aimed at limiting the impact of catastrophic events (pandemics or the deliberate release of biological agents) or in the indentification and management of disease prevention, control or mitigation measures. An important step in model assessment, involves the systematic study of the effects of modeling assumptions and such an assessment is the main theme of this dissertation.

The *conclusions* made from qualitative and or quantitative epidemiological studies of infectious diseases are derived from: 1) an assessment of the effectiveness of implemented intervention strategies and 2) the optimal allocation of limited resources. Such assessment and resource re-distribution depend on a set of *outcome (prediction) variables* that include, for example, the basic reproduction number or the final epidemic size (See **Figure 1.1** for more examples). These epidemiological quantities are derived from compartmental epidemiological models that approximately describe the underlying transmission dynamic of infectious disease epidemics. Hence we must deal with two sources of uncertainty: *intrinsic uncertainty*, that is, the uncertainty associated with observable data (incidence of infected individuals, mortality and morbidity) and *structural sensitivity* (for a detail explanation on this topic please refer to [37]). Specifically, the research in this dissertation focuses on comparing and assessing the impact of the use of exponential or non-exponential infectious period distributions on the quantitative and qualitative outcomes generated by Susceptible-Infectious-Removed (SIR) models. The novelty and usefulness of our research

relies on two facts: the *applicability* of the methodologies introduced as seen by the research in all chapters and the *flexibility* of the methodology. Our approach allows for the use and implementation of specified (arbitrary) parametric distributions for the infection period. The Gamma probability density function is used to illustrate all the methodologies developed since the Erlang distribution (a special case) is the most studied non-exponential distribution for the infectious period (See [28, 29, 31, 22, 32, 33, 37, 38] for examples).

1.2.1 Background

Successful stories of modeling of infectious diseases

What is the role modeling assumptions on the results and conclusions of qualitative or quantitative studies? How will they impact model-generated based policies aimed at improving the quality of public health policy? How can model-generated based policies guide and help decision makers? An example of this can be traced back to research carried out at the beginning of the twentieth century namely, pioneer and seminal work of Sir Ronald Ross. He introduced a Malaria transmission mathematical model in order to show that lowering the vector population below a particular threshold was enough (theoretically) for controlling this deadly disease (see [7] and references there in). Nearly ninety years after the formulation of the Ross' Malaria model, Edward H. Kaplan and others studied the dynamics of contaminated needles with the HIV as vectors in assessing the future of the HIV epidemic among populations of intravenous drug users (see [4]). Kaplan's model was used as the core of a methodology to evaluate the effectiveness of the first legal needle exchange program implemented against the HIV epidemic among the population of injecting drug users (see [5, 6]). Evidence of a significant reduction (of about 33 percent) in the transmission rate of HIV among injecting drug users was found, thus changing the public health perspective on the effectiveness of needle exchange programs, which were later

decriminalized and extended to other states of the union (see [4]).

Why the need of systematic model assessment?

The need of systematic model assessment in the study of the dynamics of infectious diseases over multiple levels of organization is supported by past experiences. The 1918-1919 influenza pandemic (known as the Spanish flu), the most devastating in recorded history, had an estimated death toll in the 20 to 100 million range [13, 14] with an estimated case fatality in the 2-6 percent levels [15, 16]. Most recently, the highly pathogenic avian influenza (HPAI) virus subtype H5N, first isolated in 1996 from farmed goose in Guangdong Province, China, resulted in 628 humans known infected cases from direct contact with infectious birds. From these 374/628 died, leading to a case fatality rate of 59.6% [18]. From these examples, we see that systematic model assessment is critical not only because of the loss of life but also because of the economic consequences linked to epidemics and pandemics. For example the cost of dealing with foot and mouth disease in Britain was estimated to be in the order of billions! Model-generated predictions suggest that the impact from an avian influenza pandemic could be in the order of billions to trillions [17].

1.3 Thesis outline

The introduction of the thesis and research problem statement are provided in Chapter one. In Chapter two the fundamental concepts needed for the understanding of the core and common frameworks used throughout the subsequent chapters are introduced. Chapter three focuses on studying the influence of input model parameters on outcome (prediction) variables within a deterministic compartmental epidemic model under Gamma distributed infectious period distributions. The stochastic aspect is explored in Chapter four via a continuous time Markov chain model. The influence on model parameter estimates under standard modeling assumptions and within a simple compartmental epidemic model

is assessed in Chapter five. Last chapter (Chapter six) provides an overall discussion and conclusions on the main contributions and results of the research in this thesis.

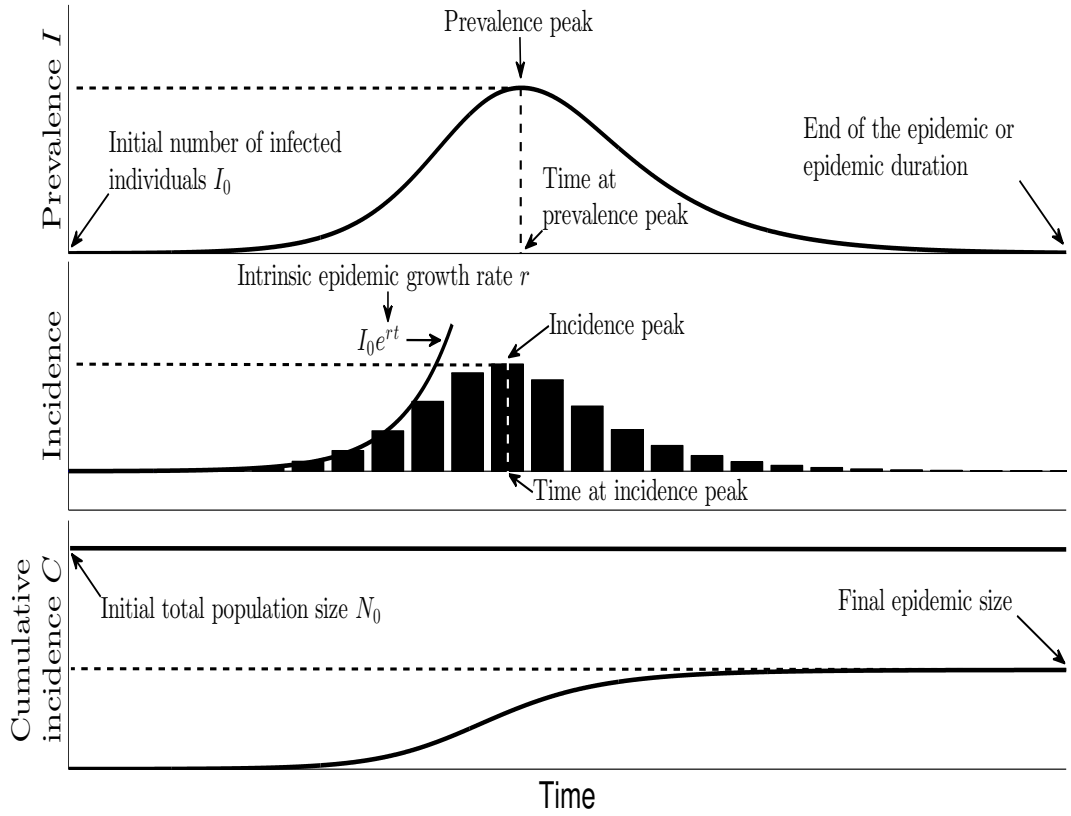


Figure 1.1: The time evolution of the prevalence of infectious individuals (top panel) and the time evolution of the cumulative incidence (bottom panel).

Chapter 2

FUNDAMENTAL CONCEPTS AND CORE FRAMEWORKS

2.1 Global framework

The landscape where an infectious disease emerges and spreads is usually a network of (sub)populations (*i.e.* a metapopulation) of individuals (humans, or poultry, or cattle, etc.). In general, the contact structure (or network) among individuals (between and within populations) and the the size of each population are dynamic and heterogenous throughout time. For example, in Switzerland the poultry size in a farm (considering commercial and non-commercial) on average is of 1,317 poultry, but it may vary from 12 to 3,807 poultry, with around 97 percent of all poultry farms having neighbor farms within one kilometer of radius (see [3]) and the whole provides an example of a metapopulation of poultry. The epidemiological models considered throughout this thesis are intended to describe the *dynamics, especially the transient dynamic* of an epidemic at either the population level when the population (or subpopulation) size(s) is (are) sufficiently large so that the assumption of homogenous mixing approximately holds.

2.2 Model description

Compartmental epidemiological models are common components used to build metatpopulation mathematical epidemiological models. The basic models consider epidemiological classes of individuals that include: *Susceptible* individuals, represented by the letter (S), may acquired the infection or disease via a ("successful") contact with an infectious individual. *Infectious* individuals are represented by the letter (I) and are the ones with the ability to spread the disease. After an infectious period, infectious individu-

als progress towards the *removed* (or *recovered*) disease stage (R). In this work we do not consider vital dynamics, that is, birth and deaths are neglected. This simple compartmental epidemiological model is known as the *Susceptible-Infectious-Recovered* (or *SIR*) model (see **Figure 2.1**). The following are the usual fundamental implicit assumptions considered in the basic SIR model (adapted from the seminal article by W. O. Kermack and A. G. McKendrick, 1927 [24]):

• (**A1**: Absence of spatiality) All individuals are in contact with each other, mathematically as a complete graph;

• (**A2**: Homogenous transmissibility) All infected individuals have the same potential to transmit the disease;

• (**A3**: Homogenous vulnerability) All susceptible individuals have the same chance of acquiring the infection;

• (**A4**: Constant rate) The transmission rate denoted by β is constant throughout the whole epidemic duration.

• (**A5**: Vital dynamic is neglected) The total population N is constant throughout the whole epidemic duration.

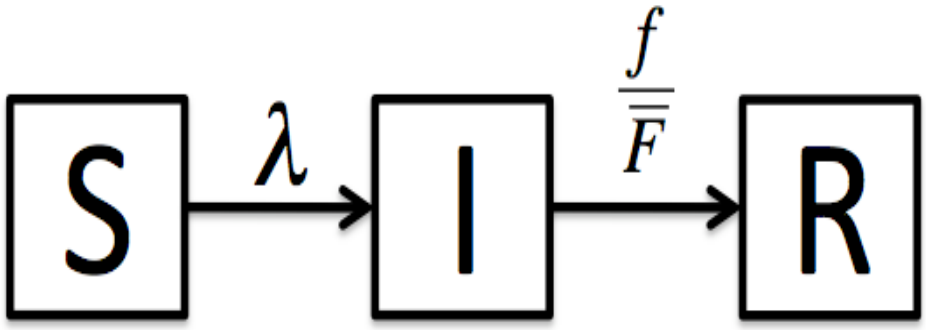


Figure 2.1: Schematic diagram of a simple age-of-infection Susceptible-Infectious-Removed model with force of infection λ and instantaneous transition rate for the infectious period $\frac{f}{\overline{F}}$.

2.2.1 Fundamental concepts and deterministic formulation

The basic and widely used deterministic Susceptible-Infectious-Recovered compartmental model is described by the nonlinear system of ordinary differential equations

$$\frac{dS}{dt}(t) = -\frac{\beta}{N} I(t)S(t) \tag{2.1}$$

$$\frac{dI}{dt}(t) = \frac{\beta}{N} I(t)S(t) - \gamma I(t) \tag{2.2}$$

$$\frac{dR}{dt}(t) = \gamma I(t) \tag{2.3}$$

where is implicitly assumed an exponential distributed infectious period. In other words, it is assumed that the infectious period is a random variable with probability density function $f(s) = \gamma e^{-\gamma s}$. In this case the recovery rate γ (also known as the **failure** or **hazard rate**) is the inverse of the mean infectious period.

But an exponential distribution for the infectious period is a far from realistic choice: In most cases one expects a bell-shaped distribution (see Figure 2.2).

For a general distributed infectious period with **probability density function** $f(s)$ (where s, the **age-of-infection**, is the time elapsed since infection), the probability that an individual remains infected after a time s is given by the **survivor function** $\bar{F}(s) = 1 - F(s)$ where $F(s)$ is the **cumulative distribution function** $F(s) = \int_0^s f(t)dt$.

The **hazard rate** is now a function of the age of infection (except for the case of exponential distribution), $\gamma(s) = \frac{f(s)}{\bar{F}(s)}$.

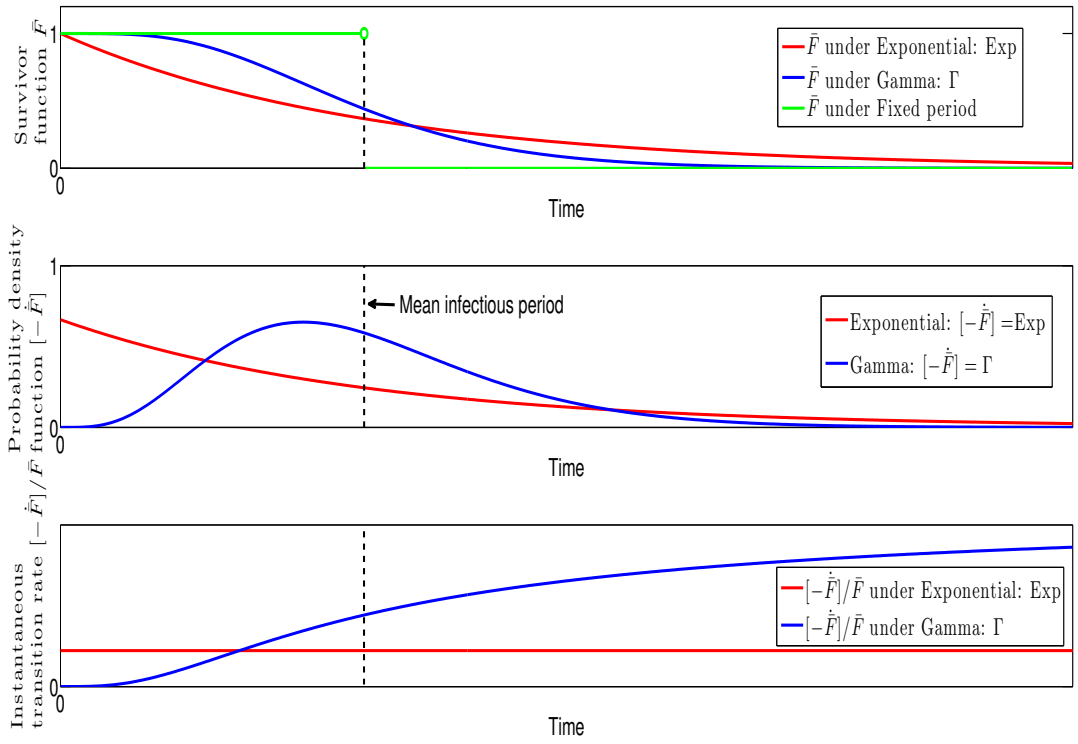


Figure 2.2: Examples of survivor functions (top panel), probability density functions (middle panel) and hazard rate functions (bottom panel) for the infectious period.

Assuming the standard force of infection the age-of-infection Susceptible-Infected-Removed model becomes the nonlinear system of Volterra integral equations:

$$S(t) = S_0 - \int_0^t \frac{\beta}{N} I(s)S(s)ds \tag{2.4}$$

$$I(t) = I_0\bar{F}(t) + \int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)ds \tag{2.5}$$

$$R(t) = R_0 + I_0(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} I(s)S(s)(1-\bar{F}(t-s))ds = N - S(t) - I(t). \tag{2.6}$$

In the above system N is represents the total population size, in this case constant, since $N(t) = S(t)+I(t)+R(t) = S_0+I_0+R_0 = N_0$. Where $S(0) = S_0, \quad I(0) = I_0, \quad R(0) = R_0$ are the corresponding initial conditions for S, I and R. The rational of equation I is as follows: the first term, accounts for the initial number of infectious individuals I_0 (index case) whom at time t still remain as infectious accordingly to the survivor function $\bar{F}$, the

whole term then is $I(0)\bar{F}(t)$; The second term accounts for those susceptible individuals S that acquired the disease with a force of infection $\frac{\beta}{N}I$ at any time s between the initial time of the epidemic 0 and the epidemic time t and whom at an age of infection $t-s$ still remain as infectious accordingly to the survivor function $\bar{F}$, the whole term is given by $\int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)ds$.

The nonlinear system (2.4)-(2.6) can be rewritten as an equivalent nonlinear system of Volterra integro-ordinary differential equations through a differentiation under the integral sign:

$$\dot{S}(t) = -\frac{\beta}{N}I(t)S(t) \tag{2.7}$$

$$\dot{I}(t) = \frac{\beta}{N}I(t)S(t) - \left(I_0[-\dot{\bar{F}}(t)] + \int_0^t \frac{\beta}{N} I(s)S(s)[-\dot{\bar{F}}(t-s)]ds\right) \tag{2.8}$$

$$\dot{R}(t) = I_0[-\dot{\bar{F}}(t)] + \int_0^t \frac{\beta}{N} I(s)S(s)[-\dot{\bar{F}}(t-s)]ds = -(\dot{S}(t)+\dot{I}(t)). \tag{2.9}$$

We define the **point prevalence** as the total number of infectives at time t (Gerstman, B.B. 2003), that is $I(t)$. Point prevalence is not a variable we can usually measure. The **incidence** of the disease, on the other hand, is what commonly is reported and is defined as the number of new cases on some period of time T (a week, a month, or a year, depending on the disease) and in our model can be obtained as

$$\int_t^{t+T} \frac{\beta}{N} I(s)S(s)ds.$$

In this thesis we will consider Gamma distributed infectious periods. Gamma distribution presents a series of advantages. We can set its mean and variance in an independent way. Furthermore when the square of the mean equals the mean, the Gamma distribution become the exponential distribution and model (2.7)-(2.9) reduces to the basic ODE model ((2.1)-(2.3), see **appendix A**). Gamma distribution has two parameter known as the shape

(k) and scale (θ) parameters and it is defined as

$$f(t;k,\theta) = \begin{cases} \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-\frac{t}{\theta}} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases} \tag{2.10}$$

and

$$(k,\theta) = \left(\frac{\tau^2}{Var}, \frac{Var}{\tau} \right) \tag{2.11}$$

from where $\theta = \frac{\tau}{k}$.

Another useful feature of the Gamma distribution is that for positive integer values for the shape parameter k it reduces to the Erlang distribution. In this case the nonlinear model of Volterra integral equations (2.4)-(2.6) can be rewritten as an equivalent, but larger dimensional nonlinear system of ordinary differential equations. This is done via a standard method called *linear chain trickery* (see **appendix B** for details on the derivation).

Finally in the limiting case when the shape parameter tends to infinity the Gamma distribution converges in distributional sense to the *Dirac delta function*. In this last case, a nonlinear system of discrete delay differential equations is obtained from the original system of Volterra integral equations (in (2.4)-(2.6)) (see **appendix C** for details on the derivation).

2.2.2 Well-posedness of the model

The biological and mathematical well-posedness of the model above is studied by establishing the conditions for which the solutions of equations in (2.4)-(2.6) exist, are unique, non-negative and bounded.

To facilitate the presentation of the theorems below, first, the following terms are defined: Let P be a set with non negative elements, described by

$$P = \{y \in \mathbb{R} : 0 < y \leq N\} \times \{y \in \mathbb{R} : 0 \leq y < N\}^2.$$

$$\vec{x} : [0,a] \to P, \text{ where } \vec{x}(t) = \begin{bmatrix} S(t) \\ I(t) \\ R(t) \end{bmatrix}, \quad \vec{x}(0) = \begin{bmatrix} S(0) \\ I(0) \\ R(0) \end{bmatrix} = \begin{bmatrix} S_0 \\ I_0 \\ R_0 \end{bmatrix} = \vec{x}_0 \in P,$$

$$\vec{h} : [0,a] \to P, \text{ where } \vec{h}(t) = \begin{bmatrix} S_0 \\ I_0\bar{F}(t) \\ R_0 + I_0(1-\bar{F}(t)) \end{bmatrix} \text{ and}$$

$$\vec{g} : [0,a]^2 \times P \to \{y \in \mathbb{R} : -\beta N < y \leq 0\} \times \{y \in \mathbb{R} : 0 \leq y < \beta N\}^2, \text{ where}$$

$$\vec{g}(t,s,\vec{x}(s)) = \begin{bmatrix} -\frac{\beta}{N} I(s)S(s) \\ \frac{\beta}{N} I(s)S(s)\bar{F}(t-s) \\ \frac{\beta}{N} I(s)S(s)(1-\bar{F}(t-s)) \end{bmatrix}.$$

The model in (2.4)-(2.6) can now be writing in vector form as:

$$\vec{x}(t) = \vec{h}(t) + \int_0^t \vec{g}(t,s,\vec{x}(s))ds \quad \vec{x}(t_0) = \vec{x}_0. \tag{2.12}$$

The following theorems of (local) existence and uniqueness of solution for the model in (2.12) are taken and adapted from the classical books by R. K. Miller [66] (chapter one) and F. Brauer and J. A. Nohel [70] (chapter three).

Local existence

Theorem 1. Suppose $\vec{h}$ is a continuous function defined on an interval $0 \leq t \leq a$. Suppose $\vec{g}$ and $\frac{\partial \vec{g}}{\partial x_j}(j = 1,2,3)$ are continuous in the region:

$$R = \{(t,s,\vec{x}) : 0 \leq s \leq t \leq a \text{ and } |\vec{x}(t) - \vec{h}(t)| \leq b\}.$$

Then there exist $\alpha > 0$ and a continuous solution of equation (2.12) on $[0,\alpha]$.

The standard and core method of proof for the local existence, is called, *Picard successive approximations*. The proof of Theorem 1 can be follow line by line from the scalar case of Theorem 8.1 in the book by R. K. Miller [66].

Uniqueness

Theorem 2. Suppose $\vec{h}$ is a continuous function defined on an interval $0 \leq t \leq a$. Suppose $\vec{g}$ and $\frac{\partial \vec{g}}{\partial x_j}(j = 1,2,3)$ are continuous in the region:

$$R = \{(t,s,\vec{x}) : 0 \leq s \leq t \leq a \text{ and } |\vec{x}(t) - \vec{h}(t)| \leq b\}.$$

Then there exist $\alpha > 0$ and a unique continuous solution of equation (2.12) on $[0,\alpha]$.

The main tool of proof for uniqueness, is called, *Gronwall inequality*. The proof of Theorem 2 can be follow line by line from the scalar case of Theorem 8.1 in the book by R. K. Miller [66].

Positive solutions

Theorem 3. The solution S of the equation (2.4) is a strictly positive function, while the solutions I and R of equations (2.5) and (2.6), respectively, are non negative functions on their domain of existence.

Proof: Recall that $0 < S_0 \leq N_0$. Define the function $G_1(s) = -\frac{\beta}{N} I(s)$, then $S(t) = S_0 \exp \int_0^t G_1(s)ds$ and thus $S(t)$ a strictly positive function as long as it exists. Recall that $0 \leq I_0 < N_0$. Define the function $G_2(s) = \frac{\beta}{N} S(s) - \frac{1}{I(s)} \Big(I_0 f(s) + \int_0^s \frac{\beta}{N} I(x)S(x)f(s-x)dx \Big)$, then $I(t) = I_0 \exp \int_0^t G_2(s)ds$ and thus $I(t)$ a non negative function as long as it exists. Recall that $0 \leq R_0 < N_0$. Define the function $G_3(s) = \frac{1}{N_0 - (S(s)+I(s))} \Big(I_0 f(s) + \int_0^s \frac{\beta}{N} I(x)S(x)f(s-x)dx \Big)$, then $R(t) = R_0 \exp \int_0^t G_3(s)ds$ and thus $R(t)$ a non negative function as long as it exists.

Boundedness

Theorem 4. The solutions of equations (2.4)-(2.6), S, I and R are bounded on their domain of existence, as follows:

$$0 < S_\infty \leq S \leq N, \quad 0 \leq I < N, \quad 0 \leq R < N.$$

Proof: Since I and R are non negative solutions, then $S = N - (I + R) \leq N$, Define $S_\infty = \lim_{t\to\infty} S(t)$. From the final size relation (see **appendix L** for its derivation), $S_\infty \neq 0$. Since $S(t)$ is strictly positive monotonically decreasing function, then $S(t)$ is bounded below away from zero by S_∞. Thus $0 < S_\infty \leq S \leq N$. From Theorem 3, I is bounded below by zero. From the equation for I:

$$\begin{aligned} I(t) &= I_0\bar{F}(t) + \int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)ds \\ &\leq I_0 + \int_0^t \frac{\beta}{N} I(s)S(s)ds \\ &= I_0 + \int_0^t [-\dot{S}(s)]ds \\ &= S_0 + I_0 - S(t) = N_0 - (S(t) + R_0) < N_0 = N. \end{aligned}$$

Thus $0 \leq I < N$. From Theorem 3, R is bounded below by zero. By definition $R = N - (S+I) < N$. Thus $0 \leq R < N$.

2.2.3 Transient and long term dynamics of the SIR model

Figure 2.3 illustrates the transient and long term dynamic of the solutions of model (2.4)-(2.6) in the (S,I)-plane. Let $\mathscr{R}_0 = \beta\tau$, defined in details in chapter three. If $\mathscr{R}_0\frac{S_0}{N}$ is less or equal than the epidemic threshold one, then the epidemic does not occur and the solution of equation (2.5), $I(t)$, decreases from I_0 to extinction (zero) in the long term (as $t \to \infty$). Otherwise, if $\mathscr{R}_0\frac{S_0}{N} > 1$, then the epidemic does occur and I initially increases from I_0, reaches a unique maximum number of infected individuals and then decreases to extinction (zero) in the long term (as $t \to \infty$). The solution of equation (2.4) for the susceptible individuals, $S(t)$, is a non-increasing function bounded below away from zero to its limit S_∞ as $t \to \infty$. In the long term (as $t \to \infty$), all the steady-states solutions $(\frac{S_\infty}{N}, \frac{I_\infty}{N}) = (\frac{S_\infty}{N}, 0)$ are neutrally stable if $\mathscr{R}_0\frac{S_0}{N} < 1$ and neutrally unstable if $\mathscr{R}_0\frac{S_0}{N} > 1$. As shown graphically in the Figure 2.3, all the solutions $(S(t), I(t))$ of equations (2.4) and (2.5) are contained on an epidemiologically feasible region (positively invariant) $T = \{(S,I) : S >$

$0, I \geq 0, S+I \leq N\}$ color coded as yellow. The mathematical formalism of these qualitative results can be found in theorem 5.1 in the seminal paper by H. W. Hethcote, [71].

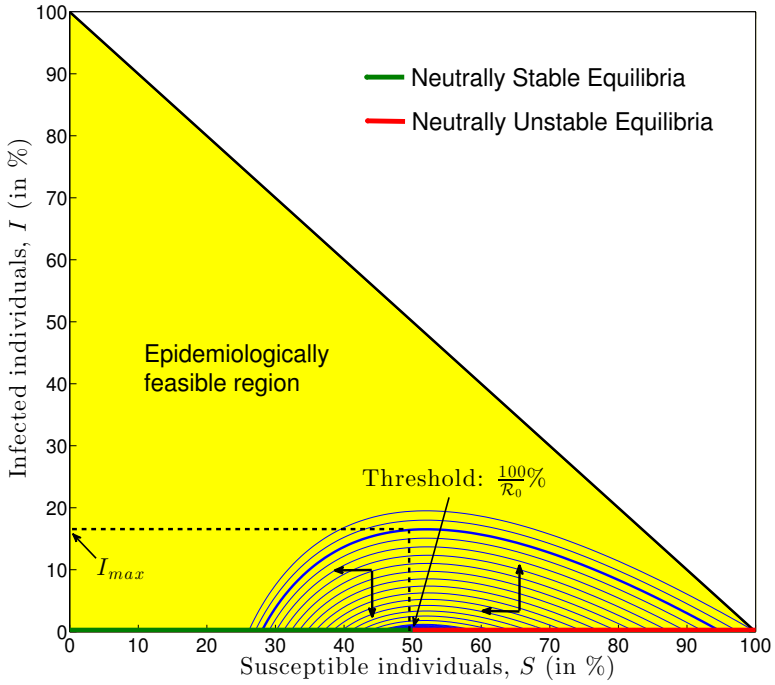


Figure 2.3: Qualitative dynamic of the age-of-infection Susceptible-Infectious-Removed model in (2.4)-(2.6).

2.3 Core numerical schemes

Throughout this thesis we developed a 4th-stage Volterra-Runge-Kutta formula of Pouzet type scheme which was used to solve numerically the nonlinear system of Volterra integral equations (see **appendix D**). We also developed a 4th-stage VIODE-Runge-Kutta formula of Pouzet type scheme used to solve numerically the nonlinear system of Volterra integro-ordinary differential equations (see **appendix E**). This type of numerical schemes are described in details in the book "The numerical solution of Volterra equations" by H. Brunner and P. J. Van der Houwen (see [65]).

2.4 Discussion

Although the simplest form of the SIR model reviewed here is considered as a phenomenological model, still can be useful (with careful) for quantitative purposes.

Chapter 3

GLOBAL UNCERTAINTY AND SENSITIVITY ANALYSES ON A SIMPLE AGE-OF-INFECTION SUSCEPTIBLE-INFECTIOUS-RECOVERED MODEL

3.1 Introduction

In the literature on the effect of non-exponential distributions for the latent and or infectious periods on disease dynamics the most commonly used distribution is the Erlang distribution, a special case for the gamma distribution, obtained when the shape parameter is restricted to take only on positive integer values (see [28, 29, 31, 22, 32, 33, 37, 38]). The choice of the Erlang distribution, allows to replace Volterra integro-ordinary differential system like (2.7)-(2.9) by a system of ordinary differential equations. It is rare to find in the literature, explicit efforts to account for the effect of the variability of the distributions via the use of different values for the shape parameter (see [29, 31, 22, 32, 33, 37]). In this chapter we move beyond the use of the Erlang distribution and carry out a general uncertainty and sensitivity analyses when the distribution used is a Gamma since we are particular interested in the study of the effects of variations in the variance of the Gamma distribution. It is worth recalling that when the variance is the square of the mean (shape parameter $k = \frac{\tau^2}{Var} = 1$) the gamma distribution is an exponential distribution and the Volterra integro-ordinary differential system (2.7)-(2.9) turns into the basic ODE model (2.1)-(2.3). Hence, our used of an extended framework and its analysis allows not only to explore the effect of variability on the outputs of the model but also the study of effects associated with model selection.

3.1.1 Research questions of interest

In order to study the influence of parameters uncertainty and model structure on disease dynamics we consider several *outcome variables* of epidemiological interest including, the peak of the prevalence of infectious individuals; time at which the peak size of the prevalence occurs; final epidemic size; epidemic duration and the basic reproduction number (see **Figure 1.1**).

Some of questions that we would like to address are:

Question one: What is the qualitative relation between the variance of the Gamma probability density function for the infectious period and the outcome variables?

Question two: What is the level of (overall) prediction precision of the compartmental epidemiological model (see **Figure 2.1**) with respect to the outcome variables?

Question three: What is the level of importance of the variance of the Gamma probability density function for the infectious period on the prediction precision of the outcome variables?

The application of a well known methodology for a global uncertainty and sensitivity analyses via the Latin Hypercube Sampling (LHS) and the estimation of two sensitivity indices to a simple age-of-infection Susceptible-Infectious-Recovered model described via a nonlinear system of Volterra integral equations in 2.4-2.6 is the *novel* contribution of this chapter.

This analysis provides:

i) Qualitative relations between the outcome variables (or prediction variables), derived from an epidemiological model, and the input model parameters, with their associated (statistical) significance (P-values),

ii) A quantification for the influence (quantitative effect) of input variables estimates on the predictability of outcome variables,

iii) An overall model prediction precision.

The knowledge generated by the analysis serves as a *research decision tool* that helps determine the type of a nonlinear system that it is more appropriate or suitable for modeling SIR dynamics. That is, do we use Volterra integral equations or ordinary differential equations? This decision is made by identifying **whether or not the variance of the infectious period distribution is a key input parameter for the predictability of the outcome variables of interest**.

Surprisingly, a *general* global uncertainty and sensitivity analyses addressing these questions had *not* been proposed yet in the field of infectious disease modeling, despite the fact that the non-exponential infectious period distributions are most likely the norm. Not carrying out an uncertainty quantification on the outcome variables as a function of the infectious period distribution may lead to serious errors or even meaningless results. It is clear, that any field in biology that relies on compartmental models can may use of the methodology presented in this chapter and this dissertation.

3.1.2 Background

A general survivor function is incorporated explicitly in a Susceptible-Infected-Removed type model via a system of nonlinear Volterra integral equations or Volterra integro-ordinary differential equations (i.e., continuous distributed delay type system). These type of equations were introduced in 1896 by Vito Volterra in a series of three papers [23]. In 1927, Voterra equations were used by W. O. Kermack and A. G. McKendrick to introduce a general framework that allowed variable infectivity levels and removal rates on an epidemiological model [24]. Empirical evidences and first attempts to estimate the distri-

bution of the incubation period of some infectious diseases using a log-normal probability density function were made in 1950, 1952 and 1966 in a series of papers by P. E. Sartwell [25, 26, 27]. In 1964, N. T. J. Bailey obtained a deterministic SEIR model with general χ^2 distributed latent and infectious periods (a special case of the Gamma distribution) as a result of a derivation of the equations for the stochastic means from a multidimensional Markov process [28]. Later in 1980, D. Anderson and R. Watson considered the general model formulated by N. T. J. Bailey in 1964 to assess the effect of the shape parameter of the Gamma distribution for the infectious period on the distribution of the final epidemic size [29]. In this work the authors also established a relation between the basic reproduction number with the intrinsic epidemic growth rate and the corresponding shape parameters from both distributions [29]. Recently, the role of the Gamma probability density function in the estimation of key epidemiological distributions have been crucial, as was the case of 2003 SARS epidemic in Hong Kong, illustrated in the work of C. A. Donnelly *et al.* [30]. In this work, the authors observed that the shape of the estimated distributions, for some disease stages, varied among age-groups and on which window of time was used for the estimations, since at different windows of time different public health interventions were applied [30]. The general relation between the basic reproduction number and the intrinsic epidemic growth rate established in 1980 by D. Anderson and R. Watson, was used in 2005 by H. J. Wearing *et al.* to study the effect of Gamma distributed latent and infectious periods on the estimates of the basic reproduction number, and other epidemiological quantities like the prevalence and incidence of infected individuals [31]. Based on the values taken by epidemiological quantities like the final epidemic size, peak size of the epidemic intensity and the control reproduction number, Z. Feng *et al.* in two papers published in 2007, illustrate that assuming different probability density functions for the latent and infectious periods on a SEIR model with quarantine and isolation as public health strategies lead to inconsistent conclusions about which strategy is more effective [22, 32]. In 2008,

P. Yan assessed the impact that estimates of the intrinsic epidemic growth from data have on the magnitude of the basic reproduction number by applying two formulas that relate these key quantities; one was established by D. Anderson and R. Watson in 1980 and it assumed a Gamma distributed latent and infectious periods and other one was established by P. Yan, in this work, and it assumed inverse-gaussian distributed latent and infectious periods [33]. In 2007 in a master thesis by C. K. Yang and later in a 2008 paper in collaboration with F. Brauer, the authors illustrated a way to calculate the basic and control reproduction numbers using multiple stage age of infection models [34, 35]. C. K. Yang and F. Brauer observed that the basic reproduction number does depend on the mean of the distribution of the diseases stages and not on the distribution, while the control reproduction number does depend on the distribution of the diseases stages [34, 35]. In the same year 2008, F. Brauer established a simpler way of deriving the final size relation, in comparison to how was calculated in 1927 by Kermack and McKendrick for the general model [36]. The results by C. K. Yang and F. Brauer in [34, 35, 36] are the mathematical justification of the inconsistent qualitative finding obtained by Z. Feng *et al.* in [22]. In 2009, A. L. Loyd, [37], assessed the structural sensitivity of a SEIR model with Gamma distributed latent and infectious periods on the estimation of the basic reproduction number by applying the general relation established by D. Anderson and R. Watson in 1980. In 2010, E. Vergu *et al.* assessed the distributional effect of the Gamma family of distributions for the infectious period, on the distribution of some epidemiological quantities obtained from realizations of a stochastic metapopulation epidemic model [38].

3.2 Methods

3.2.1 General approach

The influence of parameter uncertainty on the outcome variables (Questions one through three) are addressed using a simple age-of-infection Susceptible-Infectious-Recovered model (see 2.4-2.6) and a global uncertainty and sensitivity analyses using Latin Hypercube Sampling (LHS). Two sensitivity indices, Partial Rank Correlation Coefficient (PRCC) and Spearman's Rank Correlation Coefficient (RCC) were used to quantify the order of significance.

3.2.2 Mathematical definition of the outcome or predicted epidemiological quantities of interest

The mathematical definition of outcome or predicted epidemiological quantities of interest previously mentioned are as follow:

• An important dimensionless quantity or ratio in the epidemiology of infectious diseases is the so called the **basic reproduction number**, commonly denoted by $\mathscr{R}_0$. It represents the *average* number of secondary new infected cases produced by a *typical* infectious individual, over its entire infectious period, introduced in a *completely* susceptible population (see [68]). Mathematically, it is computed as the spectral radius of the *next generation matrix* (see [69]). It quantifies the circumstances under which an epidemic will occur. "Generally" speaking, if the basic reproduction number is strictly less than *one (the epidemic threshold or tipping point)*, then an epidemic will not occur while if it is strictly larger than one, then an epidemic will occur, thus $\mathscr{R}_0$ is critical to the characterization of the qualitative behavior of epidemic models. In addition, it helps identify the degree of intervention required to control an outbreak. For the age-of-infection SIR model in (2.4)-(2.6), the basic reproduction

model is given by

$$\mathscr{R}_0 = \beta\Big(\int_{-\infty}^{\infty} t f(t;k,\theta)dt\Big) = \beta\Big(\int_0^{\infty} \bar{F}(t;k,\theta)dt\Big) = \beta\tau, \tag{3.1}$$

where β is the transmission rate, f is the Gamma probability density function for the infectious period, $\bar{F}$ is the survivor function and τ is the mean of f (see **appendix F** for details on its derivation). Notice that the above expression for $\mathscr{R}_0$ does not depend on the variance of the infectious period distribution nor the variance of the distribution of contacts as it is in the Anderson M. Roy and Robert M. May's non-homogeneous mixing model (1992).

• The **peak size of the prevalence** of infectious individuals is given by

$$I_{peak} = \max_{t\in\Re^+} I(t). \tag{3.2}$$

• **Time at which the peak of the prevalence occurs** is given by

$$t_{peak} = \{t \in \Re^+ : \text{ and } I(t) = I_{peak}\}. \tag{3.3}$$

• **Epidemic duration** is given by

$$t_{final} = \min\{t \in \Re^+ : \text{ and } I(t) < 1\}. \tag{3.4}$$

• Number of cumulative newly infections at the end of the epidemic or **final epidemic size** is given by

$$z = C(t_{final}) = S_0 + \Big(\frac{N}{\mathscr{R}_0}\Big) W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\right). \tag{3.5}$$

In equation (3.5), W is a special function known as the Lambert W function. See **appendix G** for details on a derivation of equation (3.5).

3.2.3 Global uncertainty and sensitivity analyses

The **uncertainty analysis** allows to assess the variability or prediction imprecision of the outcome variables with respect to the uncertainty that comes from the estimates of

the input parameters (Iman & Helton, 1988). The technique is *global* in the sense that the uncertainty analysis is executed simultaneously for all the input model parameters of interest. While the **sensitivity analysis** can be considered as an extension of the global uncertainty analysis in the context that provides a rank of importance for each input parameters with respect to the prediction imprecision of the output variable of interest as well as their qualitative relations. In other words, the global uncertainty and sensitivity analyses combined provide great insights on how the variability or uncertainty in the values of the input parameters affect the values of the outcome variables (Iman & Helton, 1988).

The methodology of global uncertainty and sensitivity analyses was introduced in 1979 by M. D. McKay *et al.* (see [39]), improved by R. L. Iman *et al.* during the 80's decade with a series of eight papers and was not until 1994 that S. M. Blower (see [40]) applied it for the first time to an epidemiological model (an HIV model) described by a nonlinear system of ordinary differential equations. To my knowledge this is the first work using this methodology on an epidemiological model governed by a nonlinear system of Volterra integral equations. Global uncertainty and sensitivity analyses were carried out following these seven steps: **Step 1:** Assign a probability density functions to each of the K input model parameters; **Step 2:** Choose a total number of simulations (N_{sim}); **Step 3:** Divide the range of each of the K input parameters into N_{sim} equi-probable intervals; **Step 4:** Determine the LHS matrix; **Step 5:** Sample the values for each of the K input parameters by using the corresponding indices from the LHS matrix and execute N_{sim} simulations; **Step 6:** Perform a global uncertainty analysis; **Step 7:** Perform a global sensitivity analysis.

Four cases are considered:

Case 1A: Assumes low values for the basic reproduction number and exponentially distributed infectious period ($\mathscr{R}_0 = 1.5$ and $k = 1$ or $Var = \tau^2$).

Case 1B: Assumes low value for the basic reproduction number and Gamma distributed

infectious period ($\mathscr{R}_0 = 1.5$ and $k = 4$ or $Var = \frac{\tau^2}{4}$).

Case 2A: Assumes high value for the basic reproduction number and exponentially distributed infectious period ($\mathscr{R}_0 = 15$ and $k = 1$ or $Var = \tau^2$).

Case 2B: Assumes high value for the basic reproduction number and Gamma distributed infectious period ($\mathscr{R}_0 = 15$ and $k = 4$ or $Var = \frac{\tau^2}{4}$).

Step 1: Assignment of a probability density function to each of the K input model parameters.

The set of input model parameters ($K = 4$) consists of: β the constant transmission rate; τ the mean of the Gamma probability density function for the infectious period; I_0 the initial number of infectious individuals; and Var the variance of the Gamma probability density function for the infectious period. The assignments of a distribution for each of the input model parameters with corresponding entry values for the distributions for all four cases (Cases 1A,1B, 2A and 2B) are illustrated in **Tables 3.1 to 3.4**. The input model parameters I_0 and Var are chosen to be Gamma distributed. While the parameters β and τ are chosen to be truncated Gamma distributed, contained as $\beta > 1$ and $\tau > 1$. These constrains guarantee that an epidemic will always occur ($\mathscr{R}_0 > 1$).

Table 3.1: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period) Assignment of the probability density functions with their corresponding entry values for each input model parameter.

Model parameter	Probability density function	p.d.f.'s parameter values			
		k	θ	μ	σ^2
I_0	Gamma	50	0.1	5	$\frac{\mu_{I_0}}{10} = 0.5$
β	Truncated Gamma	12.23	0.1	$\sqrt{\mathscr{R}_0} = 1.22$	$\frac{\mu_\beta}{10} = 0.12$
τ	Truncated Gamma	12.25	0.1	$\sqrt{\mathscr{R}_0} = 1.22$	$\frac{\mu_\tau}{10} = 0.12$
Var	Gamma	15	0.1	$\mu_\tau^2 = 1.5$	$\frac{\mu_{\text{Var}}}{10} = 0.15$

Table 3.2: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period) Assignment of the probability density functions with their corresponding entry values for each input model parameter.

Model parameter	Probability density function	p.d.f.'s parameter values			
		k	θ	μ	σ^2
I_0	Gamma	50	0.1	5	$\frac{\mu_{I_0}}{10} = 0.5$
β	Truncated Gamma	12.23	0.1	$\sqrt{\mathscr{R}_0} = 1.22$	$\frac{\mu_\beta}{10} = 0.12$
τ	Truncated Gamma	12.25	0.1	$\sqrt{\mathscr{R}_0} = 1.22$	$\frac{\mu_\tau}{10} = 0.12$
Var	Gamma	3.75	0.1	$\frac{\mu_\tau^2}{4} = 0.38$	$\frac{\mu_{\text{Var}}}{10} = 0.04$

Table 3.3: Case 2A (High basic reproduction numbers and exponentially distributed infectious period) Assignment of the probability density functions with their corresponding entry values for each input model parameter.

Model parameter	Probability density function	p.d.f.'s parameter values			
		k	θ	μ	σ^2
I_0	Gamma	50	0.1	5	$\frac{\mu_{I_0}}{10} = 0.5$
β	Truncated Gamma	38.73	0.1	$\sqrt{\mathscr{R}_0} = 3.87$	$\frac{\mu_\beta}{10} = 0.39$
τ	Truncated Gamma	38.73	0.1	$\sqrt{\mathscr{R}_0} = 3.87$	$\frac{\mu_\tau}{10} = 0.39$
Var	Gamma	150	0.1	$\mu_\tau^2 = 15$	$\frac{\mu_{\text{Var}}}{10} = 0.15$

Table 3.4: Case 2B (High basic reproduction numbers and Gamma distributed infectious period) Assignment of the probability density functions with their corresponding entry values for each input model parameter.

Model parameter	Probability density function	p.d.f.'s parameter values			
		k	θ	μ	σ^2
I_0	Gamma	50	0.1	5	$\frac{\mu_{I_0}}{10} = 0.5$
β	Truncated Gamma	38.73	0.1	$\sqrt{\mathscr{R}_0} = 3.87$	$\frac{\mu_\beta}{10} = 0.39$
τ	Truncated Gamma	38.73	0.1	$\sqrt{\mathscr{R}_0} = 3.87$	$\frac{\mu_\tau}{10} = 0.39$
Var	Gamma	37.5	0.1	$\frac{\mu_\tau^2}{4} = 3.8$	$\frac{\mu_{\text{Var}}}{10} = 0.38$

Step 2: Choose a total number of simulations (N_{sim}).

The total number of simulation is: $N_{sim} = 1000$.

Step 3: Divide the range of each of the K input parameters into N_{sim} equi-probable intervals.

Let x be one of the K^{th} random input model parameters, which follows a probability density function f, cumulative distribution function F and inverse cumulative distribution function F^{-1}. Then, the N_{sim} equi-probable intervals $[x^1_{\min}, x^1_{\max}], [x^2_{\min}, x^2_{\max}], \ldots, [x^{N_{sim}}_{\min}, x^{N_{sim}}_{\max}]$ are chosen as follow:

$$\begin{aligned}
x^1_{\min} &= \min_x f(x) \text{ and } x^{N_{sim}}_{\max} = \max_x f(x), \\
x^i_{\max} &= F^{-1}\left[F(x^i_{\min}) + \frac{1}{N_{sim}}\right], \text{ since } \int_{x^i_{\min}}^{x^i_{\max}} f(x)dx = F(x^i_{\max}) - F(x^i_{\min}) = \frac{1}{N} \\
&\quad \text{for } i = 1, \ldots, N_{sim} - 1, \\
x^{i+1}_{\min} &= x^i_{\max} \text{ for } i = 1, \ldots, N_{sim} - 1.
\end{aligned}$$

Step 4: Determine the LHS matrix.

The Latin Hypercube Sampling (LHS) matrix is an N_{sim} by K matrix, in our case a 1000 by 4 matrix, where the elements in each column represent the ordered values (positive whole numbers) or indices for the values of the input model parameters after sampling just once every equi-probable interval, commonly known as sampling without replacement. Permutation on the sample of each input model parameters were executed with the purpose of reducing the correlation, if any, among the K samples. Box-plots, histograms and simple descriptive statistics such as minimum, maximum, mean, median and variance are shown in **Figures 3.1 to 3.4** and **Tables 3.5 to 3.8**.

Table 3.5: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period) Descriptive statistics from the uncertainty analysis.

Input variables	Baseline values	(sample) Statistics				
		Min.	Max.	Mean	Median	Std.
I_0	0.05	0.03	0.07	0.05	0.05	0.01
β	1.22	1.00	2.67	1.38	1.32	0.28
τ	1.22	1.00	2.58	1.38	1.32	0.29
Var	1.5	0.50	3.01	1.50	1.47	0.39

Table 3.6: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period) Descriptive statistics from the uncertainty analysis.

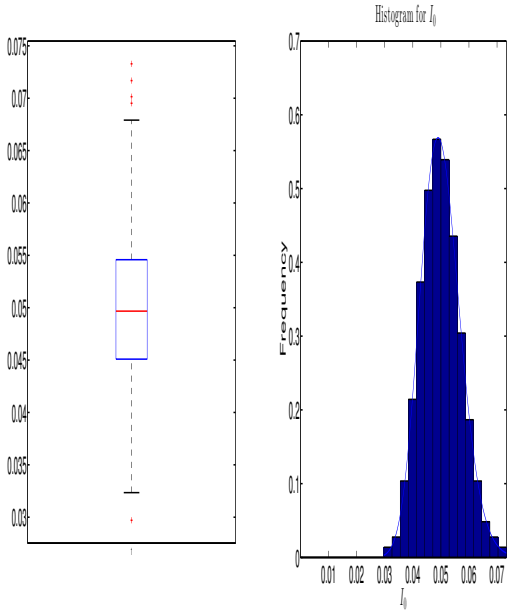
Input variables	Baseline values	(sample) Statistics				
		Min.	Max.	Mean	Median	Std.
I_0	0.05	0.03	0.08	0.05	0.05	0.01
β	1.22	1.00	2.66	1.38	1.32	0.28
τ	1.22	1.00	2.55	1.38	1.32	0.28
Var	0.38	0.03	1.36	0.38	0.34	0.19

Table 3.7: Case 2A (High basic reproduction numbers and exponentially distributed infectious period) Descriptive statistics from the uncertainty analysis.

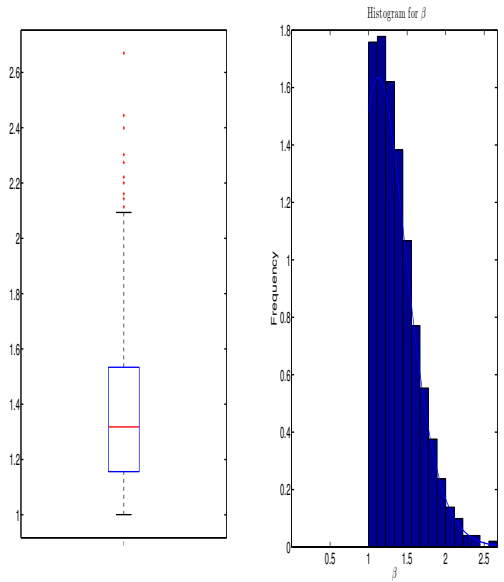
Input variables	Baseline values	(sample) Statistics				
		Min.	Max.	Mean	Median	Std.
I_0	0.05	0.03	0.08	0.05	0.05	0.01
β	3.87	2.25	6.10	3.87	3.84	0.06
τ	3.87	2.18	6.46	3.87	3.84	0.63
Var	15	11.66	18.60	15.0	14.96	1.22

Table 3.8: Case 2B (High basic reproduction numbers and Gamma distributed infectious period) Descriptive statistics from the uncertainty analysis.

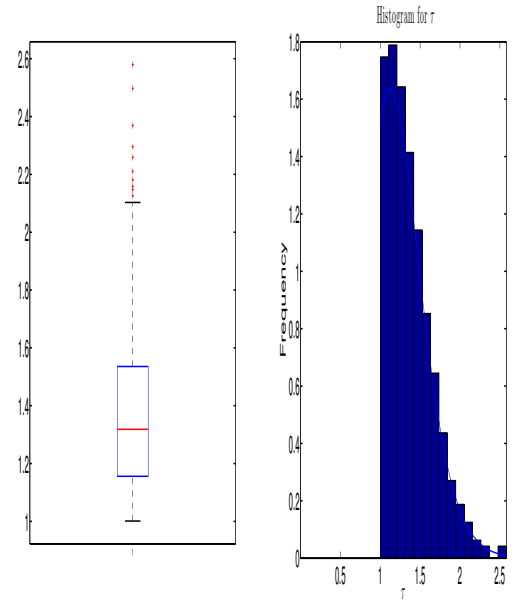
Input variables	Baseline values	(sample) Statistics				
		Min.	Max.	Mean	Median	Std.
I_0	0.05	0.03	0.07	0.05	0.05	0.01
β	3.87	2.25	6.23	3.87	3.84	0.62
τ	3.87	2.27	6.41	3.87	3.84	0.63
Var	3.75	2.08	6.04	3.75	3.72	0.61



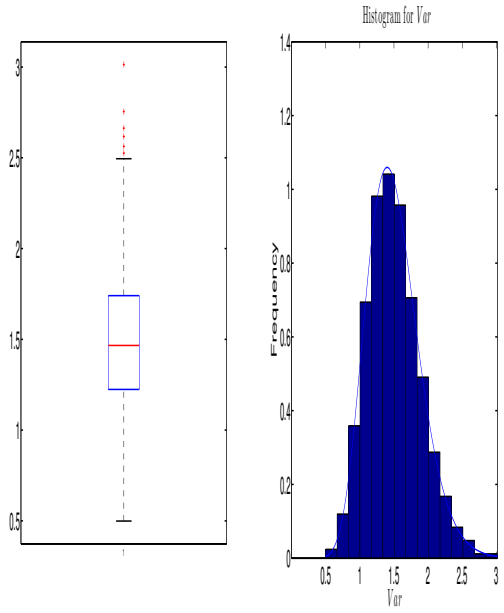
(a) Empirical distribution of I_0.



(b) Empirical distribution of β.

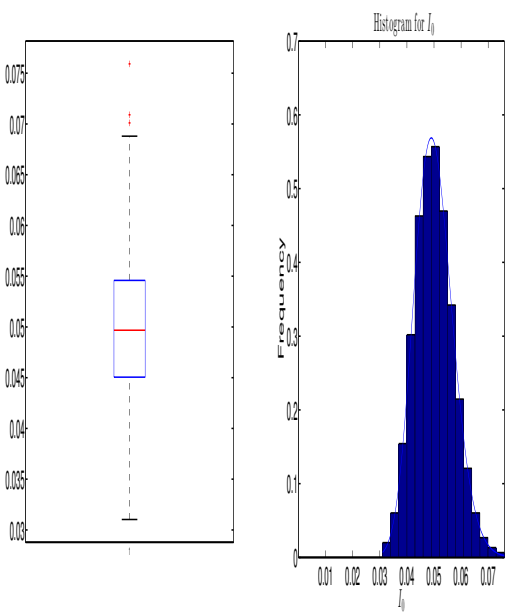


(c) Empirical distribution of τ.

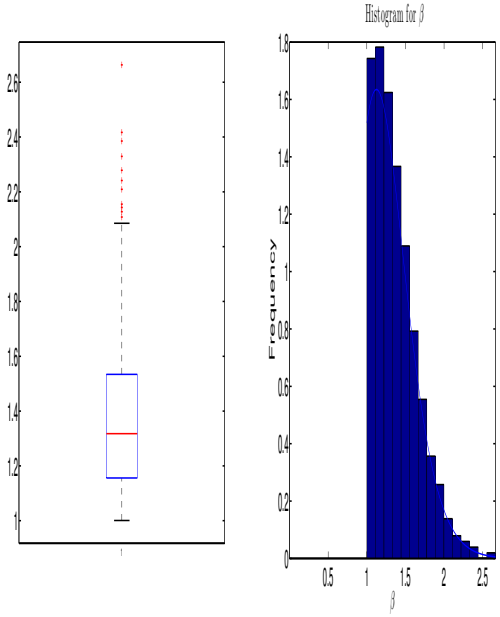


(d) Empirical distribution of Var

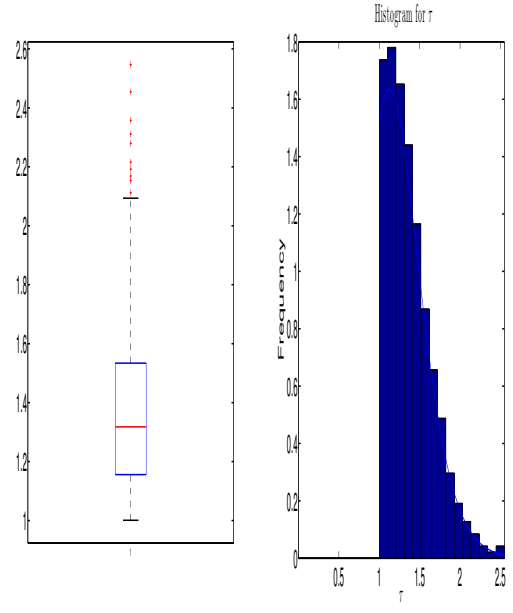
Figure 3.1: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Box-plots and histograms from the samples for each of the input model parameter.



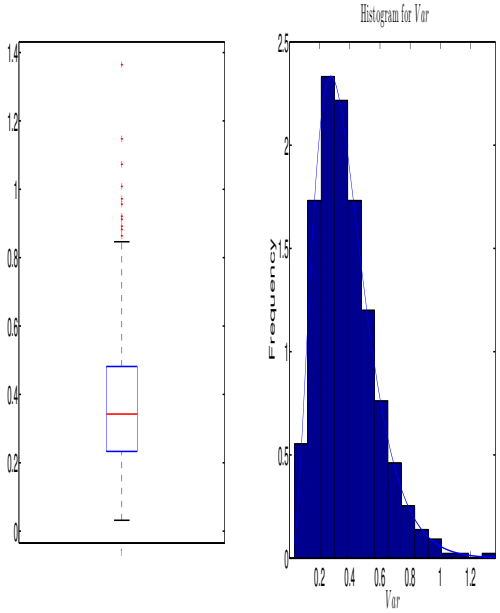
(a) Empirical distribution of I_0.



(b) Empirical distribution of β.

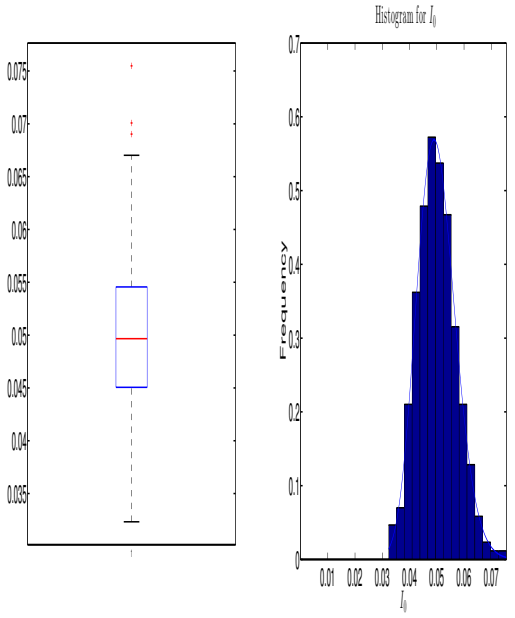


(c) Empirical distribution of τ.

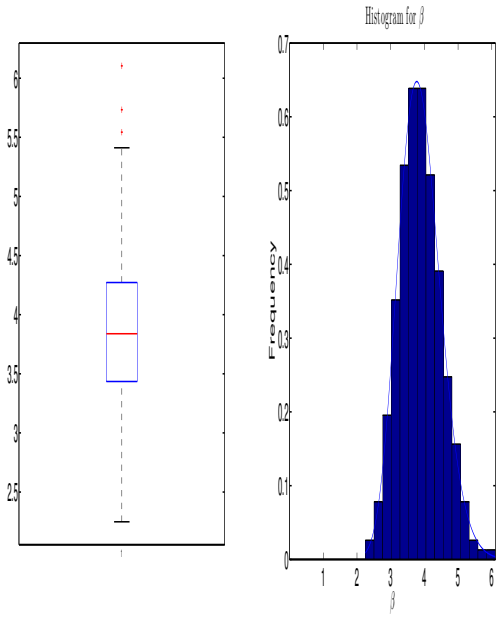


(d) Empirical distribution of Var

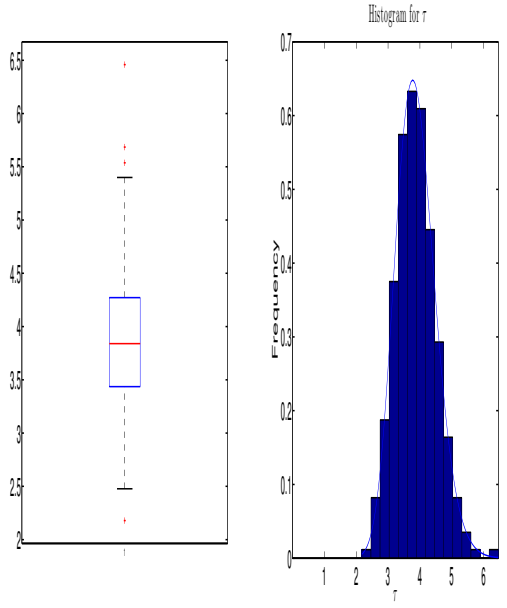
Figure 3.2: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Box-plots and histograms from the samples for each of the input model parameter.



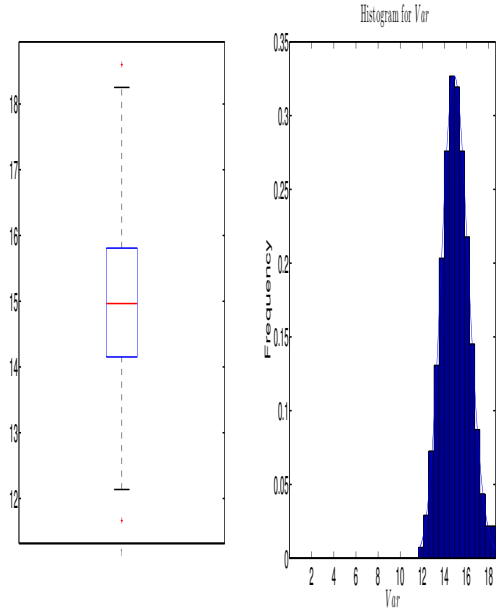
(a) Empirical distribution of I_0.



(b) Empirical distribution of β.

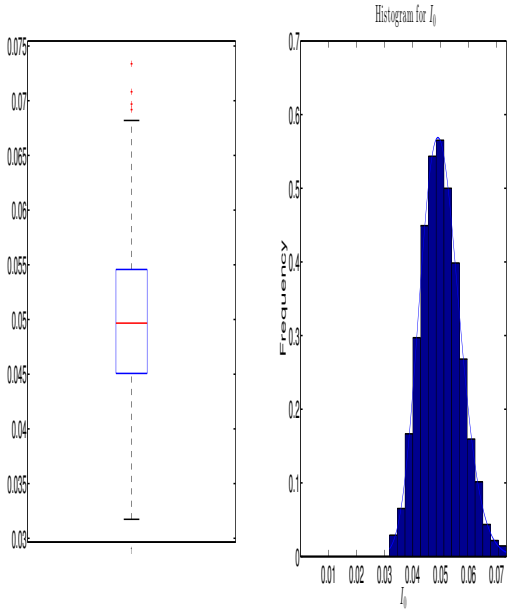


(c) Empirical distribution of τ.

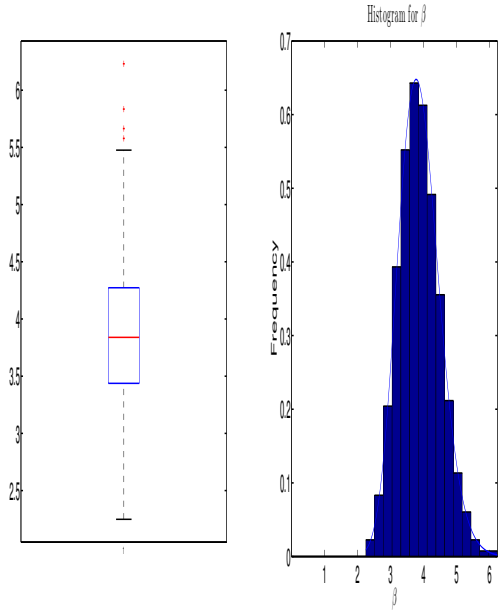


(d) Empirical distribution of Var

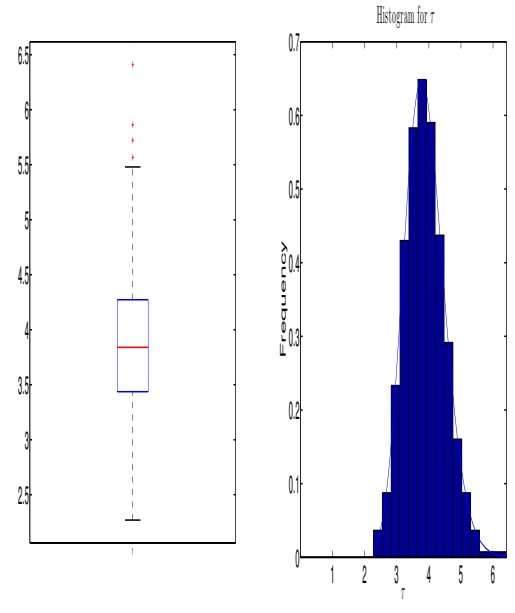
Figure 3.3: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Box-plots and histograms from the samples for each of the input model parameter.



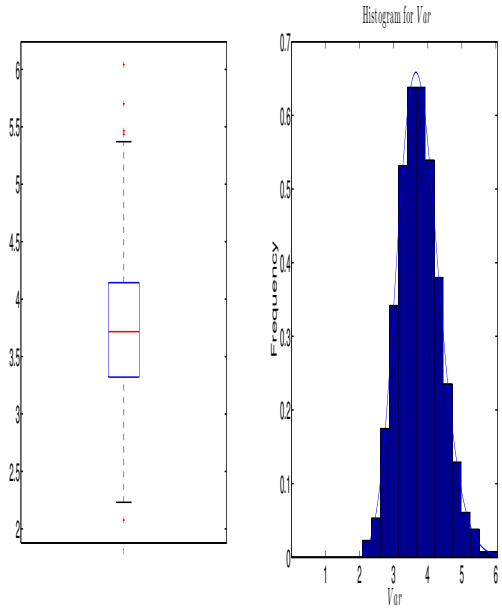
(a) Empirical distribution of I_0.



(b) Empirical distribution of β.



(c) Empirical distribution of τ.



(d) Empirical distribution of Var

Figure 3.4: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Box-plots and histograms from the samples for each of the input model parameter.

Step 5: Use the sampled values for each of the K input parameters and perform N_{sim} simulations.

The K sample sets for the input model parameters are used to performed N_{sim} simulations where the outcome or prediction variables are storage in column vectors.

Step 6: Perform a global uncertainty analysis.

Characterization of each sample obtained for the outcome variables is done via boxplots, histograms and simple descriptive statistics such as minimum, maximum, mean, median and variance. These are shown in **Figures 3.5 to 3.8** and **Tables 3.9 to 3.12**.

Step 7: Perform a sensitivity analysis.

We also computed two sensitivity indices: Partial rank correlation coefficient (PRCC) and Spearman rank correlation coefficient (RCC). The magnitude of a sensitivity index measures how strong the qualitative relation is between an input model parameter and an outcome variable. That is, it quantifies the statistical influence of the estimate of an input model parameter to the prediction precision of an outcome variable. The sign of the sensitivity index indicates how is the qualitative relation between an input model parameter and an outcome variable; a positive value indicates that the value of the output variable increases as the value of the input variable increases, otherwise, the value of the outcome variable decreases as the value of the input variable increases. The associated probability value (P-value) of the sensitivity index determines the statistical significance of the qualitative relation between an input model parameter and the outcome variable. Since the samples for the input model parameters are not normally distributed, a two tailed non-parametric statistical hypothesis test with null hypothesis under the assumption that the sensitivity index equals zero is used. The PRCC sensitivity index assumes that the qualitative relation between an input model parameter and an outcome variable is monotone. The monotonicity assumption is validated via the scatters plots.

3.2.4 Numerical implementation of the Global uncertainty and sensitivity analyses

The seven steps needed to perform the Global uncertainty and sensitivity analyses were coded in MATLAB 8.1.0 (R2013a, The MathWorks). The truncated distributions were obtained with the MATLAB built-in function *truncate* and the corresponding cumulative distribution and inverse cumulative distribution functions with *cdf* and *icdf*. The random number generators with Gamma and truncated Gamma distributions were obtained by *random*. Random permutation for the vector of input model parameter values were done with *randperm*. The descriptive statistics were obtained with: *min, max, mean, median, var*. The numerical recipes to calculate the PRCC and its corresponding non parametric statistical hypothesis test are explained in detail in the appendix A of [40]. These were written in MATLAB with the following functions *sort, tril, inv, tinv, tcdf*. The numerical recipe to calculate the RCC and its corresponding non parametric statistical hypothesis test were taken from [72] (*spear*) and re-written in MATLAB.

3.3 Results

3.3.1 Interpretation of results from the global uncertainty analysis

Statements from the uncertainty analysis are based on simple descriptive statistics (see **Tables 3.9 to 3.12**), box-plots and empirical distributions (see **Figures 3.5 to 3.8**). The global uncertainty analysis illustrates that the overall prediction precision of the age-of-infection SIR model with respect to the basic reproduction number, in all four cases, is moderate. In the cases when the basic reproduction number is low (Cases 1A and 1B), the overall prediction precision of the age-of-infection SIR model with respect to outcome variables such as the peak size and peak timing of the prevalence of infectious individuals, final epidemic size and epidemic duration is low. While in the cases when the basic reproduction number is high (Cases 2A and 2B), the overall prediction precision of the

age-of-infection SIR model with respect to the peak size and peak timing of the prevalence of infectious individuals and final epidemic size is moderate. However, for the epidemic duration, the overall prediction precision of the age-of-infection SIR model is low under the exponential assumption (Case 2A) and moderate under the gamma assumption (Case 2B).

Table 3.9: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Descriptive statistics from the uncertainty analysis.

Outcome variables	Baseline values	(sample) Descriptive Statistics						
		Min.	Max.	Mean	Median	Std.	CV= $\frac{Std.}{Mean}$	VMR= $\frac{Var.}{Mean}$
Prevalence peak size (%)	6.34	0.15	68.22	15.20	12.84	11.52	0.76	8.73
Prevalence peak time	15.70	15.70	65.20	13.67	11.10	8.19	0.60	4.91
Final epidemic size (%)	58.32	11.55	99.43	69.39	74.21	18.84	0.27	5.12
Epidemic duration	40.70	13.80	161.10	36.20	30.55	19.37	0.54	10.36
Basic reproduction number	1.50	1.06	5.20	1.89	1.83	0.54	0.29	0.16

Table 3.10: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Descriptive statistics from the uncertainty analysis.

Outcome variables	Baseline values	(sample) Descriptive Statistics						
		Min.	Max.	Mean	Median	Std.	CV= $\frac{Std.}{Mean}$	VMR= $\frac{Var.}{Mean}$
Prevalence peak size (%)	10.08	0.16	76.36	21.53	18.52	15.06	0.70	10.53
Prevalence peak time	10.70	3.70	35	9.54	8.60	3.80	0.40	1.52
Final epidemic size (%)	58.33	8.68	99.18	69.26	72.28	18.45	0.27	4.92
Epidemic duration	26.10	9.30	101.30	23.41	20.90	9.87	0.42	4.16
Basic reproduction number	1.50	1.04	4.85	1.89	1.77	0.56	0.30	0.17

Table 3.11: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Descriptive statistics from the uncertainty analysis.

Outcome variables	Baseline values	(sample) Descriptive Statistics						
		Min.	Max.	Mean	Median	Std.	CV= $\frac{Std.}{Mean}$	VMR= $\frac{Var.}{Mean}$
Prevalence peak size (%)	75.27	33.56	95.42	73.34	74.68	10.62	0.14	1.54
Prevalence peak time	2.90	1.90	6.80	3.03	3	0.54	0.18	0.10
Final epidemic size (%)	100.00	99.67	100	99.998	100	0.02	1.9×10^{-4}	3.6×10^{-6}
Epidemic duration	37.90	31.10	51.40	38.41	38.20	3.19	0.08	0.26
Basic reproduction number	15.00	5.74	27.02	15.02	14.75	3.50	0.23	0.81

Table 3.12: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Descriptive statistics from the uncertainty analysis.

Outcome variables	Baseline values	(sample) Descriptive Statistics						
		Min.	Max.	Mean	Median	Std.	CV= $\frac{Std.}{Mean}$	VMR= $\frac{Var.}{Mean}$
Prevalence peak size (%)	94.46	69.14	99.66	92.84	94.22	5.08	0.05	0.28
Prevalence peak time	3.00	2	5	3.06	3	0.45	0.15	0.07
Final epidemic size (%)	100.00	99.93	100	99.998	100	0.004	4.6×10^{-5}	2.2×10^{-7}
Epidemic duration	17.50	13.90	22	17.70	17.60	1.38	0.08	0.11
Basic reproduction number	15.00	7.28	26.81	15.00	14.67	3.45	0.23	0.79

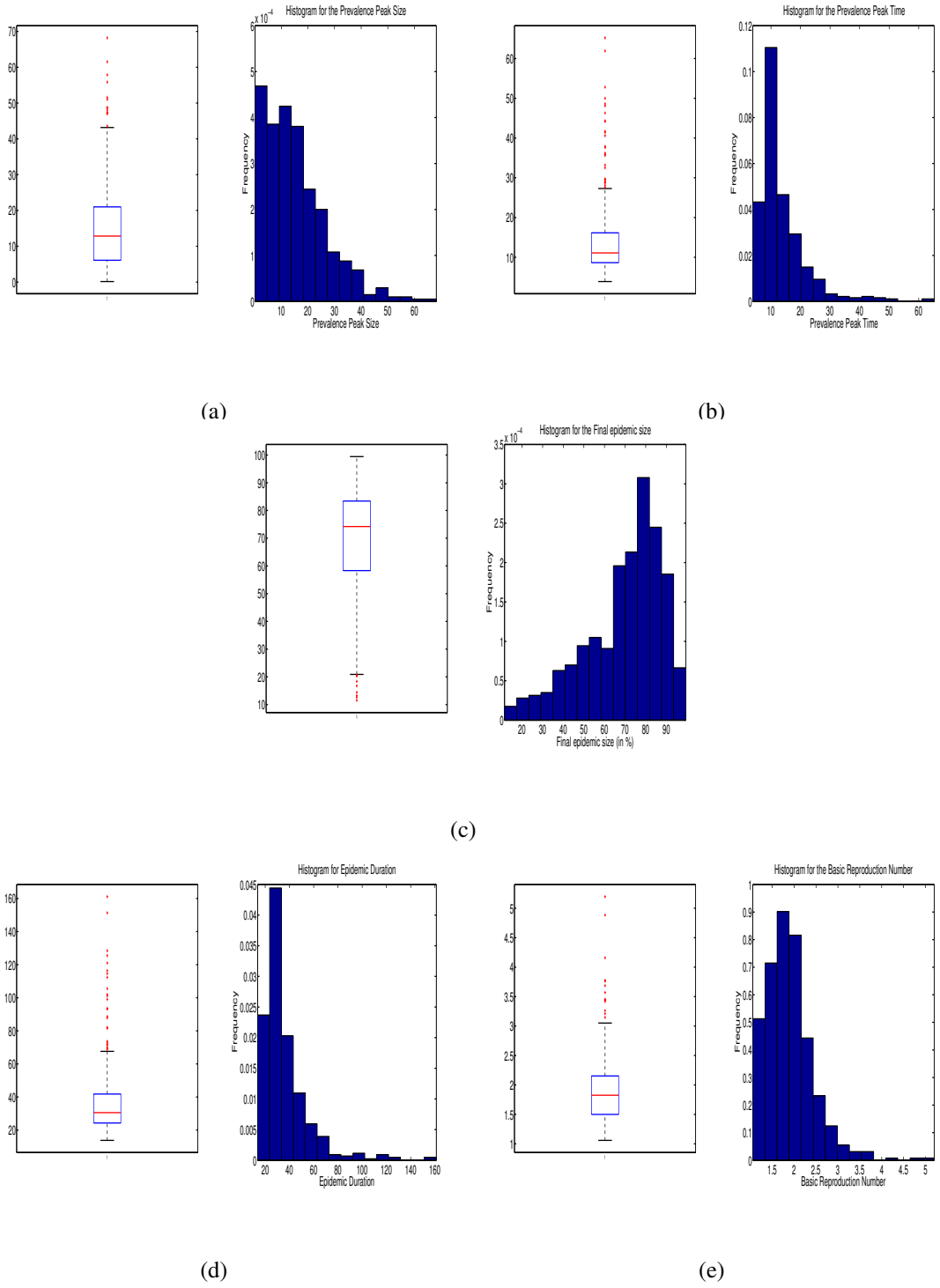
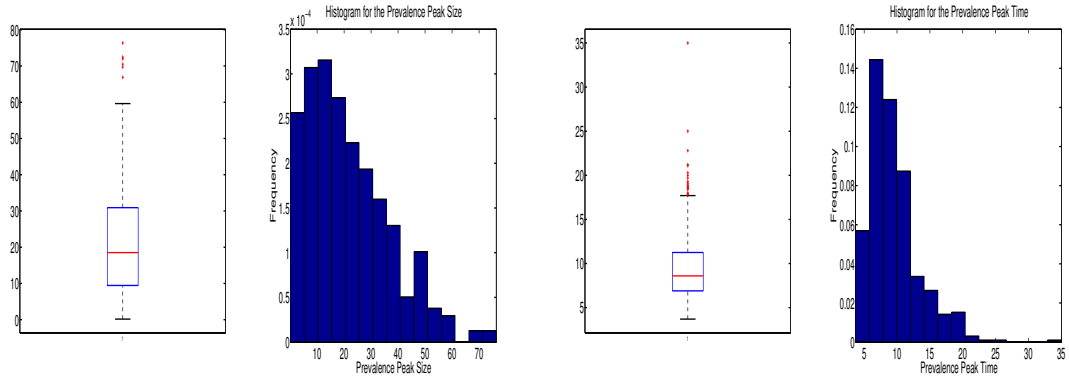
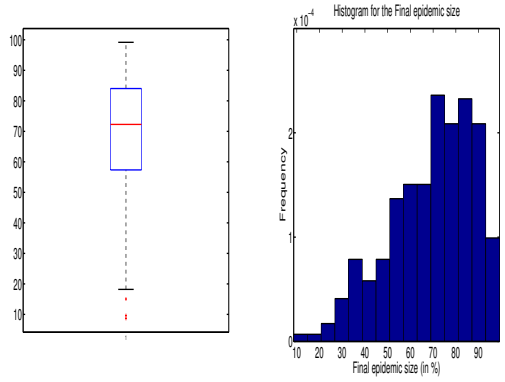


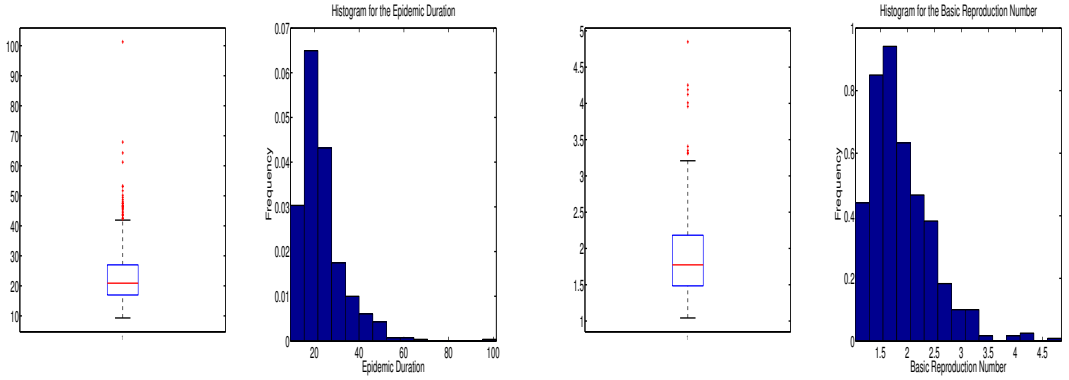
Figure 3.5: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Box-plots and empirical distributions for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number



(a) (b)

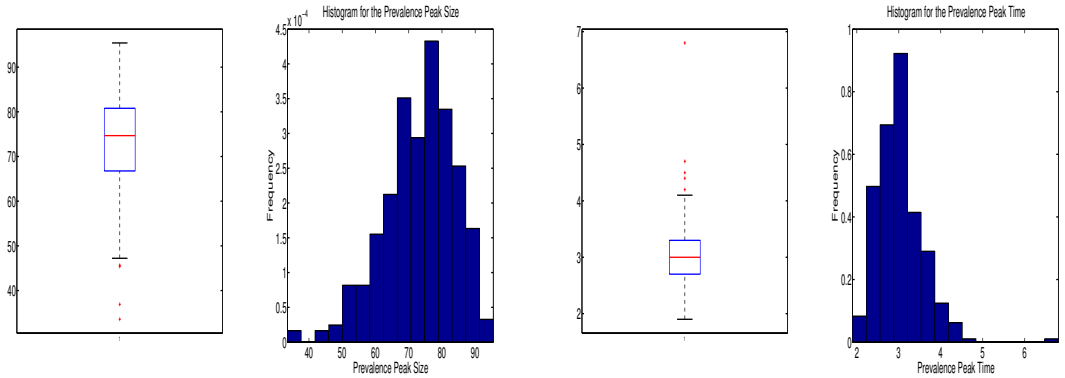


(c)

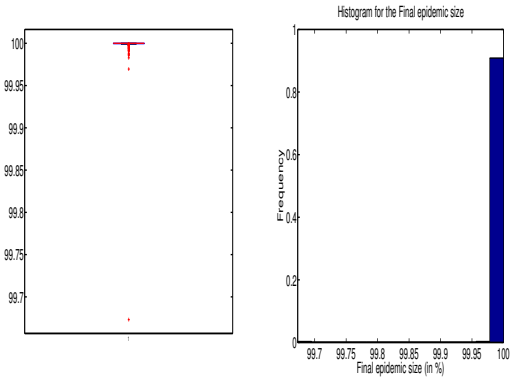


(d) (e)

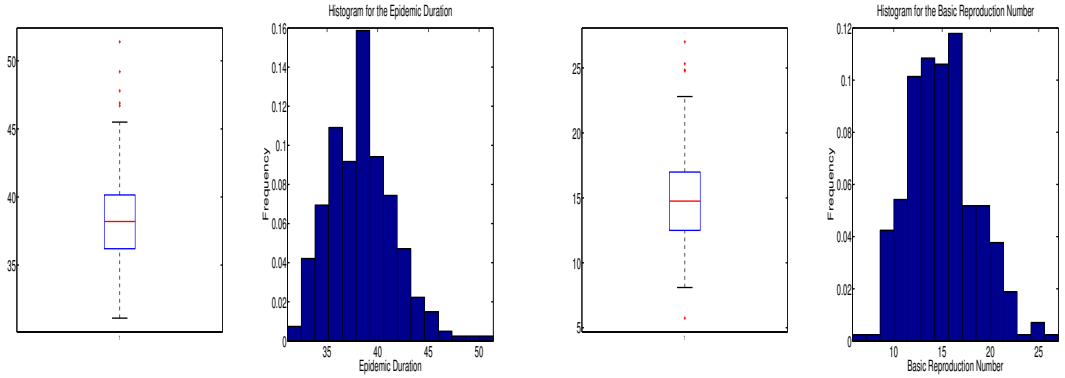
Figure 3.6: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Box-plots and empirical distributions for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number



(a) (b)



(c)



(d) (e)

Figure 3.7: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Box-plots and empirical distributions for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number

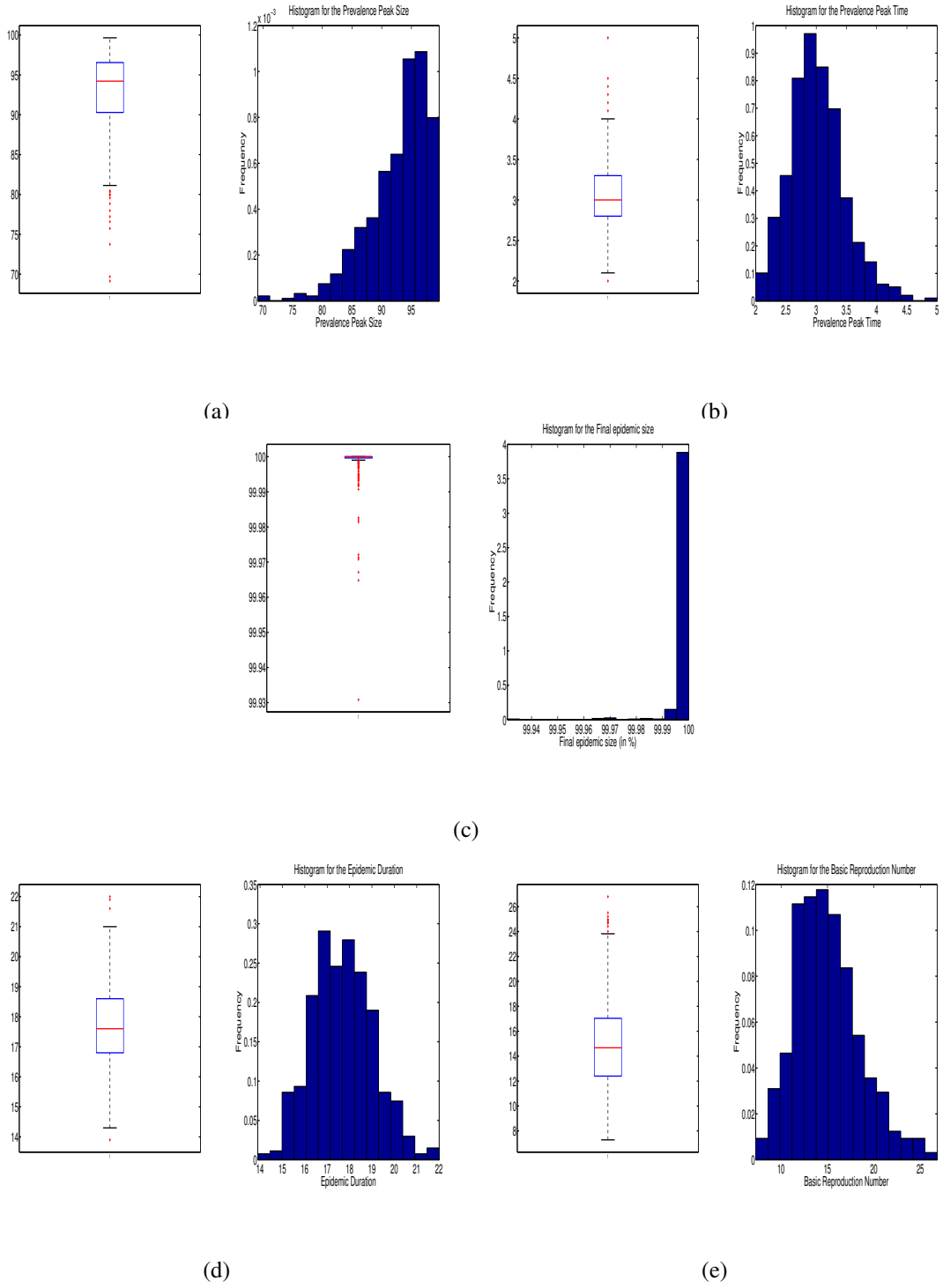


Figure 3.8: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Box-plots and empirical distributions for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number

3.3.2 Interpretation of results from the sensitivity analysis

Interpretation of the scatterplots

The univariate scatterplots reflect that the qualitative relation between each of the outcome variables such as the peak size and peak timing of the prevalence of infectious individuals, the final epidemic size, the epidemic duration and the basic reproduction number and each of the input model parameters such as the transmission rate (β) and the mean (τ) infectious period is monotone, with the exception of all the outcome variables and the initial number of infectious individuals (I_0) and the basic reproduction number and final epidemic size and the variance (*Var*) of the Gamma distribution for the infectious period (see **Figures 3.9 a)-e) to 3.12 a)-e)**). The latter results are expected since the analytical expressions for the basic reproduction number and final epidemic size do not depend on the variance (*Var*) of the Gamma distribution for the infectious period. Therefore, the monotonicity assumption for the PRCC sensitivity index is validated, except for the initial number of infectious individuals (I_0) and in the expected cases mentioned for the variance (*Var*) of the Gamma distribution for the infectious period.

The following results are all under the appropriate level of statistical significance ($\alpha = 0.05$).

Interpretation of the magnitude of the sensitivity indices

• **Peak size of the prevalence of infectious individuals.**

In all cases 1A, 1B, 2A and 2B, **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16** indicate that the input model parameters that statistically (with P-value < 0.05) influence the most to the prediction precision of the peak size of the prevalence of infectious individuals are the mean of the infectious period distribution τ, followed by the transmission rate β and then the variance of the infectious period distribution *Var* in this order.

• **Time at which the peak of the prevalence occurs.**

For the time at which the peak of the prevalence occurs, the input model parameters that statistically (with P-value < 0.05) influence the most to its prediction precision are the transmission rate β, in the first place, then the mean of the infectious period distribution τ followed by: the variance of the infectious period distribution *Var* and the initial number of infectious individuals I_0 for low values of the basic reproduction number (cases 1A and 1B, see **Tables 3.13 and 3.14** and **Figures 3.13 and 3.14**), or the initial number of infectious individuals I_0 and the variance of the infectious period distribution *Var* for high values of the basic reproduction number (cases 2A and 2B, see **Tables 3.15 and 3.16** and **Figures 3.15 and 3.16**).

• **Epidemic duration.**

For low basic reproduction numbers (cases 1A and 1B), the input model parameters that statistically (with P-value < 0.05) influence the most on the prediction precision of the epidemic duration are the transmission rate β followed by the mean of the infectious period distribution τ and then the variance of the infectious period distribution *Var*, in this order (see **Tables 3.13 and 3.14** and **Figures 3.13 and 3.14**). While for high basic reproduction numbers and under the exponential assumption (case 2A) the order changes with the mean of the infectious period distribution τ, ranked first, followed by the variance of the infectious period distribution *Var* and then the transmission rate β (see **Table 3.15** and and **Figure 3.15**). For high basic reproduction numbers and under the Gamma assumption (case 2B) the order is the variance of the infectious period distribution *Var*, ranked first, followed by the transmission rate β, the mean of the infectious period distribution τ and the initial number of infectious individuals I_0 (see **Table 3.16** and **Figure 3.16**).

• **Final epidemic size.**

Under the exponential assumption (cases 1A and 2A), the input model parameters that statistically (with P-value < 0.05) influence the most on the prediction precision of

the final epidemic size are the transmission rate β and then the mean of the infectious period distribution τ (see **Tables 3.13 and 3.15** and **Figures 3.13 and 3.15**), while under the Gamma assumption (case 1B and 2B) the order is reversed (see **Tables 3.14 and 3.16** and **Figures 3.14 and 3.16**), but with the initial number of infectious individuals I_0 in third place for only high basic reproduction numbers (case 2B, see **Table 3.16**).

• **Basic reproduction number.**

The prediction precision of the basic reproduction number is statistically (with P-value < 0.05) influenced the most by the input model parameters: the transmission rate β following by the mean of the infectious period distribution τ (cases 1A, 2A and 2B, see **Tables 3.13, 3.15 and 3.16** and and **Figures 3.13, 3.15 and 3.16**), with the only exemption that for low basic reproduction numbers and under the Gamma distribution assumption (case 1B) the order is reversed (see **Tables 3.14** and **Figure 3.14**).

The variance of the infectious period distribution *Var* was not an input model parameter with statistical influence in the prediction precision of the last two outcome variables mentioned: the final epidemic size and the basic reproduction number. For the particular model considered here (see 2.4-2.6), this last result is expected by just observing the analytical expressions for final epidemic size and reproduction number do not depend on the *Var*.

Although the magnitudes from both sensitivity indices (PRCC and RCC) were different, the orders (or ranks) of statistical influence for the input model parameters on the prediction precision of outcome variables were the same.

Interpretation of the sign of the sensitivity indices

• **Sensitivity index with positive sign.** For the following qualitative relationships, the value of the outcome variable increases as the value of the input model parameter increases:

i) In all cases 1A, 1B, 2A and 2B, the value of outcome variables such as the peak size of the prevalence of infectious individuals, the final epidemic size and the basic re-

production number increases as the value of the input model parameters such as the transmission rate β and the mean of the infectious period distribution τ increases (see **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16**).

ii) In all cases 1A, 1B, 2A and 2B, the value of the epidemic duration increases as the value of the variance of the infectious period distribution *Var* increases (see **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16**).

iii) In cases 1A, 1B and 2A, the value of the time at which the peak of the prevalence of infectious individuals occurs increases as the value of the variance of the infectious period distribution *Var* increases (see **Tables 3.13 to 3.15** and **Figures 3.13 to 3.15**), with the exception of case 2B (high basic reproduction numbers and Gamma distributed infectious period, see **Table 3.16** and **Figure 3.16**).

iv) Only for high basic reproduction numbers and Gamma distributed infectious period (case 2B), the value of the time at which the peak of the prevalence of infectious individuals occurs increases as the value of the mean of the infectious period distribution τ increases (see **Table 3.16** and and **Figure3.16**).

v) Only for high basic reproduction numbers and Gamma distributed infectious period (case 2B), the value of the final epidemic size increases as the value of the initial number of infectious individuals I_0 increases (see **Table 3.16** and **Figure 3.16**).

• **Sensitivity index with negative sign.** For the following qualitative relationships the value of the outcome variable decreases as the value of the input model parameter increases:

i) In all cases 1A, 1B, 2A and 2B, the value of the epidemic duration decreases as the value of the input model parameters such as the transmission rate β and the mean of the infectious period distribution τ increases (see **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16**).

ii) In all cases 1A, 1B, 2A and 2B, the value of the peak size of the prevalence of infectious individuals decreases as the value of the variance of the infectious period distribution *Var* increases (see **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16**).

iii) In all cases 1A, 1B, 2A and 2B, the value of the time at which the peak of the prevalence of infectious individuals occurs decreases as the value of the input model parameters such as the transmission rate β and the initial number of infectious individuals I_0 increases (see **Tables 3.13 to 3.16** and **Figures 3.13 to 3.16**).

iv) In cases 1A, 1B and 2A, the value of the time at which the peak of the prevalence of infectious individuals occurs decreases as the value of the mean of the infectious period distribution τ increases (see **Tables 3.13 to 3.15** and **Figures 3.13 to 3.15**), with the exception of case 2B (high basic reproduction numbers and Gamma distributed infectious period, see **Table 3.16** and **Figure3.16**).

v) Only for high basic reproduction numbers and Gamma distributed infectious period (case 2B), the value of the time at which the peak of the prevalence of infectious individuals occurs decreases as the value of the variance of the infectious period distribution *Var* increases (see **Table 3.16**).

vi) Only for high basic reproduction numbers and Gamma distributed infectious period (case 2B), the value of the epidemic duration decreases as the value of the initial number of infectious individuals I_0 increases (see **Table 3.16** and **Figure 3.16**).

The signs obtained from both sensitivity indices PRCC and RCC were identical.

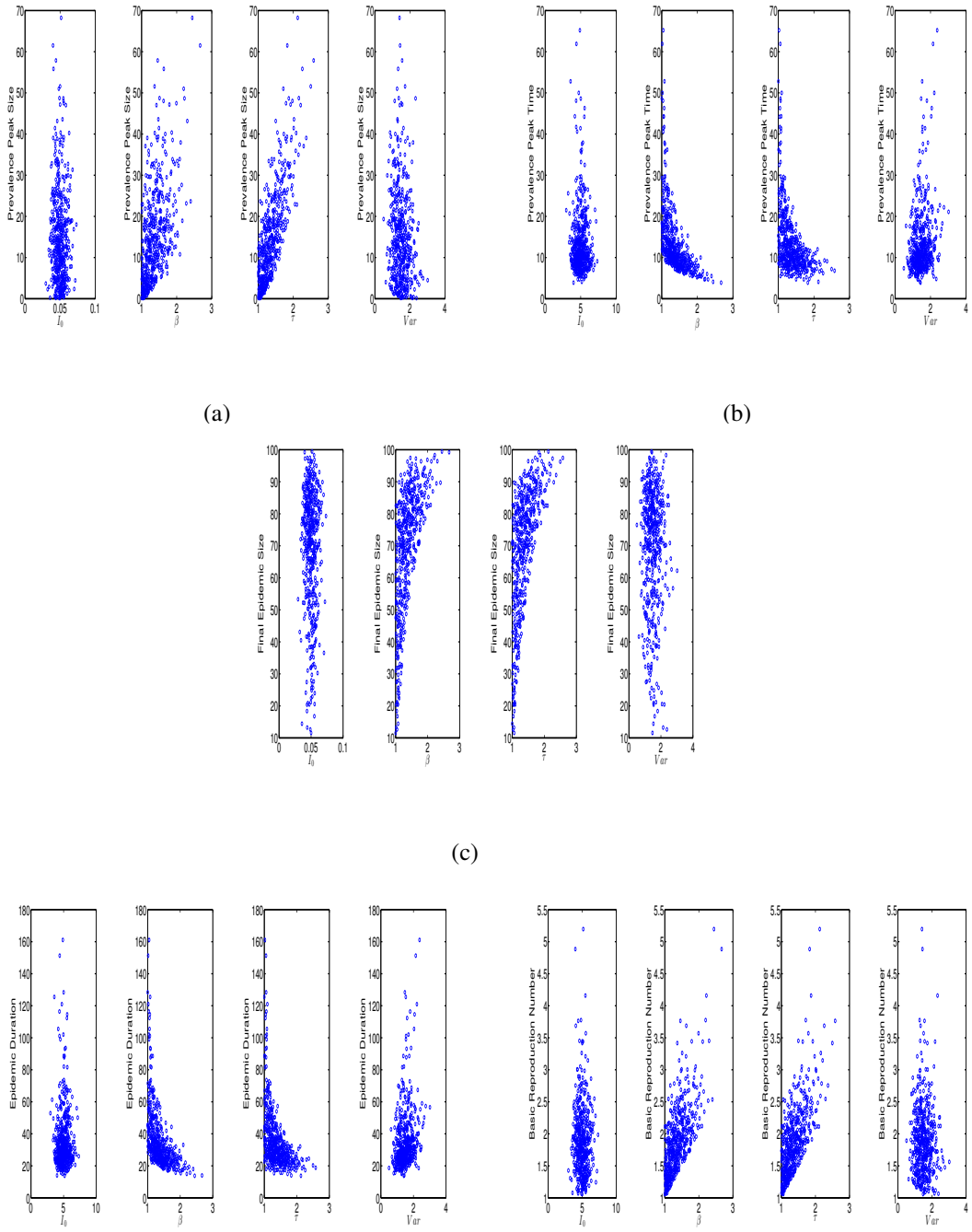


Figure 3.9: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Scatter plots for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number with respect to β, τ, I_0 and *Var*.

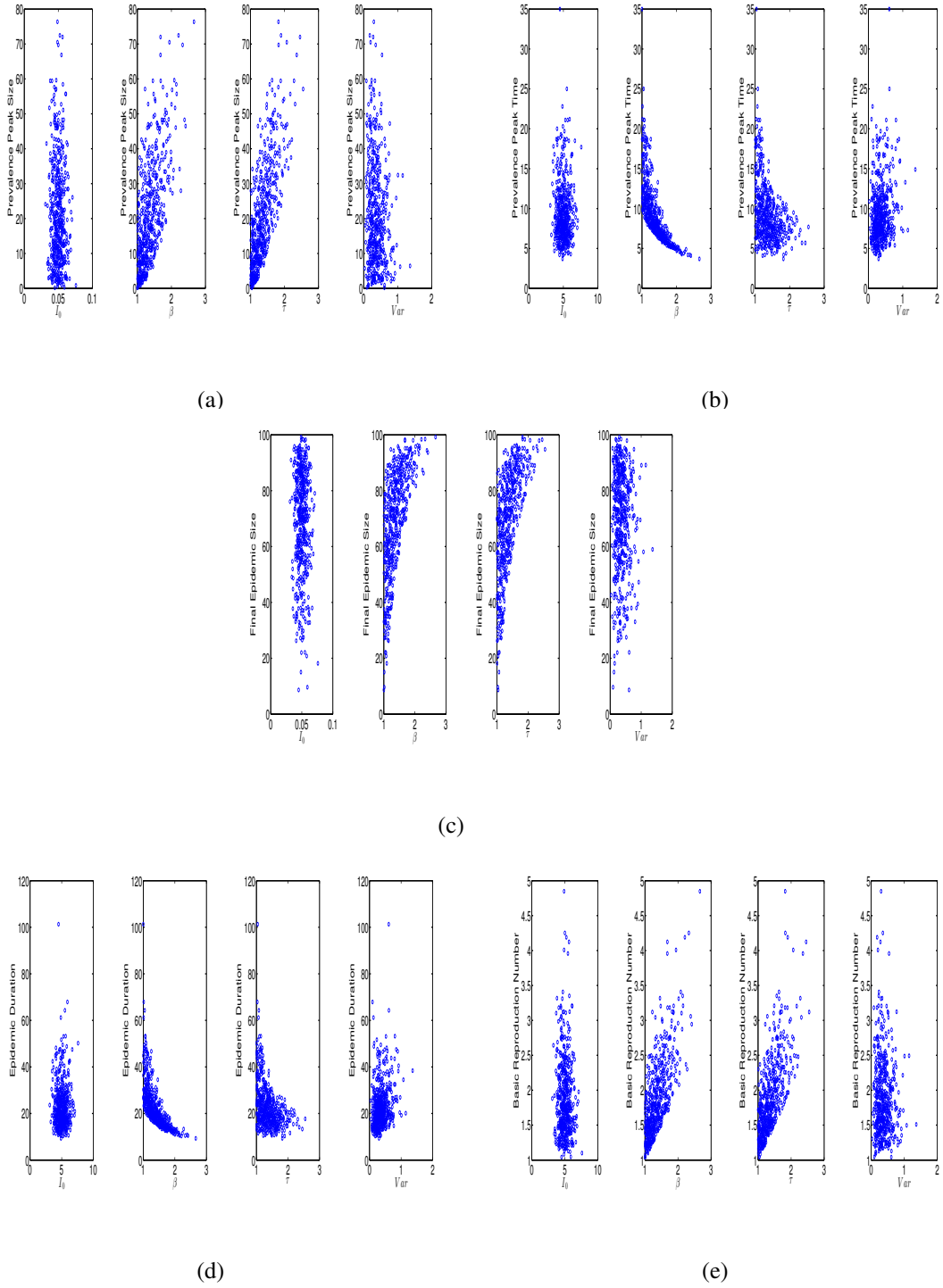


Figure 3.10: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Scatter plots for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number with respect to β, τ, I_0 and Var.

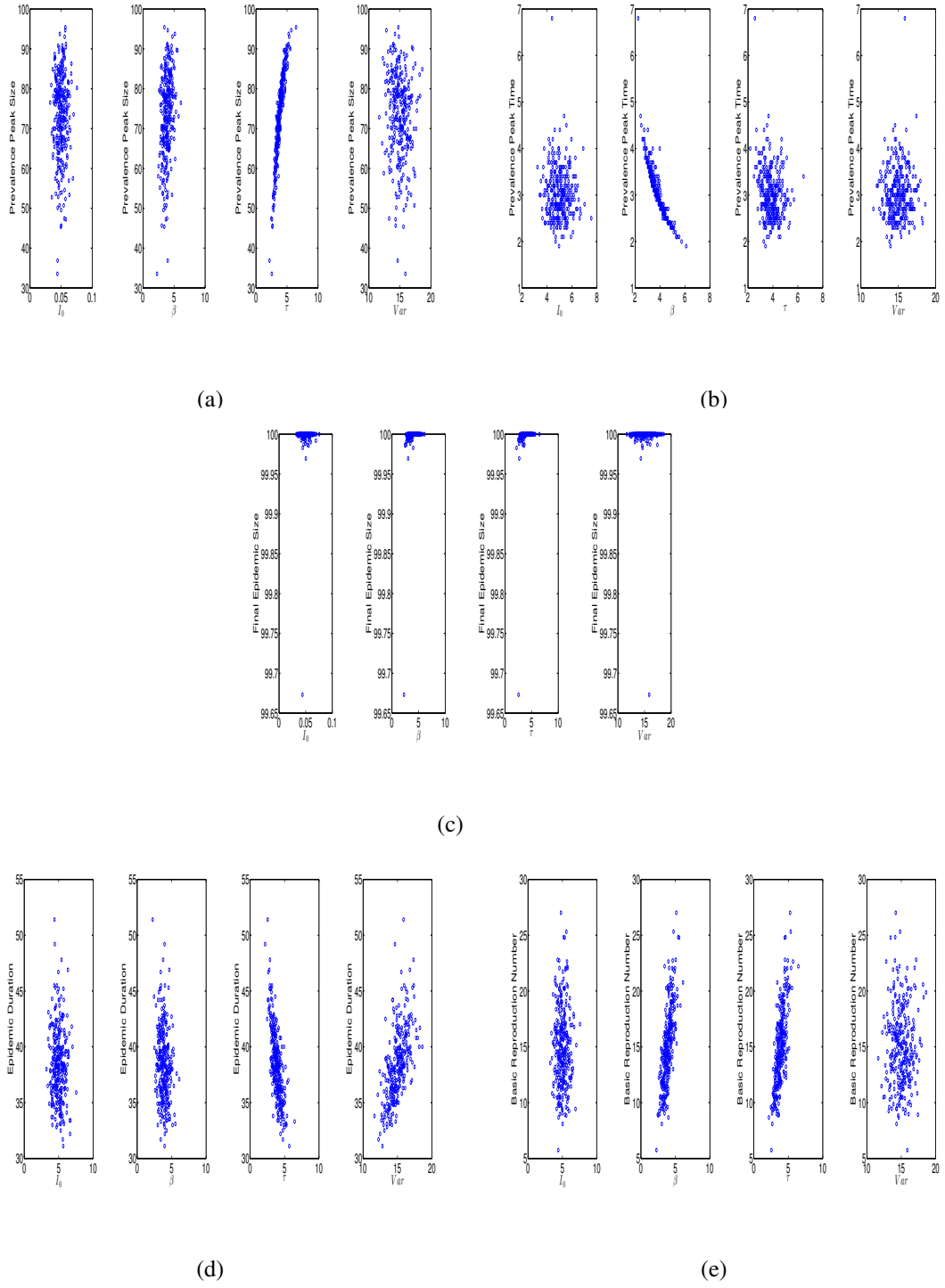
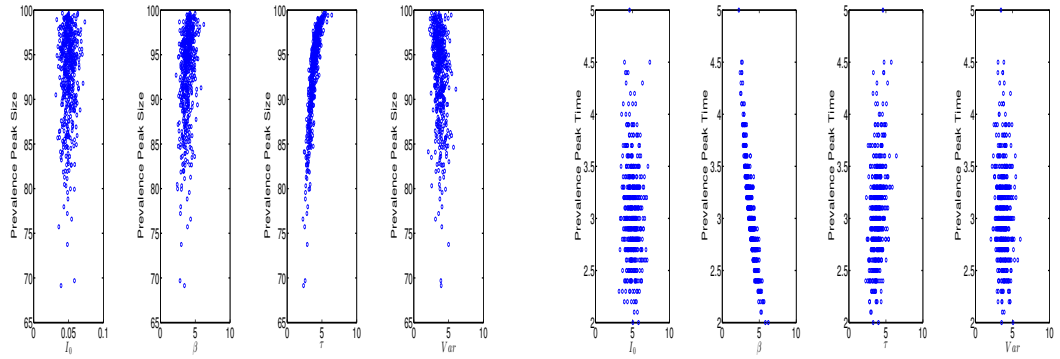
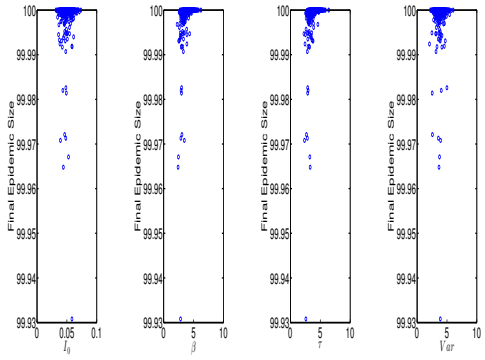


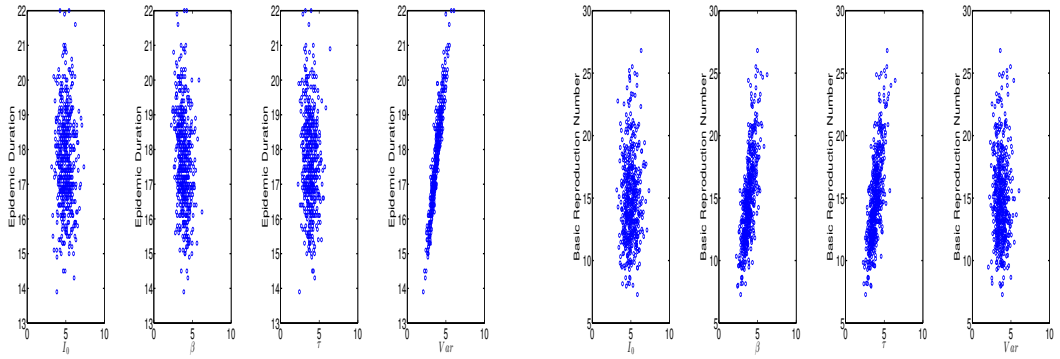
Figure 3.11: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Scatter plots for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number with respect to β, τ, I_0 and Var.



(a) (b)



(c)



(d) (e)

Figure 3.12: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Scatter plots for **a)** the peak size of the prevalence of infectious individuals, **b)** the time at which the peak of the prevalence occurs, **c)** the final epidemic size, **d)** the epidemic duration and **e)** the basic reproduction number with respect to β, τ, I_0 and Var.

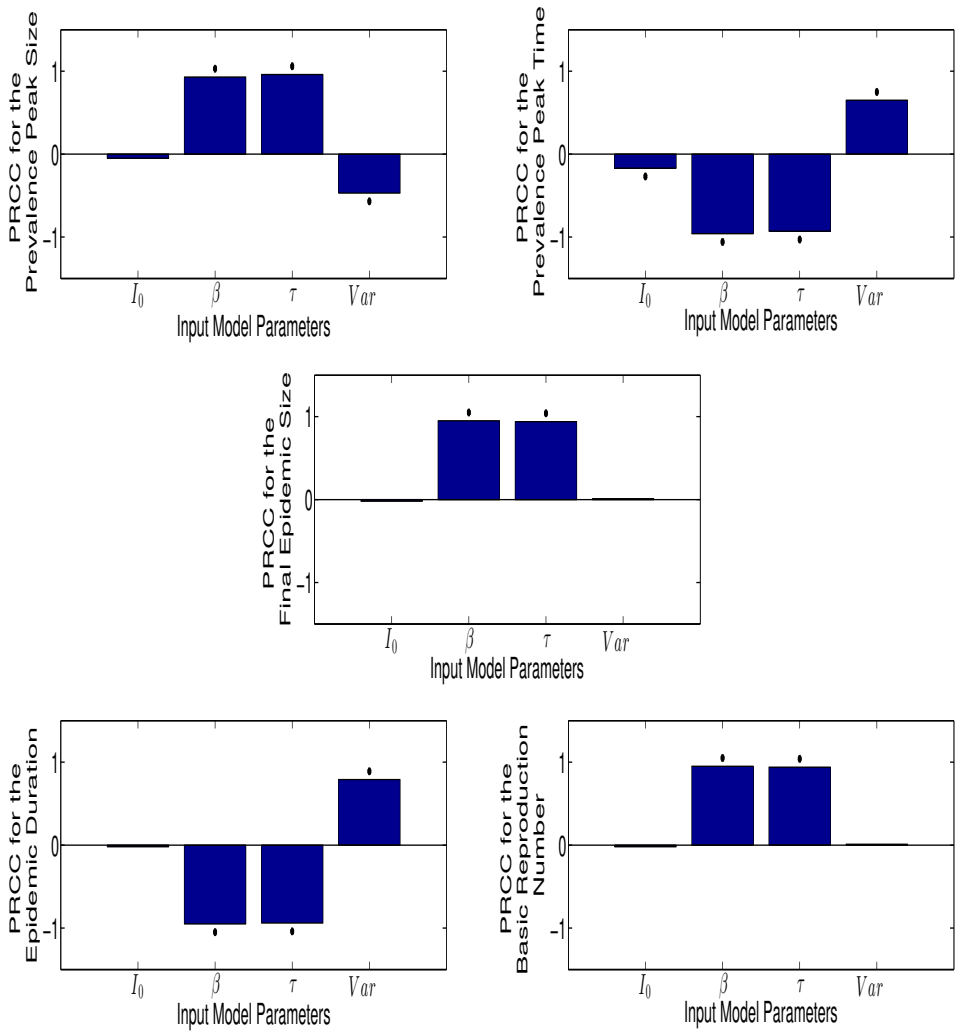


Figure 3.13: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Partial rank correlation coefficients (PRCC)

Table 3.13: Case 1A (Low basic reproduction numbers and exponentially distributed infectious period): Partial rank correlation coefficients (PRCC) and Spearman rank correlation coefficient (RCC) with associated probability values for each outcome variable

Output variables	Sensitivity indices		Input variables							
			I_0		β		τ		Var	
Prevalence peak size	PRCC(p-value)	Rank	-0.05(0.27)	**4**	0.93*(0)	**2**	0.96*(0)	**1**	-0.47*(0)	**3**
	RCC(p-value)		-0.07(0.13)		0.55*(0)		0.76*(0)		-0.15*(0.001)	
Prevalence peak time	PRCC(p-value)	Rank	-0.17*(3×10^{-4})	**4**	-0.96*(0)	**1**	-0.93*(0)	**2**	0.65*(0)	**3**
	RCC(p-value)		0.03(0.53)		-0.75*(0)		-0.54*(0)		0.21*(1×10^{-6})	
Final epidemic size	PRCC(p-value)	Rank	-0.02(0.65)	**3**	0.95*(0)	**1**	0.94*(0)	**2**	0.01(0.81)	**4**
	RCC(p-value)		-0.07(0.14)		0.67*(0)		0.66*(0)		-0.02(0.67)	
Epidemic duration	PRCC(p-value)	Rank	-0.02(0.65)	**4**	-0.95*(0)	**1**	-0.94*(0)	**2**	0.79*(0)	**3**
	RCC(p-value)		0.06(0.17)		-0.69*(0)		-0.58*(0)		0.31*(1×10^{-12})	
Basic reproduction number	PRCC(p-value)	Rank	-0.02(0.64)	**3**	0.95*(0)	**1**	0.94*(0)	**2**	0.01(0.84)	**4**
	RCC(p-value)		-0.07(0.14)		0.67*(0)		0.66*(0)		-0.02(0.67)	

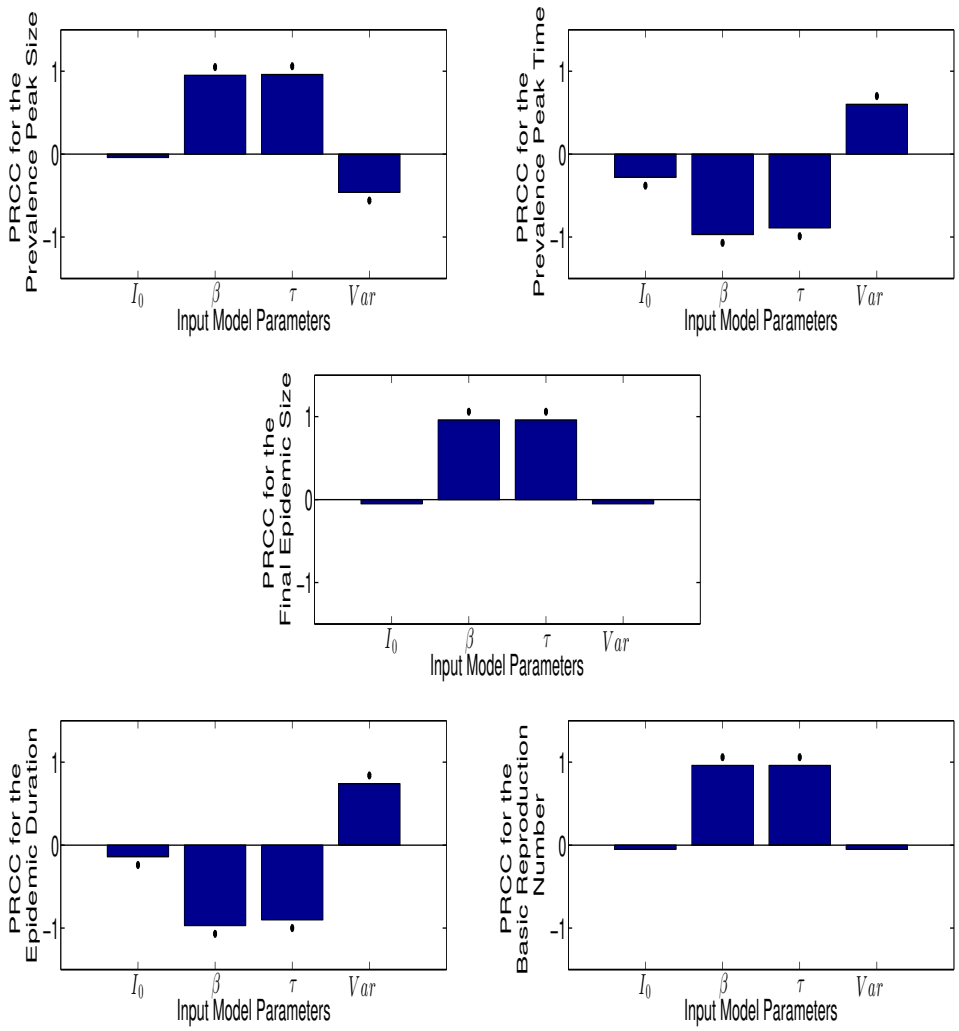


Figure 3.14: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Partial rank correlation coefficients (PRCC)

Table 3.14: Case 1B (Low basic reproduction numbers and Gamma distributed infectious period): Partial rank correlation coefficients (PRCC) and Spearman rank correlation coefficient (RCC) with associated probability values for each outcome variable

Output variables	Sensitivity indices		Input variables							
			I_0		β		τ		Var	
Prevalence peak size	PRCC(p-value)	Rank	-0.04(0.37)	**4**	0.95*(0)	**2**	0.96*(0)	**1**	-0.46*(2×10^{-16})	**3**
	RCC(p-value)		-0.02(0.7)		0.63*(0)		0.73*(0)		-0.13*(0.003)	
Prevalence peak time	PRCC(p-value)	Rank	-0.28*(5×10^{-8})	**4**	-0.97*(0)	**1**	-0.89*(0)	**2**	0.6*(0)	**3**
	RCC(p-value)		-0.03(0.56)		-0.86*(0)		-0.41*(0)		0.16*(2×10^{-4})	
Final epidemic size	PRCC(p-value)	Rank	-0.05(0.29)	**4**	0.96*(0)	**2**	0.96*(0)	**1**	-0.05(0.28)	**3**
	RCC(p-value)		-0.02(0.65)		0.68*(0)		0.69*(0)		-0.03(0.47)	
Epidemic duration	PRCC(p-value)	Rank	-0.14*(0.003)	**4**	-0.97*(0)	**1**	-0.90*(0)	**2**	0.74*(0)	**3**
	RCC(p-value)		0.004(0.93)		-0.83*(0)		-0.44*(0)		0.24*(5×10^{-8})	
Basic reproduction number	PRCC(p-value)	Rank	-0.05(0.27)	**3**	0.96*(0)	**2**	0.96*(0)	**1**	-0.05(0.28)	**4**
	RCC(p-value)		-0.21(0.65)		0.68*(0)		0.69*(0)		-0.03(0.47)	

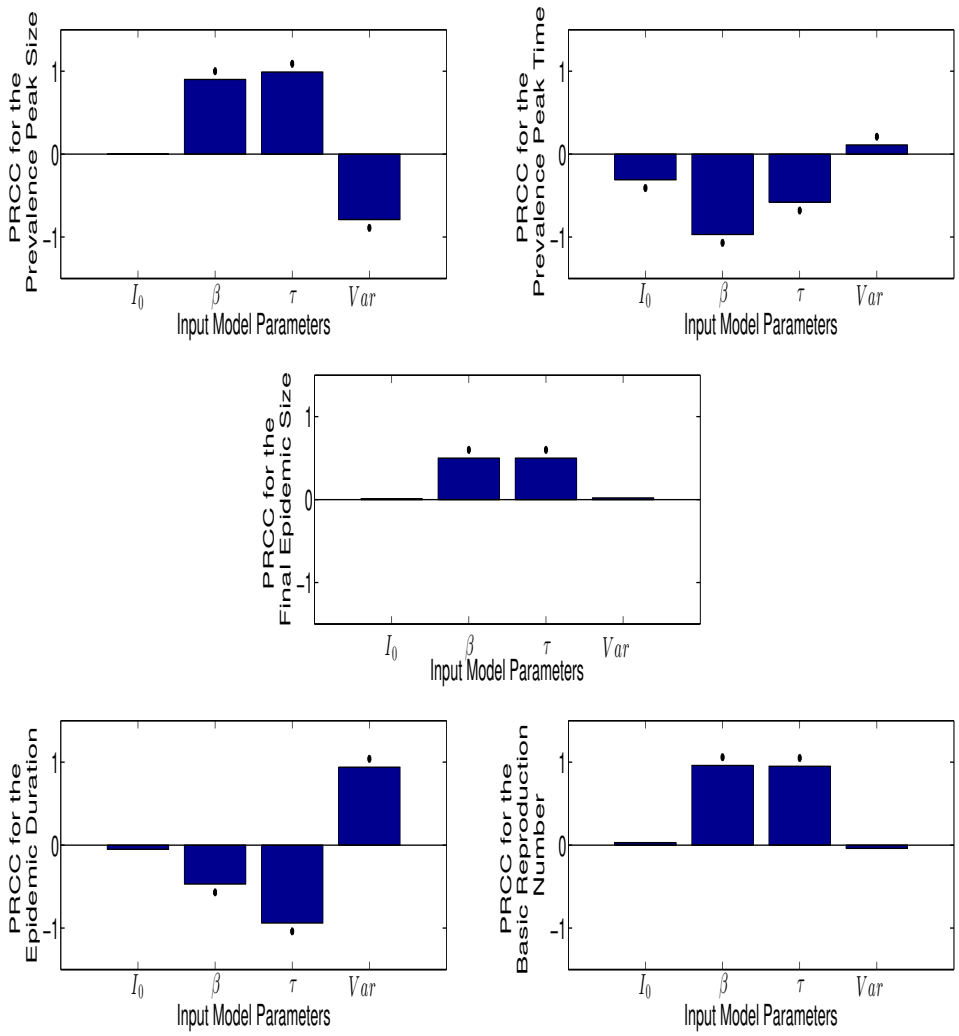


Figure 3.15: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Partial rank correlation coefficients (PRCC)

Table 3.15: Case 2A (High basic reproduction numbers and exponentially distributed infectious period): Partial rank correlation coefficients (PRCC) and Spearman rank correlation coefficient (RCC) with associated probability values for each outcome variable

Output variables	Sensitivity indices		Input variables							
			I_0		β		τ		*Var*	
Prevalence peak size	PRCC(p-value)	Rank	-0.002(0.98)	**4**	0.90*(0)	**2**	0.99*(0)	**1**	-0.79*(0)	**3**
	RCC(p-value)		0.03(0.66)		0.32*(1×10^{-8})		0.94*(0)		-0.16*(0.004)	
Prevalence peak time	PRCC(p-value)	Rank	-0.31*(5×10^{-6})	**3**	-0.97*(0)	**1**	-0.58*(5×10^{-14})	**2**	0.11*(0.04)	**4**
	RCC(p-value)		-0.03(0.65)		-0.96*(0)		-0.21*(3×10^{-4})		0.05*(0.37)	
Final epidemic size	PRCC(p-value)	Rank	0.01(0.90)	**4**	0.50*(0)	**1**	0.50*(0)	**2**	0.02(0.74)	**3**
	RCC(p-value)		0.06(0.29)		0.68*(0)		0.67*(0)		0.07(0.22)	
Epidemic duration	PRCC(p-value)	Rank	-0.05(0.38)	**4**	-0.47*(7×10^{-11})	**3**	-0.94*(0)	**1**	0.94*(0)	**2**
	RCC(p-value)		-0.05(0.41)		-0.18*(0.001)		-0.70*(0)		0.65*(0)	
Basic reproduction number	PRCC(p-value)	Rank	0.03(0.65)	**4**	0.96*(0)	**1**	0.95*(0)	**2**	-0.04(0.52)	**3**
	RCC(p-value)		-0.01(0.93)		0.73*(0)		0.69*(0)		-0.02(0.74)	

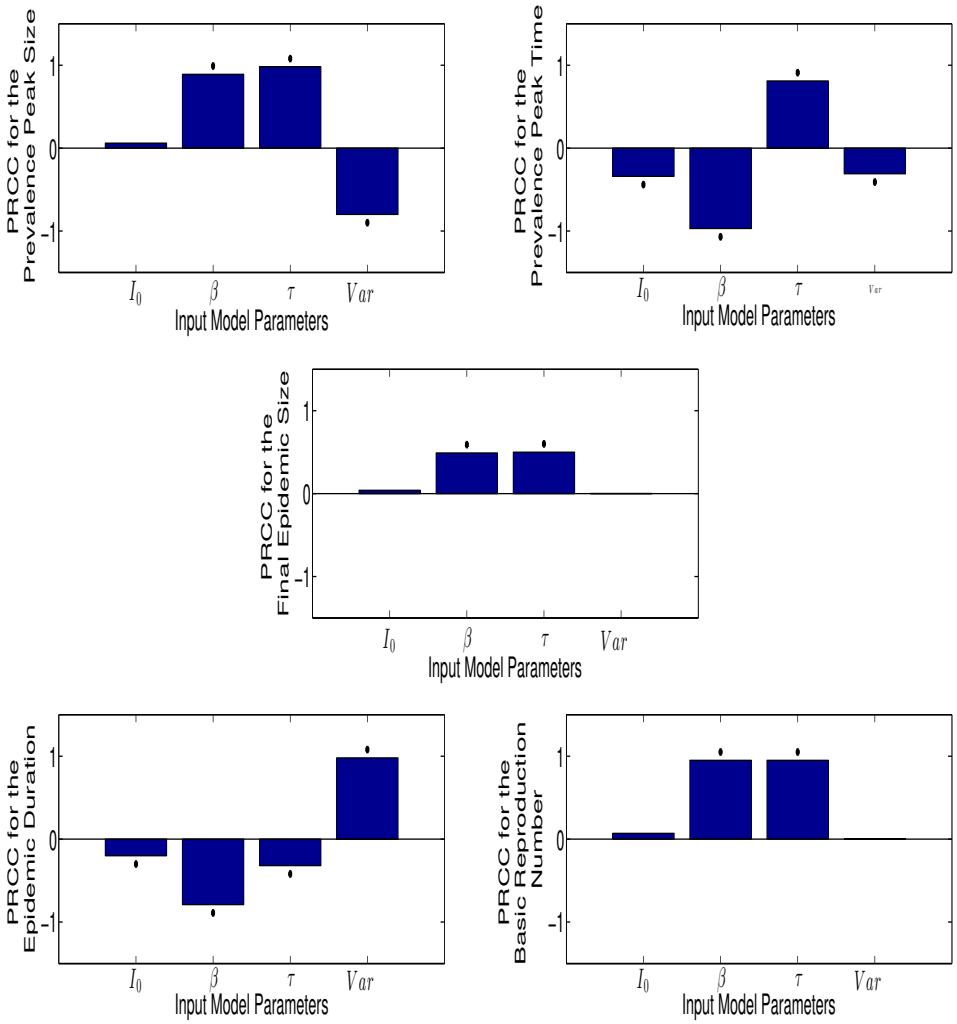


Figure 3.16: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Partial rank correlation coefficients (PRCC)

Table 3.16: Case 2B (High basic reproduction numbers and Gamma distributed infectious period): Partial rank correlation coefficients (PRCC) and Spearman rank correlation coefficient (RCC) with associated probability values for each outcome variable

Output variables	Sensitivity indices		Input variables							
			I_0		β		τ		Var	
Prevalence peak size	PRCC(p-value)	Rank	0.06(0.2)	**4**	0.89*(0)	**2**	0.98*(0)	**1**	-0.80*(0)	**3**
	RCC(p-value)		0.05(0.26)		0.37*(0)		0.88*(0)		-0.29*(6×10^{-11})	
Prevalence peak time	PRCC(p-value)	Rank	-0.34*(1×10^{-10})	**3**	-0.97*(0)	**1**	0.81*(0)	**2**	-0.31*(4×10^{-9})	**4**
	RCC(p-value)		-0.08(0.07)		-0.91*(0)		0.33*(4×10^{-14})		-0.07(0.11)	
Final epidemic size	PRCC(p-value)	Rank	0.04(0.41)	**3**	0.49*(0)	**2**	0.50*(0)	**1**	4×10^{-4}(1)	**4**
	RCC(p-value)		0.14*(0.002)		0.64*(0)		0.67*(0)		0.03(0.49)	
Epidemic duration	PRCC(p-value)	Rank	-0.20*(3×10^{-5})	**4**	-0.79*(0)	**2**	-0.32*(6×10^{-10})	**3**	0.98*(0)	**1**
	RCC(p-value)		-0.03(0.48)		-0.28*(1×10^{-10})		-0.11*(0.01)		0.94*(0)	
Basic reproduction number	PRCC(p-value)	Rank	0.07(0.1)	**3**	0.95*(0)	**1**	0.95*(0)	**2**	-3×10^{-4}(0.99)	**4**
	RCC(p-value)		0.06(0.20)		0.70*(0)		0.69*(0)		-0.04(0.32)	

3.3.3 Local graphical approach

The qualitative relations between the outcome variables considered and the variance of the distribution for the infectious period *Var* are illustrated graphically in **Figures 3.17 and 3.20**. The results from this local and graphical approach are in agreement with the qualitative results obtained from the signs of the sensitivity indices.

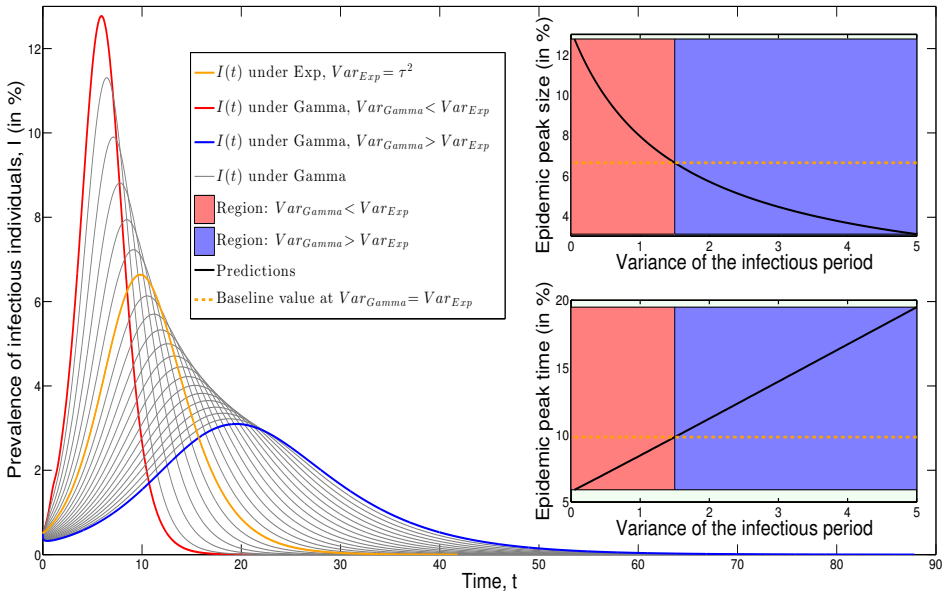


Figure 3.17: The prevalence of infectious individuals in time for various values for the variance of the Gamma distribution of the infectious period ($\mathscr{R}_0 = 1.5$).

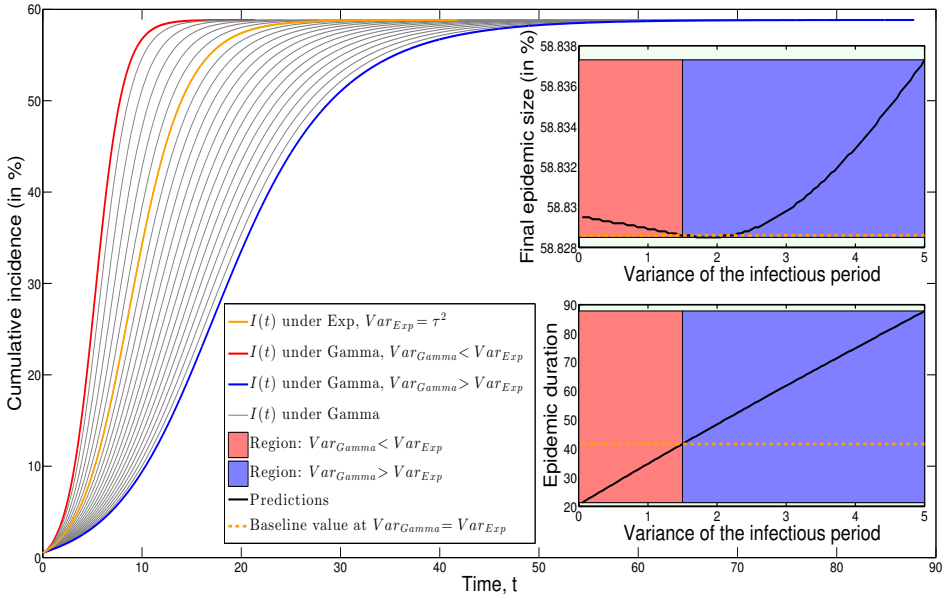


Figure 3.18: The cumulative incidence of infectious individuals in time for various values for the variance of the Gamma distribution of the infectious period ($\mathscr{R}_0 = 1.5$).

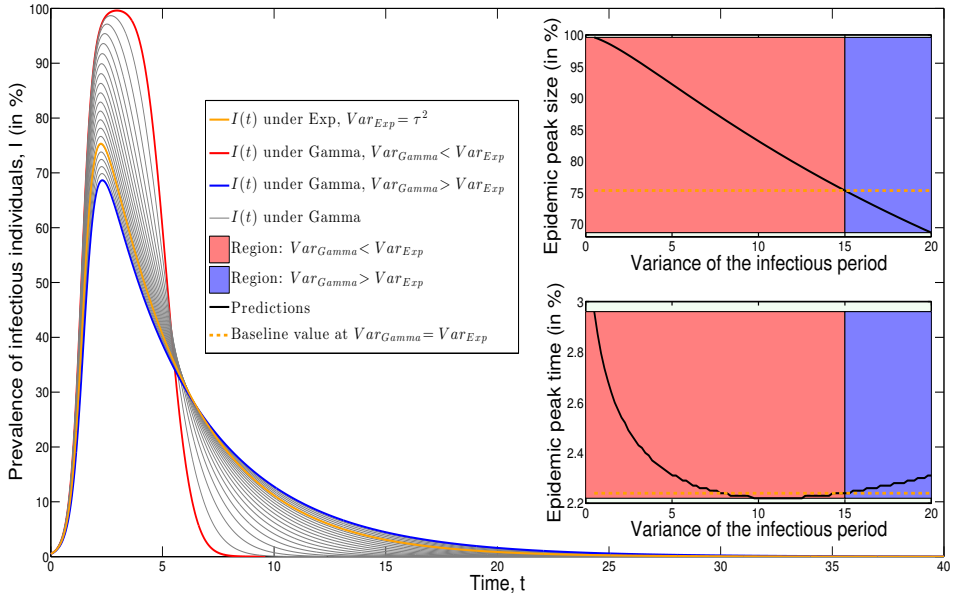


Figure 3.19: The prevalence of infectious individuals in time for various values for the variance of the Gamma distribution of the infectious period ($\mathscr{R}_0 = 15$).

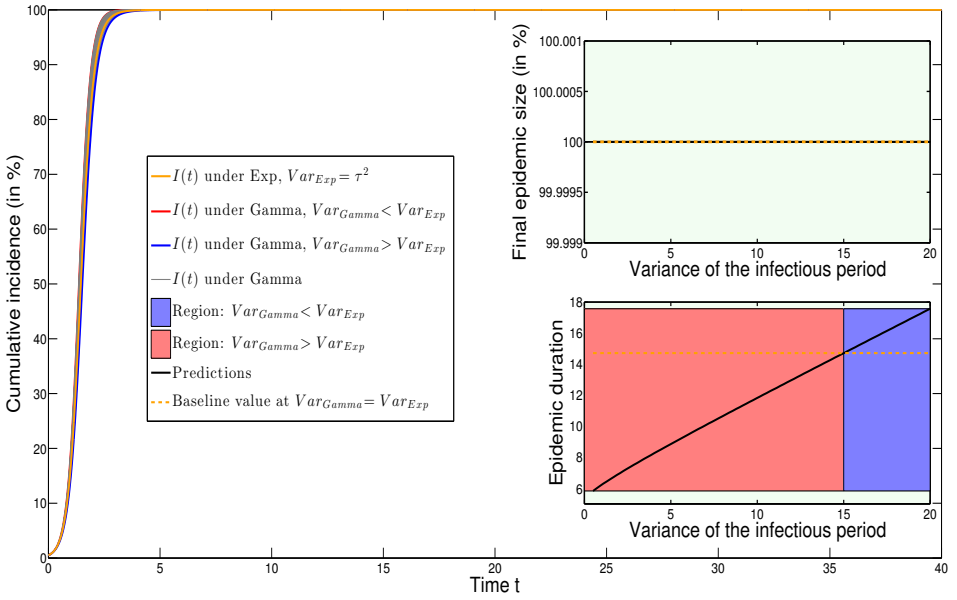


Figure 3.20: The cumulative incidence of infectious individuals in time for various values for the variance of the Gamma distribution of the infectious period ($\mathscr{R}_0 = 15$).

3.4 Conclusions

Based on the combined results from both sensitivity indices, namely the partial rank correlation coefficient (PRCC) and the Spearman's rank correlation coefficient (RCC), the variance of the Gamma distribution for the infectious period embedded in the age-of-infection SIR model is a key input parameter (statistically significant) on the predictability of the following epidemiological variables: the epidemic duration and the peak size and timing of the prevalence of infectious individuals. Therefore, for the predictability of these variables is preferable to utilize a SIR model governed by a nonlinear system of Volterra integral equations, which incorporates the variance of the Gamma distribution for the infectious period as an input model parameter, rather than a nonlinear system of ordinary differential equations; a less flexible option with constant variance for the exponentially distributed infectious period (the "standard"). While for the predictability of epidemiological variables such as the final epidemic size and the basic reproduction number the choice on which type of nonlinear system for the description of the SIR model is irrelevant, since the variance of the infectious period distribution does not play a role. Although, for the latter case, and with the aim of lowering the complexity and number operations in the numerical methods, a nonlinear system of ordinary differential equations is preferred.

3.5 Discussion

The novel application in this work serves as a research decision tool to determine which type of nonlinear system is more appropriate or suitable to utilize for the description of a model: if a Volterra integral equations or ordinary differential equations. This decision is determined by identifying whether or not the variance of a distribution, embedded in a model, is a key input parameter for the predictability of quantities of interest in a research.

The authors suggest to apply the methodology for global uncertainty and sensitivity

analyses via Latin Hypercube Sampling and the computation of appropriate sensitivity indices to epidemic models as a required tool and prior step in the research design, with the potential to prevent (or at least identify) researchers to report incomplete (with respect to the assumption over the disease stage distribution) or misguiding statements or conclusions from quantitative and or qualitative studies or tasks.

Chapter 4

THE EFFECT OF NON EXPONENTIAL DISTRIBUTED INFECTIOUS PERIOD IN A SIMPLE STOCHASTIC AGE-OF-INFECTION SUSCEPTIBLE-INFECTIOUS-RECOVERED MODEL

4.1 Introduction

4.1.1 Problem relevance

In the stochastic setting, the model prediction of an epidemic quantity is represented by a random variable, which follows a probability density function. Probability density functions are useful for calculating the probability of particular events. For example, the probability that the basic reproduction number is greater than the epidemic threshold, in other words, what is the probability that an epidemic will occur? can be computed from the probability density function for the basic reproduction number. Another example, the probability that no more than x percentage of the total population will be infected at the end of an epidemic, can be determined from the probability density function for the final epidemic size. A probability density function can be approximated by an empirically distribution, which is obtained, simulation based, via a stochastic model. Hence, the importance of assessing the effect that modeling assumptions and input model parameters have on the estimation of the empirical distribution.

4.1.2 Research question

The aim of chapter four is to determine whether the empirical distributions for epidemic quantities such as: epidemic duration, prevalence peak size, prevalence peak time and final epidemic size, obtained by assuming exponential versus non-exponential (Gamma)

distributed infectious period are drawn from the same probability distribution function.

4.1.3 Background

As previous work: in 1980, D. Anderson and R. Watson considered the general model formulated by N. T. J. Bailey in 1964 to assess the effect of the shape parameter of the Gamma distribution for the infectious period on the distribution of the final epidemic size [29]; and more recently, in 2010 E. Vergu *et al.* assessed the distributional effect of the Gamma family of distributions for the infectious period, on the distribution of some epidemiological quantities obtained from realizations of a stochastic metapopulation epidemic model [38].

4.2 Methods

4.2.1 General approach

The research aim mentioned above is addressed by first, rewriting the deterministic model governed by a system of Volterra integral equations (see equations 2.4-2.6), under the assumption that the infectious period is Erlang distributed, into a larger dimensional system of ordinary differential equations. This is done via the linear chain trickery (see **appendix B** for details on the derivation). From the latter deterministic (ODE's) model, a corresponding simple stochastic age-of-infection susceptible-infectious-recovered model is established, as a continuous time Markov chain model (see [29]). The stochastic model is solved numerically via the Gillespie's Direct algorithm (see [42]). A pair of empirical distributions for each epidemiological quantity of interest is obtained. Finally, the Kolmogorov-Smirnov test, which is a non-parametric statistical hypothesis test designed to determine whether two empirical distributions (or samples) are drawn from the same probability distribution function, is applied to each pair of empirical distributions.

4.2.2 Stochastic epidemic model: Continuous-Time Markov Chain Model

A simple stochastic age-of-infection susceptible-infectious-recovered model is developed based on a multivariate Markov jump process

$$X_t = \{(S_t, I_{1,t}, \ldots, I_{n_I,t}, R_t) : t \in \mathbb{R}_+\}$$

with state space $\mathbb{Z}_+^{n_I+2} \cup \{0\}$. The model considers a total of $n_I + 1$ events: *infection*, *progression from the infectious stage I_i to I_{i+1}* and *removal*. The stochastic rates of the process or intensities are defined as:

Table 4.1: Stochastic events and their rates.

Event	From	To	Rate
Infection	X	$X + (e_2 - e_1)^T$	$\beta SI/N$
Progression from the infectious stage I_i to I_{i+1} for $i = 1, \ldots, n_I - 1$	X	$X + (e_{i+2} - e_{i+1})^T$	$n_I \gamma I_i$
Removal	X	$X + (e_{n_I+2} - e_{n_I+1})^T$	$n_I \gamma I_{n_I}$

where $e_i = [0, \ldots, \underbrace{1}_{i^{th}}, \ldots, 0]^T$ is a unit vector from the canonical basis. Notice that the notation $X + (e_{i+1} - e_i)^T = [X_1, \ldots, X_i - 1, X_{i+1} + 1, \ldots, X_{n_I+2}]$ indicates the current state of the process X after the occurrence of an event or "jump".

The corresponding transition probabilities of the events are given by:

$$\begin{aligned}
P(X_{t+\Delta t} - X_t = (e_2 - e_1)^T) &= \frac{\beta_t}{N_t} S_t \sum_{j=1}^{n_I} I_t \Delta t + o(\Delta t) \\
P(X_{t+\Delta t} - X_t = (e_{i+2} - e_{i+1})^T) &= n_I \gamma I_{i,t} \Delta t + o(\Delta t) \\
&\quad \text{for } i = 1, \ldots, n_I - 1 \\
P(X_{t+\Delta t} - X_t = (e_{n_I+2} - e_{n_I+1})^T) &= n_I \gamma I_{n_I,t} \Delta t + o(\Delta t)
\end{aligned}$$

The absorbing states of the process X are

$$(S, I_1, \ldots, I_{n_I}, R) = (s, \overbrace{0, \ldots, 0}^{n_I}, N(t_{extinction}) - s),$$

where $s \in \{1, ..., N(t_1)\}$, and the other states are transient. The waiting times or jump times $0 < W_1 < W_2 < \ldots$ have increments (interevent $\{T_i\}$) exponentially distributed with parameter μ and are given by

$$P(T_i = W_i - W_{i-1} > t | W_j, j \leq i-1) = e^{-t\mu(W_{i-1})},$$

where

$$\begin{aligned}
\mu(W_{i-1}) &= \left(\frac{\beta(W_{i-1})}{N(W_{i-1})} S(W_{i-1}) \Big(\sum_{j=1}^{n_I} I_j(W_{i-1}) \Big) + n_I \gamma I_1(W_{i-1}) + n_I \gamma I_2(W_{i-1}) + \cdots + n_I \gamma I_{n_I}(W_{i-1}) \right)^{-1}, \\
&= \left(\frac{\beta(W_{i-1})}{N(W_{i-1})} S(W_{i-1}) \Big(\sum_{j=1}^{n_I} I_j(W_{i-1}) \Big) + n_I \gamma \Big(\sum_{j=1}^{n_I} I_j(W_{i-1}) \Big) \right)^{-1} \\
&= \left(\frac{\beta(W_{i-1})}{N(W_{i-1})} S(W_{i-1}) I(W_{i-1}) + n_I \gamma I(W_{i-1}) \right)^{-1}.
\end{aligned}$$

The implementation of the stochastic model (Gillespie's Direct algorithm [42]) was written in MATLAB (R2013a, The MathWorks).

4.2.3 Empirical distributions

For each epidemiological quantity of interest previously defined in equations 3.1, 3.2, 3.3, 3.4, 3.5, a pair of empirical distributions is obtained; one under the standard assumption that the infectious period is distributed exponentially and the other one is under the Gamma (or Erlang) distribution assumption (See **Figure 4.1**). Empirical distributions were obtained by carrying out numerical simulations of the stochastic model (1,000 realizations) with the following parameter values: model parameters: $\mathscr{R}_0 = 1.5$ and $\mathscr{R}_0 = 15$, $\beta = \tau = \sqrt{\mathscr{R}_0}$, $n_I = 1$ and $n_I = 4$, Var $= \frac{\tau^2}{n_I}$; initial conditions: $S_0 = 990$, $I_{1,0} = 10$, $I_{i,0} = 0$, for $i = 2, \ldots, n_I - 1$, $R_0 = 0$, $N_0 = N_t = 1{,}000$.

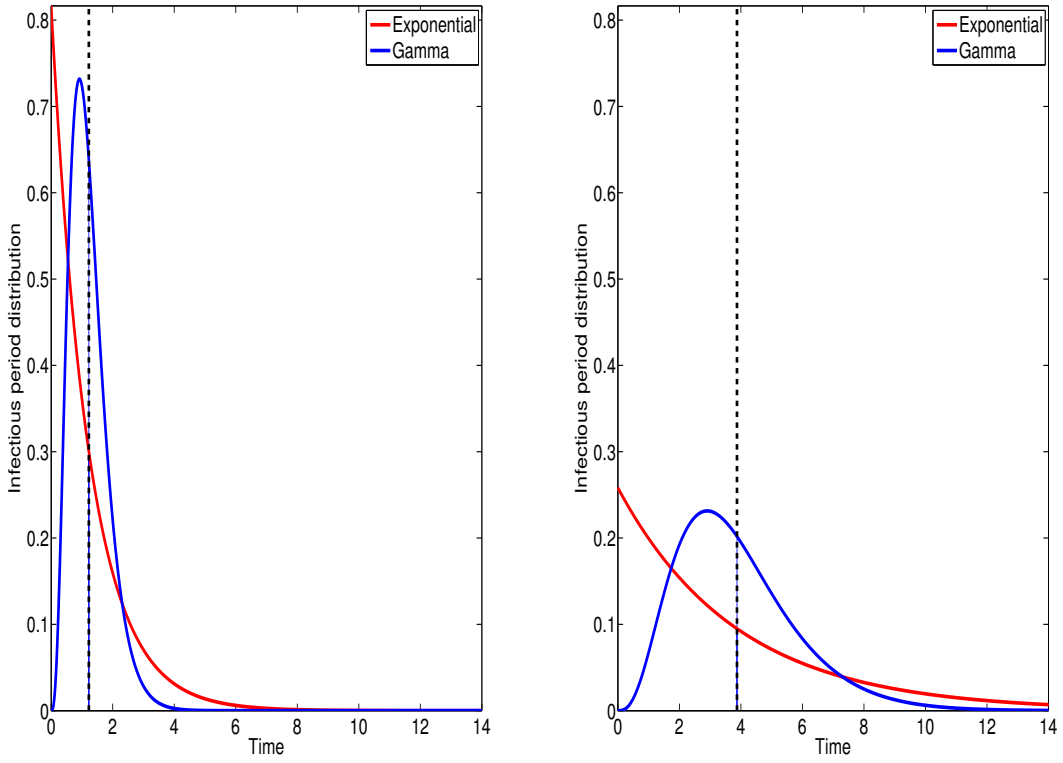


Figure 4.1: Infectious period distributions: (left panel) Exponential ($n_I = 1$) and Erlang ($n_I = 4$) distributions, both under low basic reproduction numbers ($\mathscr{R}_0 = 1.5$) and (right panel) Exponential ($n_I = 1$) and Erlang ($n_I = 4$) distributions, both under high basic reproduction numbers ($\mathscr{R}_0 = 15$).

4.2.4 Statistical hypothesis test

Each pair of empirical distributions were tested statistically by applying the non-parametric Kolmogorov-Smirnov test, which is designed to determine whether two empirical distributions (or samples) are drawn from the same probability distribution function. The MATLAB (R2013a, The MathWorks) built-in routine *kstest2* was used for this task.

4.3 Results

Based on simple descriptive statistics (see **Table 4.3**) and empirical distributions (see **Figures 4.2 to 4.9**) for each of epidemiological quantities considered (final epidemic

size, epidemic duration, prevalence peak size and prevalence peak time): the uncertainty of these is high for the case when the basic reproduction number is relatively low (around $\mathscr{R}_0 = 1.5$). In contrast, for a high basic reproduction number (around $\mathscr{R}_0 = 15$), the uncertainty of the variables mentioned previously is low, with the exemption of the epidemic duration, which still remain high. These results are invariant from the two distribution chosen and assumed for the infectious period, Exponential and Erlang.

Based on outcomes from the Kolmogorov-Smirnov tests: for low to moderate basic reproduction numbers (around $\mathscr{R}_0 = 1.5$), all the pairs of empirical distributions for the epidemiological quantities of interest resulted as statistically dissimilar, with a level of significance of $\alpha = 0.05$. While for high basic reproduction number (around $\mathscr{R}_0 = 15$), the only pair of empirical distributions that resulted as statistically ($\alpha = 0.05$) dissimilar was the distribution corresponding to the epidemic duration.

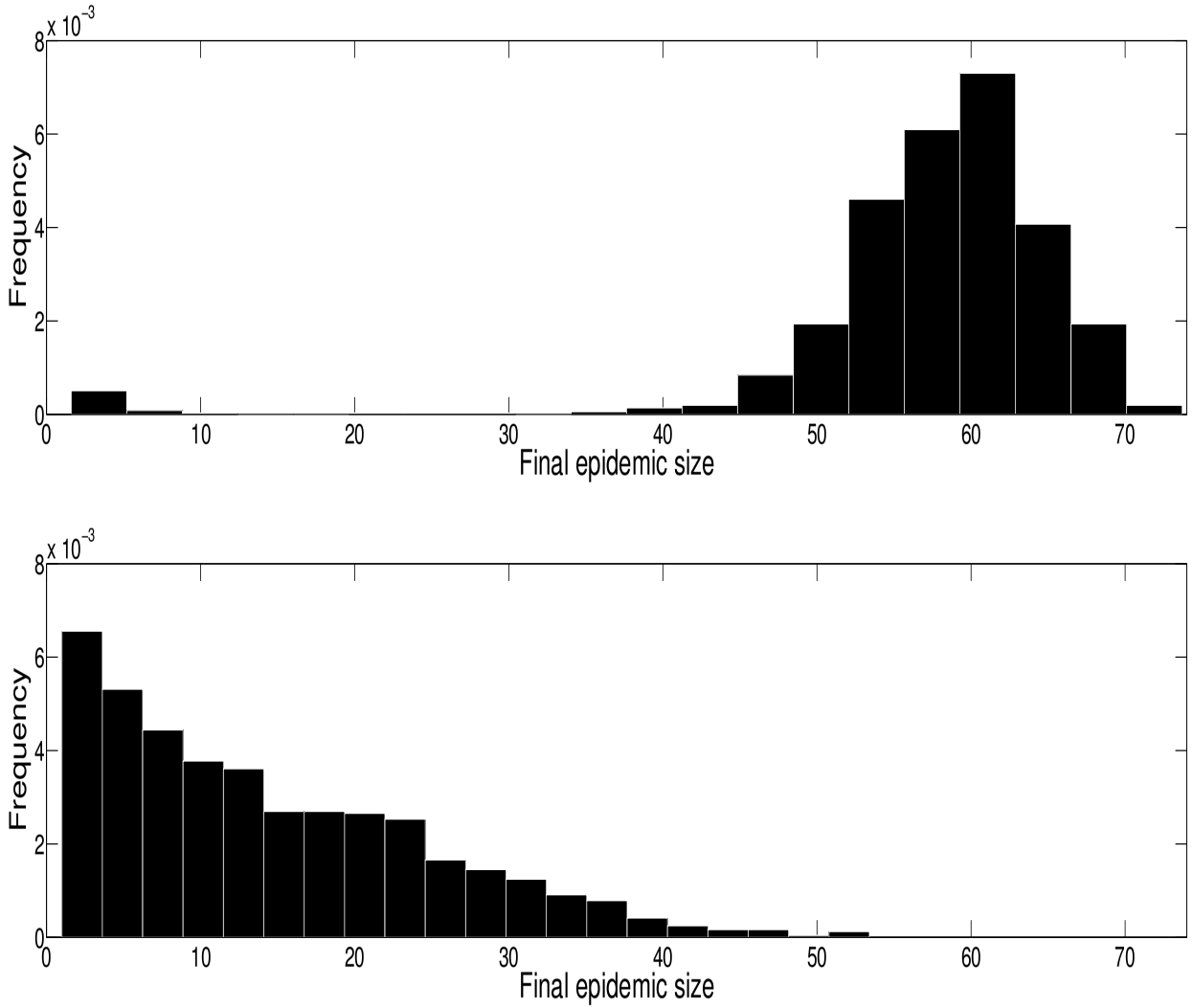


Figure 4.2: Empirical distribution for the final epidemic size under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under low basic reproduction number ($\mathscr{R}_0 = 1.5$).

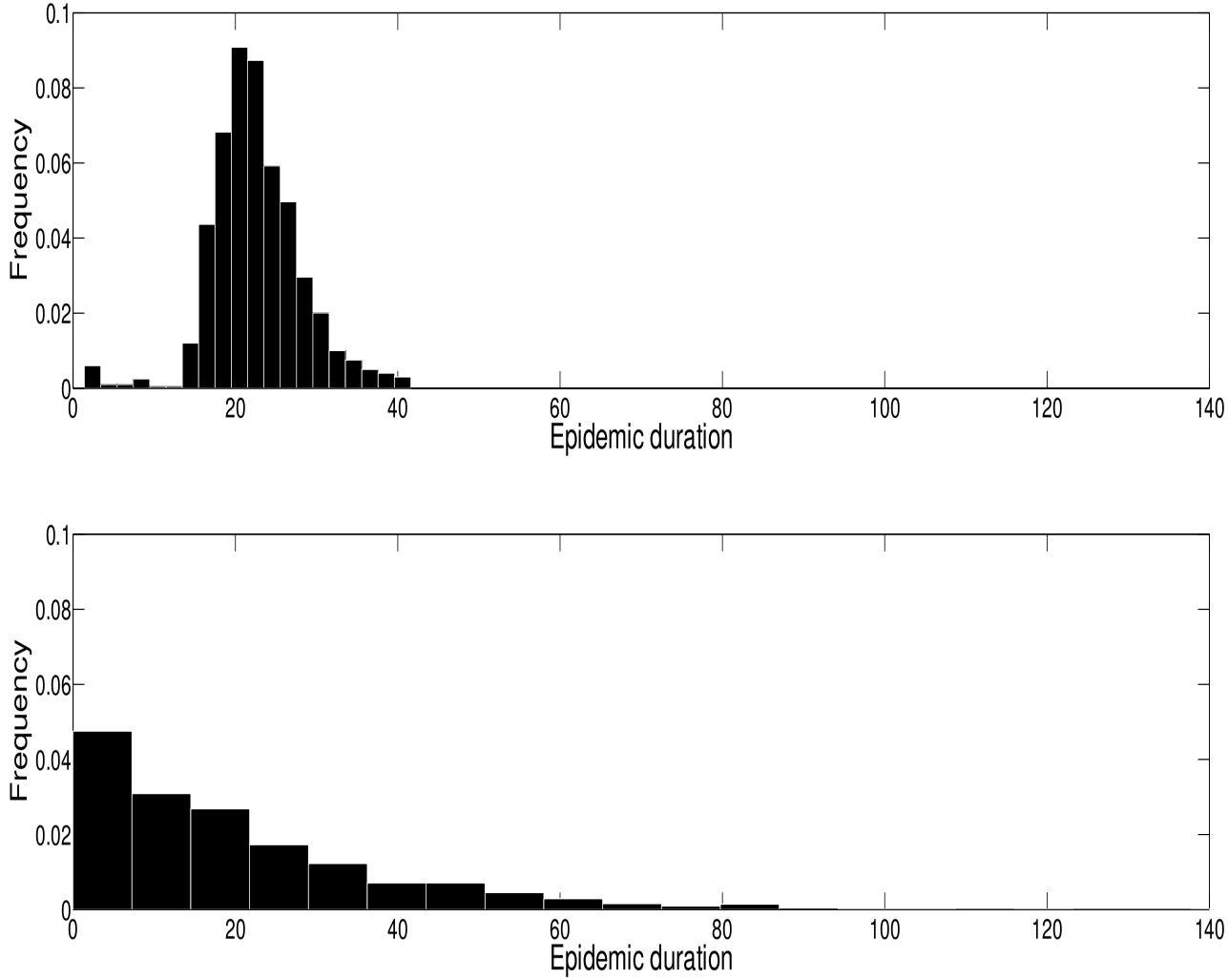


Figure 4.3: Empirical distribution for the epidemic duration under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under low basic reproduction number ($\mathscr{R}_0 = 1.5$).

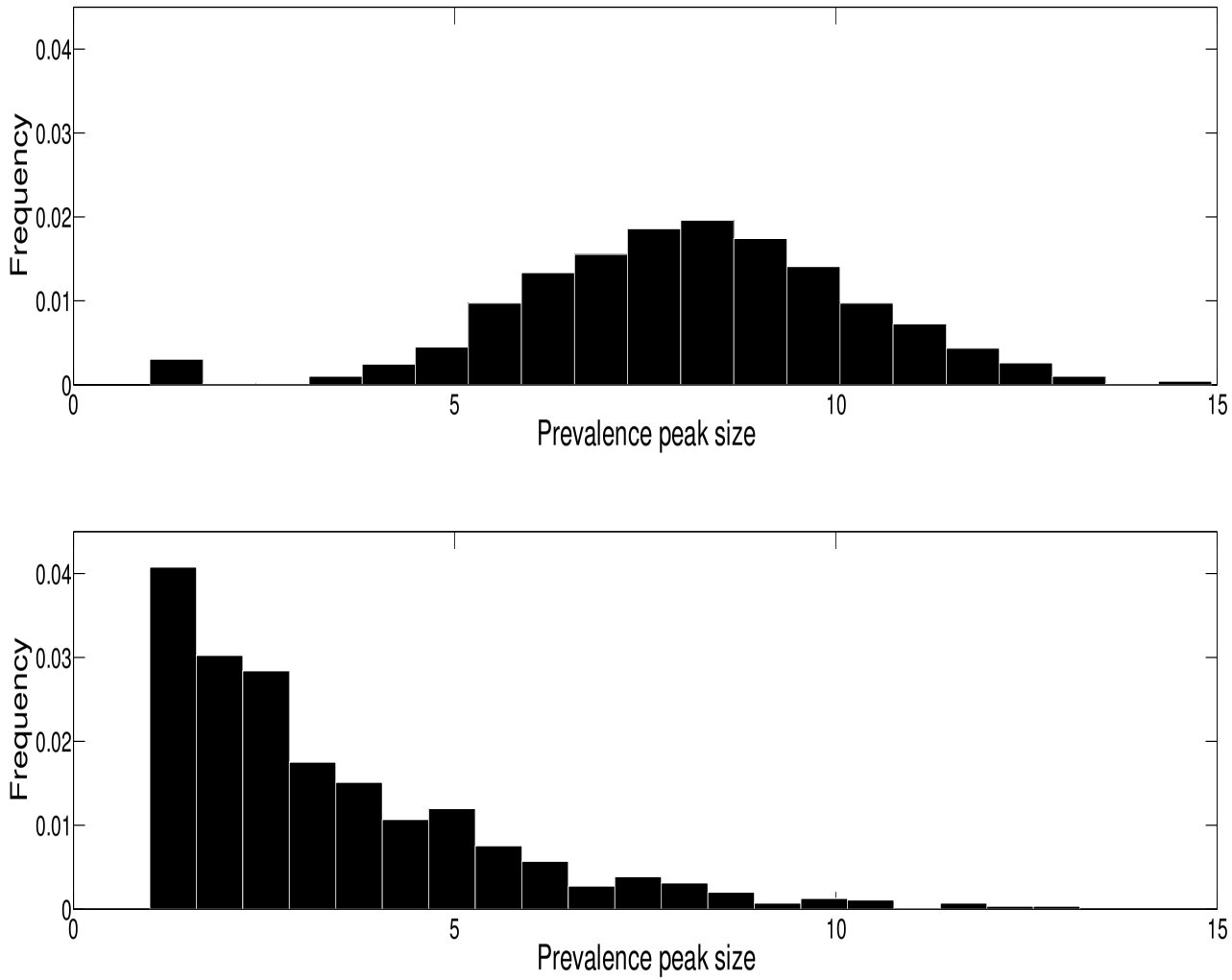


Figure 4.4: Empirical distribution for the prevalence peak size under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under low basic reproduction number ($\mathscr{R}_0 = 1.5$).

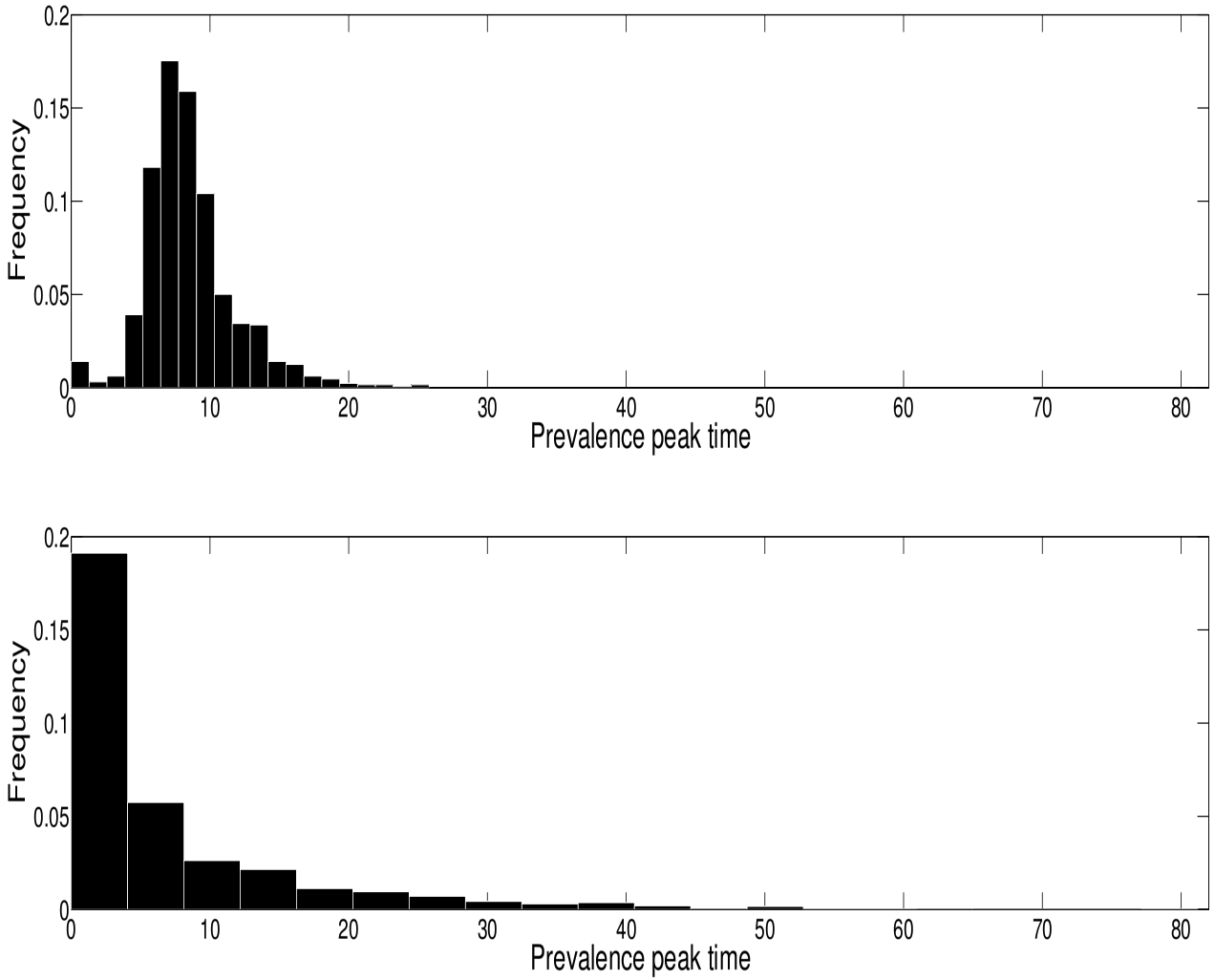


Figure 4.5: Empirical distribution for the prevalence peak time under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under low basic reproduction number ($\mathscr{R}_0 = 1.5$).

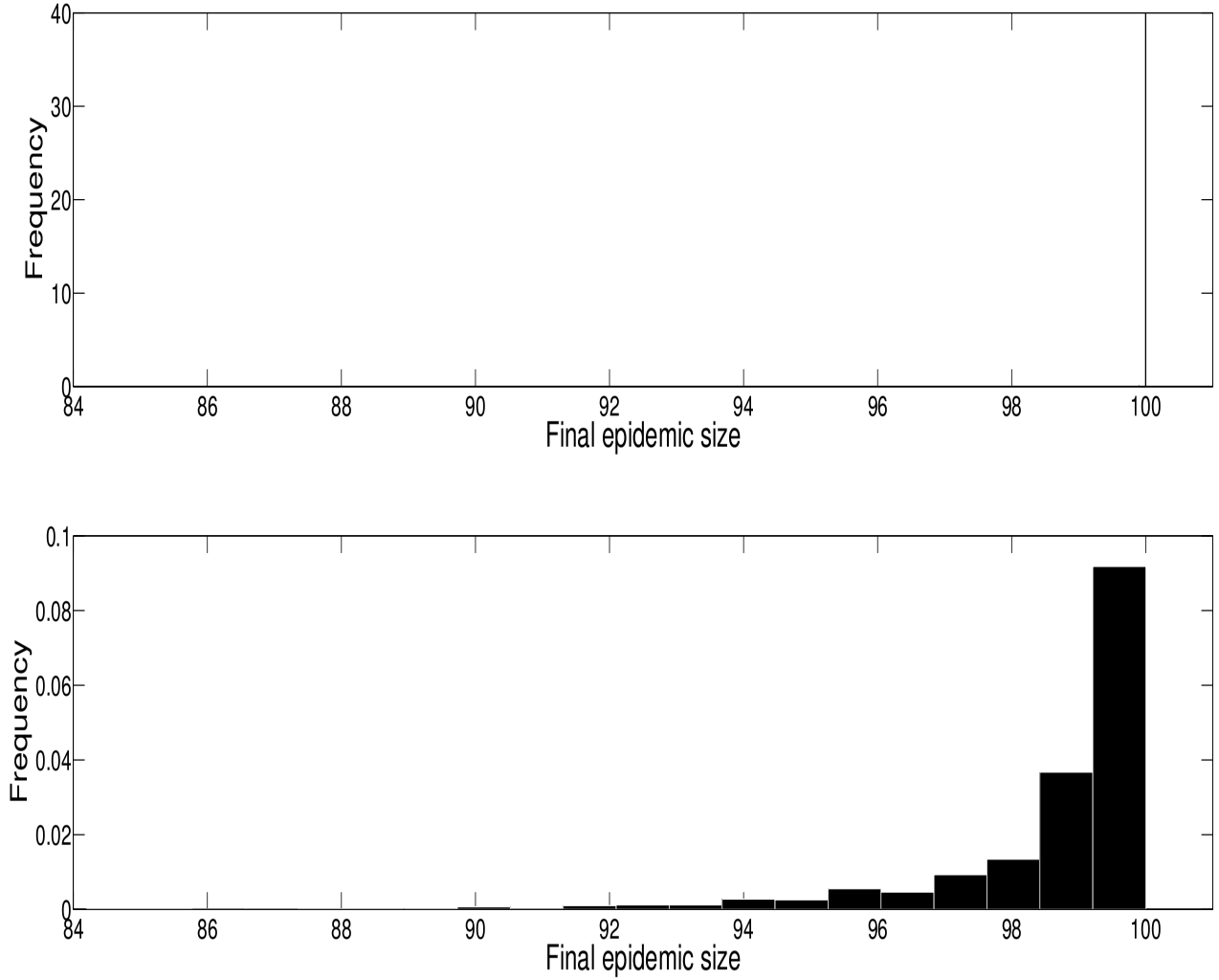


Figure 4.6: Empirical distribution for the final epidemic size under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under high basic reproduction number ($\mathscr{R}_0 = 15$).

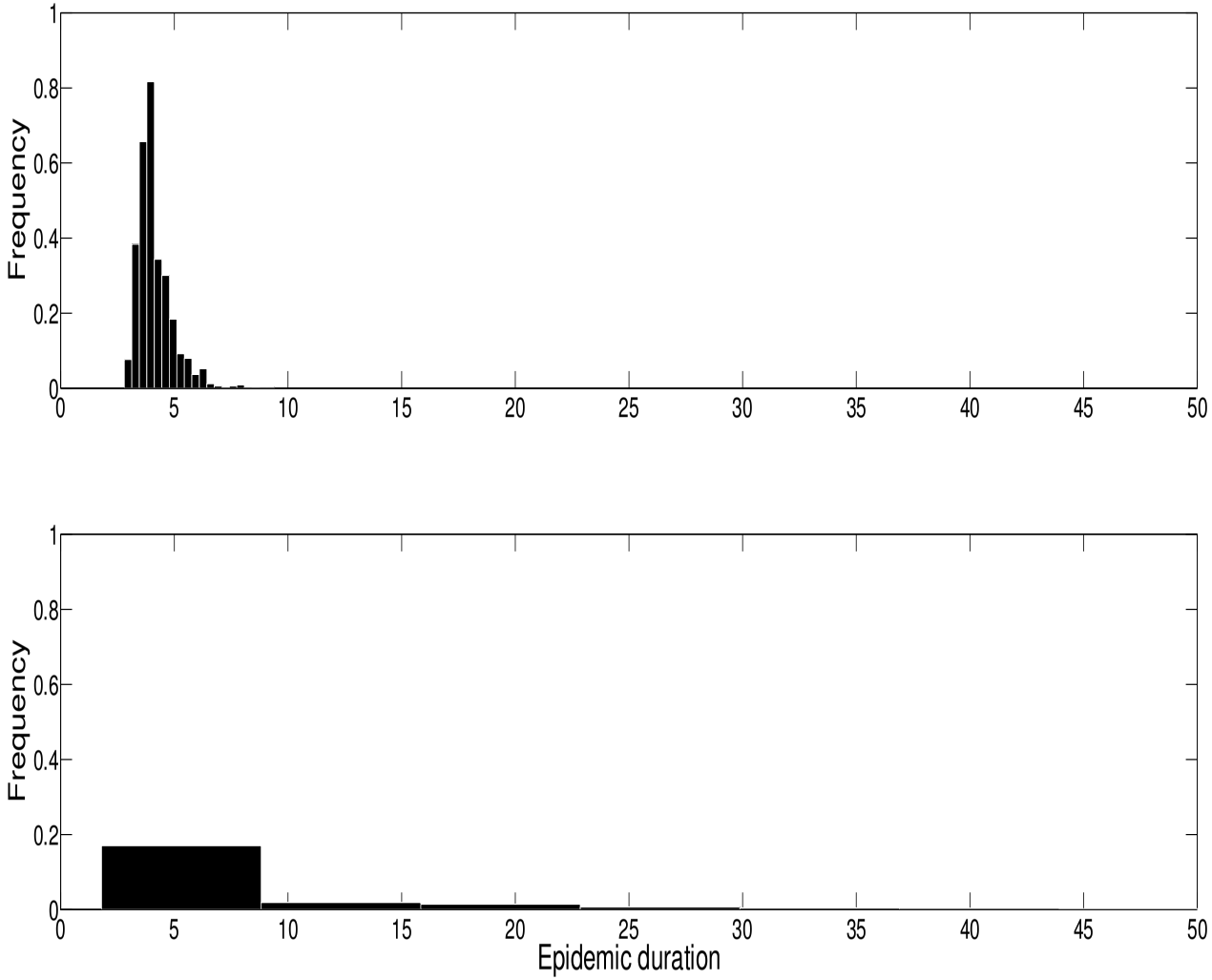


Figure 4.7: Empirical distribution for the epidemic duration under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under high basic reproduction number ($\mathscr{R}_0 = 15$).

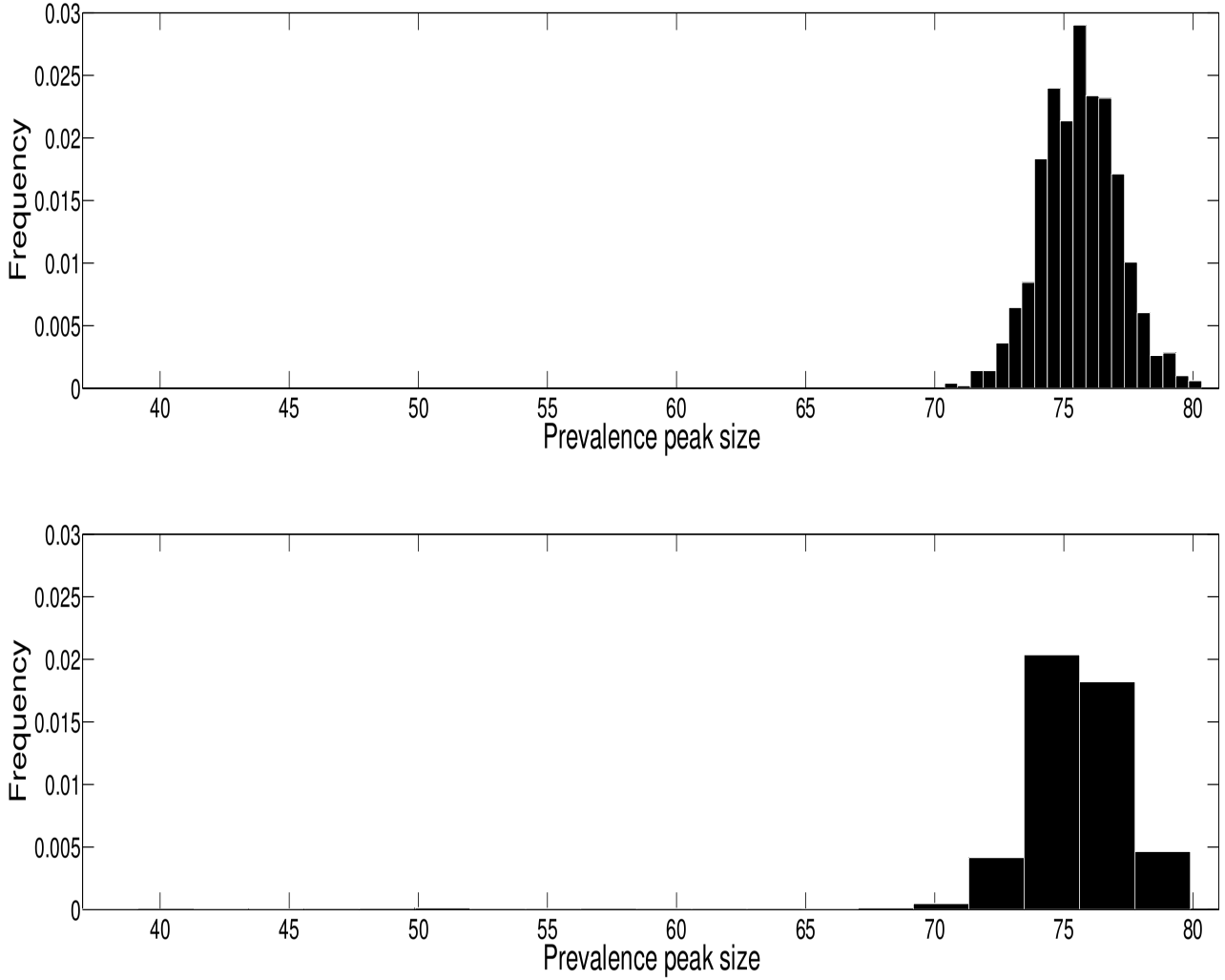


Figure 4.8: Empirical distribution for the prevalence peak size under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under high basic reproduction number ($\mathscr{R}_0 = 15$).

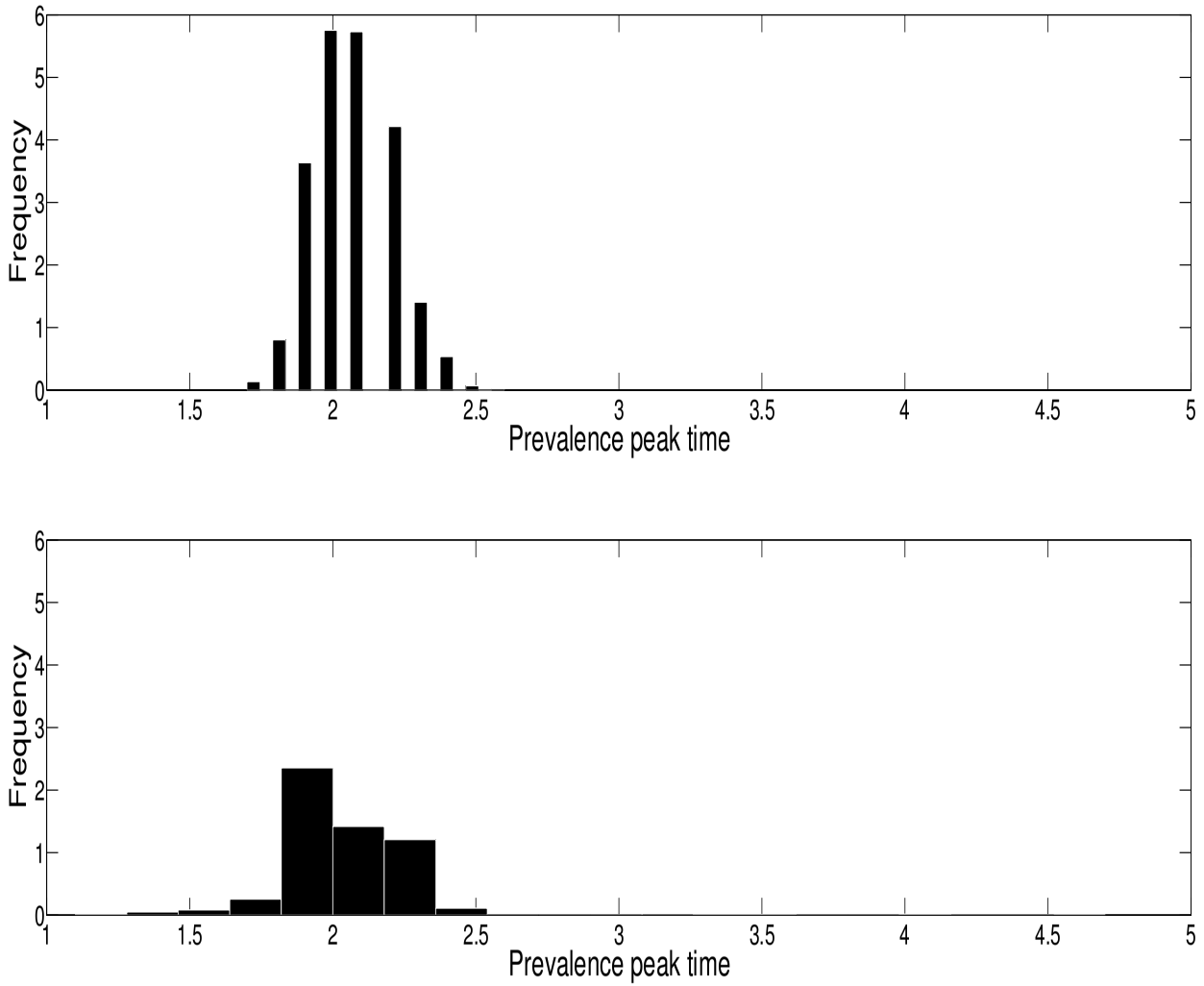


Figure 4.9: Empirical distribution for the prevalence peak time under the Exponential ($n_I = 1$) distribution (top panel) and Erlang ($n_I = 4$) distribution (bottom panel). Both under high basic reproduction number ($\mathscr{R}_0 = 15$).

Table 4.2: Statistics, P-value and test conclusion of the Kolmogorov-Smirnov test for each of the epidemiological variables of interest.

Case	Epidemiological variables	Statistics (P-value)	Conclusion
Low $\mathscr{R}_0$	Final epidemic size	0.96(0)	H_0 is rejected
	Epidemic duration	$0.46(2\times10^{-93})$	H_0 is rejected
	Prevalence peak	$0.76(2\times10^{-251})$	H_0 is rejected
	Prevalence peak time	$0.56(4\times10^{-140})$	H_0 is rejected
High $\mathscr{R}_0$	Final epidemic size	0.947(0)	Fails to reject H_0
	Epidemic duration	$0.387(3\times10^{-66})$	H_0 is rejected
	Prevalence peak	0.047(0.21)	Fails to reject H_0
	Prevalence peak time	0.035(0.57)	Fails to reject H_0

Table 4.3: Simple descriptive statistics for each of the epidemiological variables of interest under the Exponential ($n_I = 1$) distribution and Erlang ($n_I = 4$) distribution. The values of those statistics associated with the number of infectious cases, are reported as a percentage of the total initial population size ($N = 1000$).

Case	Epidemiological variables	Case	min	max	mean	median	mode	std	var
Low $\mathscr{R}_0$	Final epidemic size	Exp.	1.6	73.7	57.35	58.95	61.4	9.99	99.71
		Γ	1.0	53.4	14.3417	12.1	1.8	10.57	111.79
	Epidemic duration	Exp.	1.4	41.6	22.56	22.1	22.3	5.67	32.11
		Γ	0	145.0	20.18	15.1	0.4	19.07	363.49
	Prevalence peak	Exp.	1.0	14.93	8.06	8.12	1.00	2.25	5.05
		Γ	1.0	13.2	3.36	2.71	1.0	2.16	4.66
	Prevalence peak time	Exp.	0	25.8	8.57	8.0	7.7	3.31	10.93
		Γ	0	81.20	7.23	3.15	0.1	10.03	100.69
High $\mathscr{R}_0$	Final epidemic size	Exp.	99.9	100.0	99.99	100.0	100.0	0.004	0.00002
		Γ	842	100.0	98.59	99.3	99.8	1.95	3.82
	Epidemic duration	Exp.	2.8	9.4	4.13	3.9	3.8	0.8	0.64
		Γ	1.8	142.2	9.56	3.4	2.9	15.05	226.36
	Prevalence peak	Exp.	70.39	80.33	75.6	75.59	70.39	1.5	2.27
		Γ	37.0	79.9	75.15	75.5	75.4	3.59	12.85
	Prevalence peak time	Exp.	1.7	2.6	2.07	2.1	2	0.14	0.02
		Γ	1.1	4.7	2.06	2.1	2	0.25	0.06

4.4 Conclusions

The only distinction between the two different established stochastic models is the value for the variance of the (Erlang) distribution of the infectious period. The first model, assumes exponentially distributed infectious period, or equivalently, $Var = \tau^2$ and a second model, assumes Erlang distributed infectious period, or equivalently, $Var = \frac{\tau^2}{4}$.

Hence, for relatively low basic reproduction number (around $\mathscr{R}_0 = 1.5$), the variance of the infection period distribution, indirectly, is a key statistical significant ($\alpha = 0.05$) input model parameter in the estimation of empirical distributions of epidemiological variables such as: final epidemic size, epidemic duration, prevalence peak size and prevalence peak time. In other words, for relatively low basic reproduction number (around $\mathscr{R}_0 = 1.5$), the distribution of the infection period is an important aspect to be considered in the estimation of the probability of a particular event, defined from the empirical distribution of any of the random variables of interest.

4.5 Discussion

In general, the framework introduced in this work can be applicable to other models, with aim to determine whether or not the variance (or shape) of a Erlang distribution of a stage, embedded in a model, is a key input parameter, on the estimation of the probability of a particular event, definied from the empirical distribution of a random variable of interest.

Chapter 5

PARAMETER ESTIMATION ON A SIMPLE AGE-OF-INFECTION SUSCEPTIBLE-INFECTIOUS-RECOVERED MODEL

5.1 Introduction

5.1.1 Background

Parameter estimation from observed data using the least-square estimation procedure has become a highly popular and useful tool in many scientific fields. The method helps to determine estimates of significance for non-obvious quantities (*i.e.* transmission rate, mean and variance for the infectious period distribution, etc.) obtained from mathematical models that describe the underlying mechanisms of a particular process, thus allowing the possibility of predictions. In infectious diseases epidemiology, it has been applied in studies of emerging and re-emerging infectious diseases such as Ebola [43], the Spanish flu pandemic [44, 45], Degue fever [46], SARS [47], and Pneumococcal diseases [48], just to mention a few. Typically, a parameter estimation problem is subject to a deterministic epidemiological model governed by a system of nonlinear ordinary differential equations, which implicitly assume exponential distributed disease stages. In this chapter, the standard (exponentially distributed infectious period case) and a general case with respect to the infectious period distribution are considered, where a parameter estimation problem is subject to a deterministic epidemiological model described by a nonlinear system of Volterra integral equations. On the non-exponentially distributed disease stages case, there is not much work done. There are two related studies that can be mentioned. A first study, in 2005, where H. J. Wearing *et al.* observed that the exponentially distributed latent and infectious periods assumption leads to overoptimistic results in comparison to the Erlang distributed

latent and infectious periods [31]. This observation was based on predictions of the basic reproduction number calculated from observed field data, indirectly, through the estimation of parameters of various epidemiological models under the assumption of different combinations of the Erlang probability density functions for the latent and infectious periods [31]. In their work, all the epidemiological models considered are described by nonlinear systems of ordinary differential equations. A disadvantage of their approach (linear chain trickery) is that it makes impossible the estimation of shape parameter or equivalently the variance of the distribution for the latent and or infectious period, as input model parameters. More recently, a second study, in 2011, B. P. Holder and C. A. A. Beauchemin obtained different estimates for key input parameters in SEIR models. The estimates are computed indirectly via solving the least-square problem with deterministic models under the assumption of different distributions for the diseases stages, such as: Exponential, fixed period (or "δ" distribution), Normal and Log-normal [49]. In their work, the epidemiological models considered are described by a nonlinear system of Volterra integral equations. The parameters of the probability density functions are not estimated. They used a "modified Euler technique" as for the numerical scheme to solves the nonlinear system of Volterra integral equations, which is not as accurate as using a 4th-stage Volterra-Runge-Kutta formula of Pouzet type (see **appendix D**), and consequently the numerical inaccuracy could have an impact on the reported results and main conclusions.

5.1.2 Research questions of interest

Chapter five is devoted to the exploration of the following research questions.

Question one: What is the **quantitative effect** of the standard modeling assumption of exponentially distributed infectious period on the estimates of input parameter and associated predictions of outcome variables?

Question two: Can the probability density function for the infectious period of a particular

infectious disease be identifiable from epidemiological data, indirectly, by using a SIR model governed by a nonlinear system of Volterra integral equations?

5.2 Methods

5.2.1 General approach

The research questions are mainly addressed by pursuing estimations of input parameters of a simple age-of-infection Susceptible-Infectious-Recovered model (2.4-2.6), indirectly, from artificially generated incidence data (see **Figures 5.1**), via the least-square estimation procedure.

5.2.2 Model description

We considered the same simple age-of-infection Susceptible-Infectious-Recovered model in (2.4-2.6) described in details in chapter two.

5.2.3 Description of data

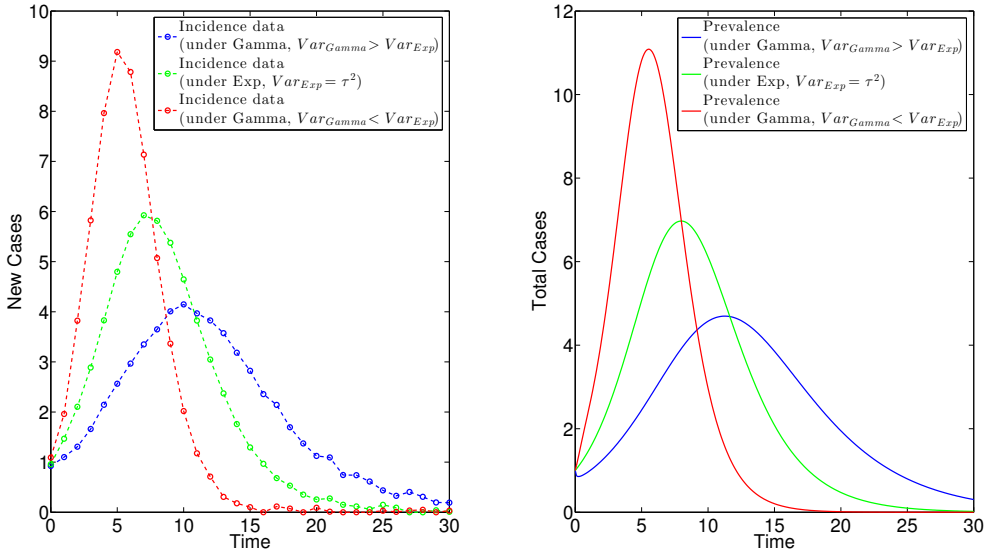
Sets of artificially (or synthetic) generated data for the *incidence* are obtained by specifying an explicit error term (or function) in the statistical model (see 5.1), which is a function of the deterministic states variables of the age-of-infection SIR model in (2.4-2.6). Two scenarios are considered for the sets *incidence* data: low and high transmissibility. Each scenario has three cases, which assume three types of infectious period distribution (see **Figures 5.1**):

i) A Gamma probability density function with a larger variance than the variance for the Exponential distribution (shape parameter $k = 0.5$);

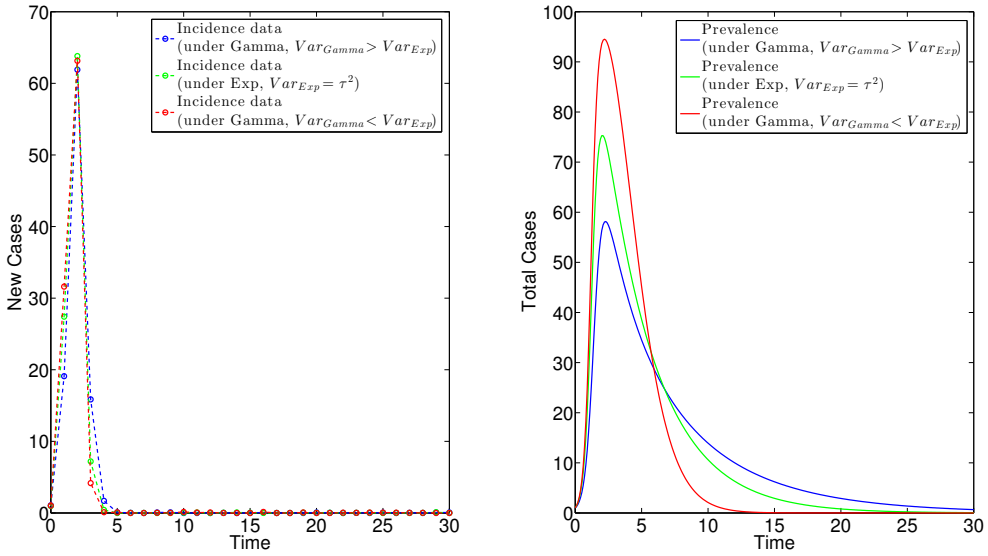
ii) An Exponential distribution (the standard assumption $k = 1$);

iii) A Gamma distribution with a lower variance than the variance for the Exponential distribution (shape parameter $k = 4$).

The model parameters used to generate the incidence data are provided in **Tables 5.1, 5.4 and 5.5**.



(a) Low transmissibility, $\mathscr{R}_0 = 1.5$.



(b) High transmissibility, $\mathscr{R}_0 = 15$.

Figure 5.1: Incidence data (left panels) and prevalence baselines (left panels) of infectious individuals for low (top panels) and high (bottom panels) transmissibility under different infectious period distributions.

5.2.4 Parameter estimation or inverse problem

The least-square approximation is applied to the age-of-infection SIR model in (2.4-2.6) and the different synthetic sets of incidence data (see **Figures 5.1**). Two sets of model input parameters to estimate are defined: i) a first set $\vec{\theta}_{0,\text{Exp}}$, where the variance of the assumed Exponentially distributed infectious period is fixed ($Var = \tau^2$) and for which case the model is described as a nonlinear system of ordinary differential equations (see **appendix A**); and a second set $\vec{\theta}_{0,\Gamma}$, where the variance of the assumed Gamma distributed infectious period is an input model parameter. The sets $\vec{\theta}_{0,\text{Exp}}$ and $\vec{\theta}_{0,\Gamma}$ are given by:

$$\vec{\theta}_{0,\text{Exp}} = (\beta, \tau)^T \in \mathbb{R}_+^{p_{\text{Exp}}} \text{ and } \vec{\theta}_{0,\Gamma} = (\beta, \tau, Var)^T \in \mathbb{R}_+^{p_\Gamma},$$

with $p_{\text{Exp}} = 2$ and $p_\Gamma = 3$, where $\mathbb{R}_+^{p_{\text{Exp}}}$ and $\mathbb{R}_+^{p_\Gamma}$ are known as the corresponding *sets of admissible parameter values*. To simplify the notation, denote $\vec{\theta}_0$ as the set of "true" parameter values of either $\vec{\theta}_{0,\text{Exp}}$ or $\vec{\theta}_{0,\Gamma}$. The *statistical model* for the random process Y_i and its realization y_i representing the incidence data are defined as:

$$Y_i = f(t_i, \vec{\theta}_0) + f(t_i, \vec{\theta}_0)^\xi \varepsilon_i \quad \text{and} \quad y_i = f(t_i, \vec{\theta}_0) + f(t_i, \vec{\theta}_0)^\xi \varepsilon_i \quad \text{for } i = 1, \ldots, n, \tag{5.1}$$

(see [64]) with $n = 31$. The mathematical model (or *regression function*) for the "observed" incidence cases at weeks/days t_i is

$$f(t_i, \vec{\theta}) = \begin{cases} C(t_1, \vec{\theta}) & \text{if } i = 1, \\ C(t_i, \vec{\theta}) - C(t_{i-1}, \vec{\theta}) & \text{if } 2 \le i \le n. \end{cases} \tag{5.2}$$

with C representing the cumulative incidence (see equation (3.5)). The observation errors ε_i are assumed to be independent random variables from the same unspecified probability density function with mean $E[\varepsilon_i] = 0$ and fixed variance $Var[\varepsilon_i] = \sigma_0^2$ (see [53, 64, 48, 54]). Equivalently, for the random process Y_i the mean is $E[Y_i] = f(t_i, \vec{\theta}_0)$ and the variance is $Var[Y_i] = f(t_i, \vec{\theta}_0)^{2\xi} \sigma_0^2$ (see [64]). The value of ξ determines the structure of the error

function of the statistical model in (5.1). This was chosen to be $\xi \equiv 0$ (ordinary least-squares), meaning that the variance of the random process Y_i remains constant through time ($Var[Y_i] = Var[\varepsilon_i] = \sigma_0^2$). This error structure is known as *absolute noise* (see [64]).

The random variable $\vec{\theta}_{LS}$, which is the *least squares estimator* for $\vec{\theta}_0$, and its realization $\hat{\vec{\theta}}_{LS}$ are obtained respectively by solving the minimization problem based on the cost functionals:

$$J_n(\vec{Y},\vec{\theta}) = \sum_{i=1}^{n} w_i |Y_i - \mathrm{f}(t_i,\vec{\theta})|^2 \qquad \text{and} \qquad J_n(\vec{y},\vec{\theta}) = \sum_{i=1}^{n} \hat{w}_i |y_i - \mathrm{f}(t_i,\vec{\theta})|^2, \tag{5.3}$$

$$\vec{\theta}_{LS} = \underset{\vec{\theta}\in\mathbb{R}^p_+}{\arg\min}\, J_n(\vec{Y},\vec{\theta}) \qquad \text{and} \qquad \hat{\vec{\theta}}_{LS} = \underset{\vec{\theta}\in\mathbb{R}^p_+}{\arg\min}\, J_n(\vec{y},\vec{\theta}) \tag{5.4}$$

where the weights w_i and $\hat{w}_i$ are given by

$$w_i = (\mathrm{f}(t_i,\vec{\theta}))^{-2\xi} \quad \text{and} \quad \hat{w}_i = (\mathrm{f}(t_i,\hat{\vec{\theta}}))^{-2\xi} \quad \text{for } i = 1,\ldots,n. \tag{5.5}$$

The estimates for $\hat{\vec{\theta}}_{\mathrm{Exp}}$ and $\hat{\vec{\theta}}_{\Gamma}$ for the sets $\vec{\theta}_{0,\mathrm{Exp}}$ and $\vec{\theta}_{0,\Gamma}$ are provided in **Tables 5.1, 5.4 and 5.5**.

An algorithm for the implementation of the least squares approach is described in the following pseudo-code:

Input Set the maximum number of iterations ($N_{iter} = 25$); set the tolerance ($TOL = 10^{-q}$), where $q(=5)$ is the resolution desired for convergence; set the number of iterations to zero ($j = 0$); set the initial guess values $\hat{\vec{\theta}}^{(0)}(\equiv \vec{1})$, where the superscript represents the number of iterations; and the weights are set to $\hat{w} \equiv \vec{1}$.

step 1 Do { step 2 to step 4 } While$((j \le N_{iter}) \& (||\hat{\vec{\theta}}^{(j-1)} - \hat{\vec{\theta}}^{(j)}||_2 \ge \mathrm{TOL}))$.

step 2 Compute the $j+1$ estimate $\hat{\vec{\theta}}^{(j+1)}$ for the estimator $\hat{\vec{\theta}}_{LS}$ by solving

$$\hat{\vec{\theta}}^{(j+1)} = \underset{\hat{\vec{\theta}}\in\mathbb{R}^p_+}{\arg\min}\, J_n(\vec{y},\hat{\vec{\theta}}^{(j)}). \tag{5.6}$$

step 3 Update the weights $\hat{w}_i = (\mathrm{f}(t_i, \hat{\vec{\theta}}^{(j+1)}))^{-2\xi}$ for $i = 1, \ldots, n$.

step 4 Increment the number of iteration by one $j = j + 1$.

Output Set the estimator $\hat{\vec{\theta}}_{LS} = \hat{\vec{\theta}}^{(j)}$.

The MATLAB 8.1.0 (R2013a, The MathWorks) built-in function *fminsearch* was used to solve equation (5.6). Alternative functions are *lsqnonlin* and *lsqcurvefit*.

5.2.5 Residual plots

The validity of the assumptions of the statistical model (5.1) is studied through the observation on the pattern from the *residuals plots*: If the pattern on the residuals (a realization of ε_i) over time is a random (scattered all over the domain without a clear or obvious trend), then it suggests that the errors (ε_i) are independent; Otherwise, if the pattern on the estimated model ($f(t, \hat{\vec{\theta}}_{LS})$) versus the residuals is a non-increasing random pattern, then it suggests that the assumption that the variance of the random process Y_i remains constant through time ($Var[Y_i] = \sigma_0^2$) is plausible (see for example **Figures 5.3 and 5.4**).

5.2.6 Confidence intervals for $\vec{\theta}_{0,Exp}$ and $\vec{\theta}_{0,\Gamma}$

This subsection is devoted to describe how the confidence intervals of the estimated sets of parameters $\vec{\theta}_{0,\mathrm{Exp}}$ and $\vec{\theta}_{0,\Gamma}$ can be obtained. To simplify the notation, let $\vec{\theta}_0$ be either $\vec{\theta}_{0,\mathrm{Exp}}$ or $\vec{\theta}_{0,\Gamma}$ and p be either p_{Exp} or p_Γ. Given that the regularity and sampling conditions are satisfied, then according to asymptotic theory (as $n \to \infty$) the least-squares estimator $\vec{\theta}_{LS}^n$ follows a p-multivariable normal distribution with mean $E[\vec{\theta}_{LS}^n] \approx \vec{\theta}_0$ and variance-covariance matrix $Var[\vec{\theta}_{LS}^n] \approx \sum_0^n$:

$$\vec{\theta}_{LS} = \vec{\theta}_{LS}^n \sim \mathscr{N}_p(\vec{\theta}_0, \sum_0^n) \approx \mathscr{N}_p(\vec{\theta}_0, \sigma_0^2[\chi^{nT}(\vec{\theta}_0)\chi^n(\vec{\theta}_0)]^{-1}),$$

where σ_0^2 is the variance of the errors ε_i (for $i = 1, \ldots, n = 31$) and $\chi^n(\vec{\theta}_0)$ is the sensitivity matrix of the mathematical model in (5.2) (see [53, 48, 54] and the references therein).

The sensitivity matrix is defined as $\chi(\vec{\theta}) = \chi^n(\vec{\theta}) = \{\chi^n_{ij}\}$, where $\chi^n_{ij}(\vec{\theta}) = \frac{\partial f(t_i,\vec{\theta})}{\partial \theta_j}$ for $i = 1,\ldots,n$ and $j = 1,\ldots,p$ are known as the sensitivity equations of $f(t_i,\vec{\theta})$ with respect to $\vec{\theta}$ [55] (see **appendix H** for a detailed derivation).

The variance σ_0^2 is approximated by computing the bias adjusted estimate

$$\sigma_0^2 \approx \hat{\sigma}^2(\hat{\vec{\theta}}_{LS}) = \frac{1}{n-p} J_n(\vec{y}, \hat{\vec{\theta}}_{LS}), \tag{5.7}$$

and an estimate of the variance-covariance matrix $\sum_0^n$ is given by

$$\hat{\sum}^n(\hat{\vec{\theta}}_{LS}) = \left[\frac{1}{\hat{\sigma}^2(\hat{\vec{\theta}}_{LS})} \chi^T(\hat{\vec{\theta}}_{LS})\chi(\hat{\vec{\theta}}_{LS})\right]^{-1}, \tag{5.8}$$

where the standard error for each $\hat{\theta}_{LS_j}$ is given by $SE_j(\hat{\vec{\theta}}_{LS}) = \sqrt{\hat{\sum}^n_{jj}(\hat{\vec{\theta}}_{LS})}$ (see [53, 48, 54, 56]).

Finally, an expression for the confidence interval for each θ_{0_j} at a level of significance α is provided by the following expression:

$$P\{\hat{\theta}_{LS_j} - t_{1-\frac{\alpha}{2}} SE_j(\hat{\vec{\theta}}_{LS}) < \theta_{0_j} < \hat{\theta}_{LS_j} + t_{1-\frac{\alpha}{2}} SE_j(\hat{\vec{\theta}}_{LS})\} = 1-\alpha, \tag{5.9}$$

where $t_{1-\frac{\alpha}{2}}$ is Student-t distribution statistic with $n-p$ degrees of freedom [53]. The 95^{th} percentile confidence intervals for the sets $\vec{\theta}_{0,\text{Exp}}$ are given in **Table 5.1**.

5.2.7 Estimation of the epidemiological quantities of interest and their confidence intervals

The epidemiological quantities of interest such as: The peak size of the prevalence of infectious individuals; Time at which the peak of the prevalence occurs; Final epidemic size; Epidemic duration and Basic reproduction number (see **Figure 1.1**) are estimated as described in the methods section of chapter three but using the estimates $\hat{\vec{\theta}}_{\text{Exp}}$ and $\hat{\vec{\theta}}_{\Gamma}$.

The variance of the epidemiological quantities mentioned above can be estimated in two ways. The propagation of error method or "δ"-method is used for those epidemiological

quantities that has explicit analytical functional description with the model parameters, such as the basic reproduction number and the final epidemic size (see equations (3.1) and (3.5)). While for the rest of the epidemiological quantities an indirect method such as *bootstrapping* can be applied.

Propagation of error method or "δ"-method

Let f and g be the analytical description for the basic reproduction number and the final epidemic size, respectively given by $f(\vec{q}_0) = \mathscr{R}_0$ and $g(\vec{q}_0) = z$ (see equations (3.1) and (3.5), respectively), with $\vec{q}_0 = (\beta_0, \tau)^T$ being the vector of "true" parameter values with corresponding variance-covariance matrix $\sum_0$. Then an approximation of the variance $Var(\hat{\mathscr{R}}_0)$ and $Var(\hat{z})$ is obtained via the equations:

$$Var(\hat{\mathscr{R}}_0) \approx \nabla f(\hat{\vec{q}})^T \hat{\sum}^n \nabla f(\hat{\vec{q}}) \quad \text{and} \quad Var(\hat{z}) \approx \nabla g(\hat{\vec{q}})^T \hat{\sum} \nabla g(\hat{\vec{q}}) \tag{5.10}$$

with standard error $SE(\hat{\mathscr{R}}_0) \approx \sqrt{Var(\hat{\mathscr{R}}_0)}$ and $SE(\hat{z}) \approx \sqrt{Var(\hat{z})}$ (see [59] for details on the "δ"-method).

Let x_0 be the "true" value for either of the five epidemiological quantities of interest mentioned above with mean $\hat{x}$ and standard error $SE(\hat{x})$, then the confidence intervals for x_0 at a level of significance α is obtained by

$$P\{\hat{x} - t_{1-\frac{\alpha}{2}} SE(\hat{x}) < x_0 < \hat{x} + t_{1-\frac{\alpha}{2}} SE(\hat{x})\} = 1 - \alpha. \tag{5.11}$$

The estimates, standard errors and 95^{th} percentile confidence intervals for the five epidemiological quantities of interest mentioned above are given in **Tables 5.2 and 5.3**.

5.2.8 Residual sum of squares (RSS) based test or ANOVA test

An ANOVA (statistical) test is used to assess the validity of the common modeling assumption of exponential distributed infectious period. This is executed indirectly by testing if the value for the variance of the assumed Gamma distributed infectious period equals the corresponding value for the exponential distribution, $Var = \tau^2$.

Let Q be the set of admissible parameters for a vector of parameters $\vec{\theta}_\Gamma = (\beta, \tau, Var)^T$, then $Q = \mathbb{R}_+^{p_\Gamma}$, with $p_\Gamma = 3$. Let Q_H be a subset of the set Q, with description

$$Q_H = \{\vec{\theta}_\Gamma \in Q : H\vec{\theta}_\Gamma = c\},$$

where H is a $r \times p_\Gamma$ matrix with r = 1 defined as $H = (0,0,1)$, and where c is a constant that takes the value for the variance of the exponential distribution for the infectious period, c = τ^2. The *null hypothesis* H_0 for the statistical test is then:

$$H_0 : \vec{\theta}_{0,\Gamma} \in Q_H \quad \text{or equivalently} \quad H_0 : Var = \tau^2.$$

A logically equivalent alternative interpretation for the null hypothesis H_0 is that the SIR model should be described by a nonlinear system of ordinary differential equations. While the *alternative hypothesis* H_a, given by:

$$H_a : \vec{\theta}_{0,\Gamma} \notin Q_H \quad \text{or equivalently} \quad H_a : Var \neq \tau^2$$

is interpreted as the SIR model being described by a nonlinear system of Volterra integral equations, with the exception of having exponentially distributed infectious period.

The estimator $\vec{\theta}_{LS}$ for $\vec{\theta}_{0,\Gamma}$ and its realization $\hat{\vec{\theta}}_{LS}$ are defined in (5.4) and the estimator $\vec{\theta}_H$ for $\vec{\theta}_{0,\Gamma}$ and its realization $\hat{\vec{\theta}}_H$ are given by:

$$\vec{\theta}_H = \underset{\vec{\theta} \in Q_H}{\arg\min} J_n(\vec{Y}, \vec{\theta}) \qquad \text{and} \qquad \hat{\vec{\theta}}_H = \underset{\vec{\theta} \in Q_H}{\arg\min} J_n(\vec{y}, \vec{\theta}). \tag{5.12}$$

A test statistic T_n and its realization $\hat{T}_n$ are respectively defined as:

$$T_n(\vec{Y}) = n(J_n(\vec{Y},\vec{\theta}_H) - J_n(\vec{Y},\vec{\theta}_{LS})), \tag{5.13}$$

$$\hat{T}_n(\vec{y}) = n(J_n(\vec{y},\hat{\vec{\theta}}_H) - J_n(\vec{y},\hat{\vec{\theta}}_{LS})). \tag{5.14}$$

The above test statistics is non-negative, since $J_n(\vec{y},\hat{\vec{\theta}}_H) \geq J_n(\vec{y},\hat{\vec{\theta}}_{LS})$.

An additional test statistic U_n is defined as a function of the test statistics T_n as:

$$U_n(\vec{Y}) = \frac{T_n(\vec{Y})}{J_n(\vec{Y},\vec{\theta}_{LS})} \quad \text{with realization} \quad \hat{U}_n(\vec{y}) = \frac{\hat{T}_n(\vec{y})}{J_n(\vec{y},\hat{\vec{\theta}}_{LS})}. \tag{5.15}$$

The veracity of the model comparison statistical (ANOVA) test depends on two plausible assumptions, under regularity and the way in which the sample or data is collected (see [53] and references their for more details):

- The estimator $\vec{\theta}_{LS}$ converges to $\vec{\theta}_{0,\Gamma}$ with probability one as $n \to \infty$;
- If the null hypothesis H_0 is true, then U_n converges in the distributional sense to $U \sim \chi^2(r)$ as $n \to \infty$ where r represent the degrees of freedom in the χ^2 distribution and is determined by number of constrains imposed to the vector of parameter $\vec{\theta}_\Gamma$ or the number of rows in the H matrix.

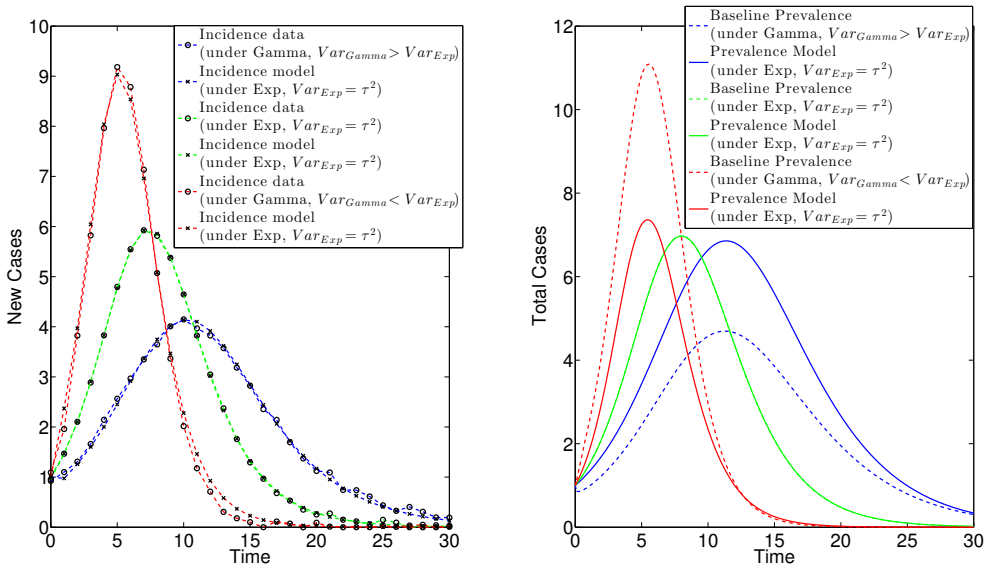
The statistical test is stated is as follow: If the probability-value $P\{U \sim \chi^2(r) \geq \hat{U}_n\}$ is strictly less than the α level of significance, then there is sufficient evidence to reject the null hypothesis H_0 and thus accept the alternative hypothesis H_a, meaning that the exponential probability density function is not a suitable option for the infectious period distribution which is equivalent to say that the SIR model should not be described by a nonlinear system of ordinary differential equations; otherwise rejection of the the null hypothesis H_0 fails, concluding that there is not sufficient evidence that suggest that a the SIR model should be described by a nonlinear system of Volterra integral equations.

5.3 Results

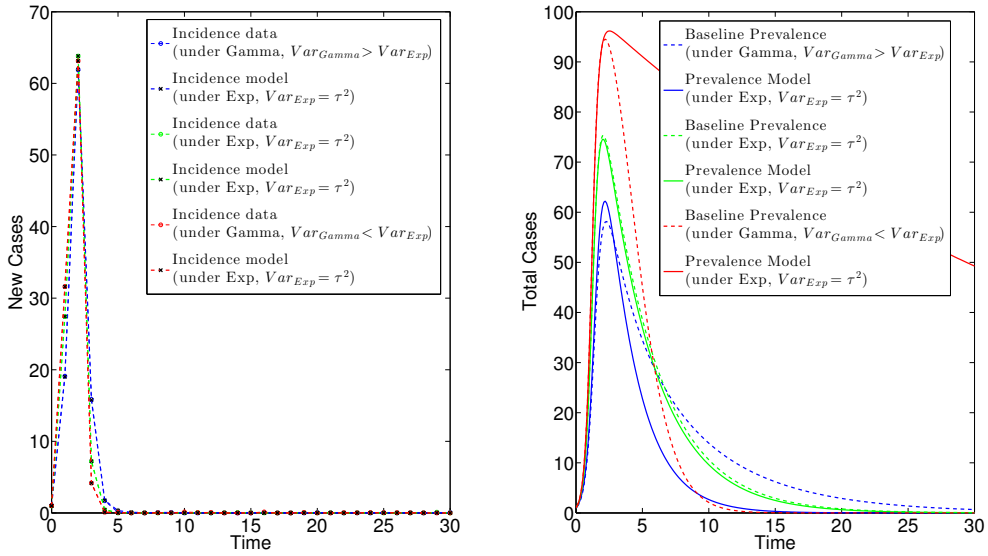
5.3.1 Results using the mathematical model f assuming Exponentially distributed infectious period

The mathematical model f used to describe the observable incidence cases, under the assumption that the infectious period is exponentially distributed, provided a good fit to the data for the scenarios where the transmissibility is either low or high and for all the cases where the "true" (or baseline) infectious period distribution is either exponential or non-exponential (Gamma) (see **Figure 5.2**, left panels). The parameter estimates for the set $\hat{\vec{\theta}}_{\text{Exp}} = (\hat{\beta}, \hat{\tau})^T$ are in agreement with the "true" values ($\vec{\theta}_{0,\text{Exp}}$), on both scenarios, only for the case where the "true" (or baseline) infectious period distribution is exponential (see **Table 5.1**). For high transmissibility, the estimate of mean infectious period τ is a sensitive parameter for the cases where the "true" (or baseline) infectious period distribution is non-exponential (Gamma) (see **Table 5.1**). In both scenarios, the predicted prevalence of infected individuals does not capture the transient trend of the "true" prevalence whenever the baseline infectious period distribution is non-exponential (Gamma) (see **Figure 5.2**, right panels), with an exception for high transmissibility, only at the very early stage of the epidemic (see **Figure 5.2**, bottom-right panel). When the baseline infectious period distribution is Exponential, as expected, the predicted prevalence is in agreement with the "true" transient trend of the prevalence (see **Figure 5.2**, right panels). For the low transmissibility scenario and all cases the prediction of the outcome variables such as: the prevalence peak time, final epidemic size, epidemic duration and basic reproduction number are in close agreement with the "true" values (see **Table 5.2**). For the high transmissibility scenario and all cases the prediction of the outcome variables such as: the prevalence peak time and final epidemic size are in close agreement with the "true" values (see **Table 5.3**). For high transmissibility, predicted values for the epidemic duration and basic reproduction number

tend to be underestimated whenever the baseline variance of the Gamma distributed infectious period is greater than the variance of the Exponential distribution, on the contrary, these values tend to be overestimated whenever the baseline variance of the Gamma distributed infectious period is less than the variance of the Exponential distribution (see **Table 5.3**). The opposite occur to the predicted values for the prevalence peak size, but for both scenarios, these tend to be overestimated whenever the baseline variance of the Gamma distributed infectious period is greater than the variance of the Exponential distribution, or on the hand, overestimated whenever the baseline variance of the Gamma distributed infectious period is less than the variance of the Exponential distribution (see **Tables 5.2 and 5.3**). In all scenarios and cases considered, the random pattern of the residuals over time (see **Figures 5.3 and 5.4**, left panels) provide strong evidence that validates the statistical modeling assumption that the errors (ε) are independent. Also, the non-increasing random pattern of the residuals ε versus the estimated incidence model f suggests that the assumption that the variance of the random process Y_i remains constant through time ($Var[Y_i] = \sigma_0^2$) (see **Figures 5.3 and 5.4**, right panels).



(a) Low transmissibility, $\mathscr{R}_0 = 1.5$.



(b) High transmissibility, $\mathscr{R}_0 = 15$.

Figure 5.2: Incidence data (left panels, in dotted-dash lines) and prevalence baselines (left panels, in dash line) of infectious individuals for low (top panels) and high (bottom panels) transmissibility under different infectious period distributions. The predictions from the SIR model under Exponentially distributed infectious period are in solid lines.

Table 5.1: Estimates and 95% confidence intervals for $\vec{\theta}_{0,\text{Exp}}$. The low and high transmissibility cases, $\mathscr{R}_0 = 1.5$ and $\mathscr{R}_0 = 15$.

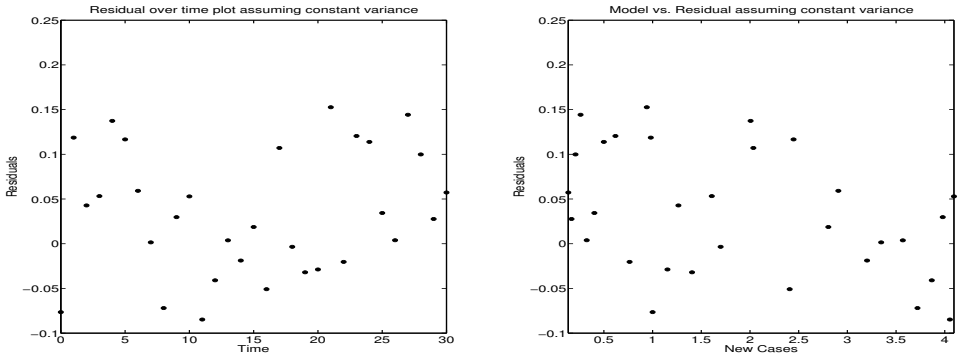
Case	Parameter	Infectious Period Distribution assumed for the generated data	True Value	Estimate (95% CI)
Low $\mathscr{R}_0(=1.5)$	β	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	$\sqrt{1.5}$	$0.86(95\%\text{ CI: } 0.85, 0.87)$
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	$\sqrt{1.5}$	$1.223(95\%\text{ CI: } 1.219, 1.227)$
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	$\sqrt{1.5}$	$1.79(95\%\text{ CI: } 1.77, 1.82)$
	τ	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	$\sqrt{1.5}$	$1.73(95\%\text{ CI: } 1.70, 1.75)$
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	$\sqrt{1.5}$	$1.227(95\%\text{ CI: } 1.220, 1.233)$
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	$\sqrt{1.5}$	$0.85(95\%\text{ CI: } 0.83, 0.86)$
High $\mathscr{R}_0(=15)$	β	**Gamma**$(k=0.5, \theta=\frac{\sqrt{15}}{0.5}), Var=30$	$\sqrt{15}$	$3.53(95\%\text{ CI: } 3.52, 3.54)$
		Gamma$(k=1, \theta=\sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var=15$	$\sqrt{15}$	$3.881(95\%\text{ CI: } 3.875, 3.886)$
		Gamma$(k=4, \theta=\frac{\sqrt{15}}{4}), Var=3.75$	$\sqrt{15}$	$3.89(95\%\text{ CI: } 3.88, 3.90)$
	τ	**Gamma**$(k=0.5, \theta=\frac{\sqrt{15}}{0.5}), Var=30$	$\sqrt{15}$	$2.31(95\%\text{ CI: } 2.25, 2.37)$
		Gamma$(k=1, \theta=\sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var=15$	$\sqrt{15}$	$3.74(95\%\text{ CI: } 3.64, 3.83)$
		Gamma$(k=4, \theta=\frac{\sqrt{15}}{4}), Var=3.75$	$\sqrt{15}$	$41.24(95\%\text{ CI: } 19.86, 62.63)$

Table 5.2: Estimates and 95% confidence intervals for the predicted epidemiological variables from the SIR model under the Exponentially distributed infectious period assumption. The low transmissibility case, $\mathscr{R}_0 = 1.5$.

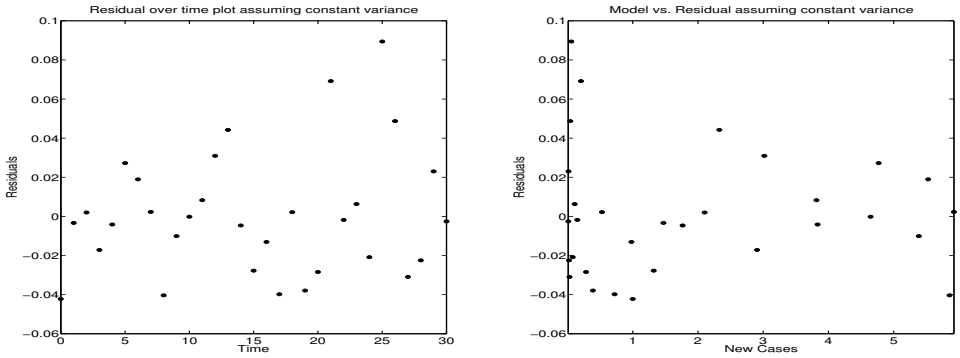
Case	Outcome variable	Infectious Period Distribution assumed for the generated data	True Value	Estimate (95% CI)
Low $\mathscr{R}_0(=1.5)$	Prevalence peak size (%)	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	4.70	6.82(95% CI: 6.33, 7.31)
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	6.9722	6.9718(95% CI: 6.78, 7.17)
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	11.0866	7.3424(95% CI: 6.5943, 8.0881)
	Prevalence peak time	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	11.3	11.3(95% CI: 10.71, 11.96)
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	8	8(95% CI: NaN, NaN)
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	5.5	5.4(95% CI: 5.19, 5.67)
	Final epidemic size (%)	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	58.94	58.28(95% CI: 43.35, 73.20)
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	59.323	59.320(95% CI: 53.50, 65.13)
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	59.13	60.67(95% CI: 39.73, 81.59)
	Epidemic duration	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	23.8	25(95% CI: 21.98, 27.93)
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	17.6	17.6(95% CI: 16.92, 18.32)
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	12.2	12(95% CI: 10.996, 13.004)
	Basic reproduction number	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	1.5	1.491(95% CI: 1.485, 1.497)
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	1.5	1.4998(95% CI: 1.497, 1.502)
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	1.5	1.52(95% CI: 1.512, 1.530)

Table 5.3: Estimates and 95% confidence intervals for the predicted epidemiological variables from the SIR model under the Exponentially distributed infectious period assumption. The high transmissibility case, $\mathscr{R}_0 = 15$.

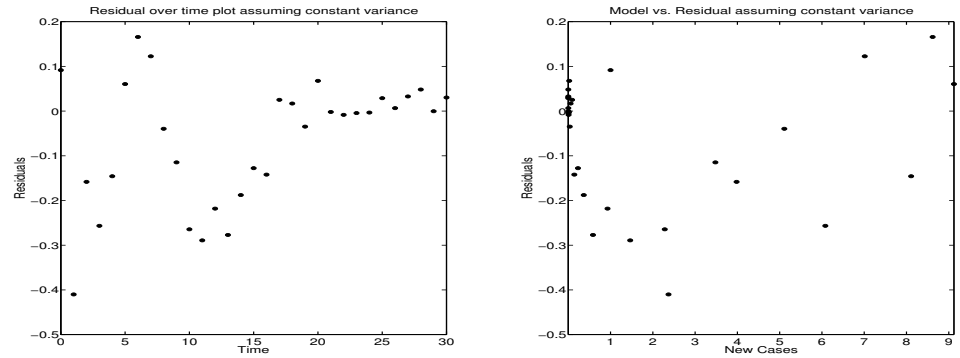
Case	Outcome variable	Infectious Period Distribution assumed for the generated data	True Value	Estimate (95% CI)
High $\mathscr{R}_0 (= 15)$	Prevalence peak size (%)	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	58.15	62.12(95% CI: 44.79, 79.45)
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	75.28	74.70(95% CI: 59.51, 89.89)
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	94.53	96.22(95% CI: 27.29, 164.82)
	Prevalence peak time	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	2.3	2.2(95% CI: 2.2, 2.2)
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	2	2(95% CI: NaN, NaN)
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	2.2	2.5(95% CI: 2.36, 2.67)
	Final epidemic size (%)	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	100.22	99.97(95% CI: 99.71, 100.23)
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	100.2924	99.9999(95% CI: 99.9991, 100.0008)
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	100.89	100(95% CI: 100, 100)
	Epidemic duration	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	27.3	12.3(95% CI: 10.72, 13.86)
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	19.2	18.6(95% CI: 14.91, 22.25)
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	11	30(95% CI: NaN, NaN)
	Basic reproduction number	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	15	8.15(95% CI: 7.46, 8.82)
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	15	14.51(95% CI: 12.13, 16.87)
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	15	160.36(95% CI: $-3114.52, 3435.14$)



(a) The case of low transmissibility, $\mathscr{R}_0 = 1.5$ and Gamma distributed infectious period as: Gamma$(k = 0.5, \theta = \frac{\sqrt{1.5}}{0.5}), Var = 3.0$.

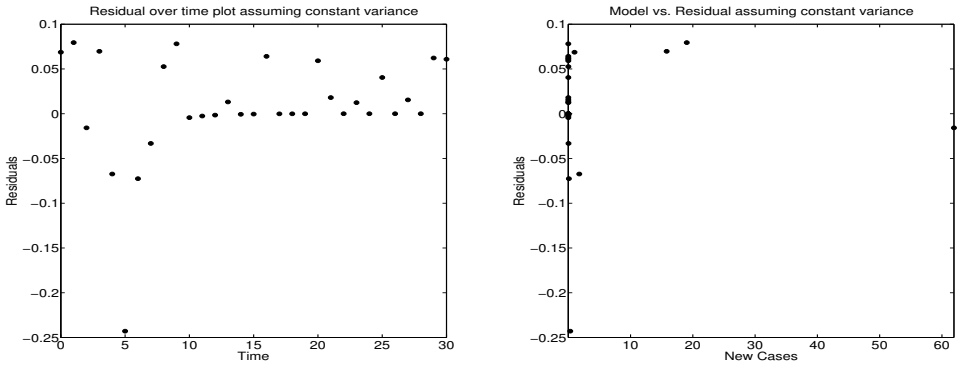


(b) The case of low transmissibility, $\mathscr{R}_0 = 1.5$ and Exponentially distributed infectious period as: Exp$(\frac{1}{\sqrt{1.5}}), Var = 1.5$.

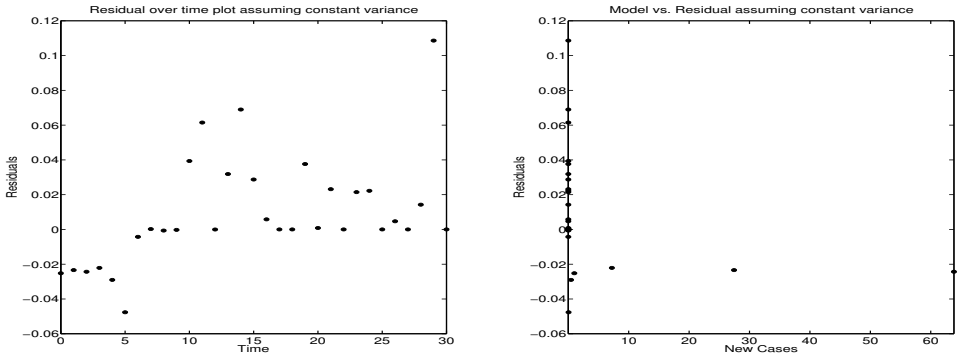


(c) The case of low transmissibility, $\mathscr{R}_0 = 1.5$ and Gamma distributed infectious period as: Gamma$(k = 4, \theta = \frac{\sqrt{1.5}}{4}), Var = 0.375$.

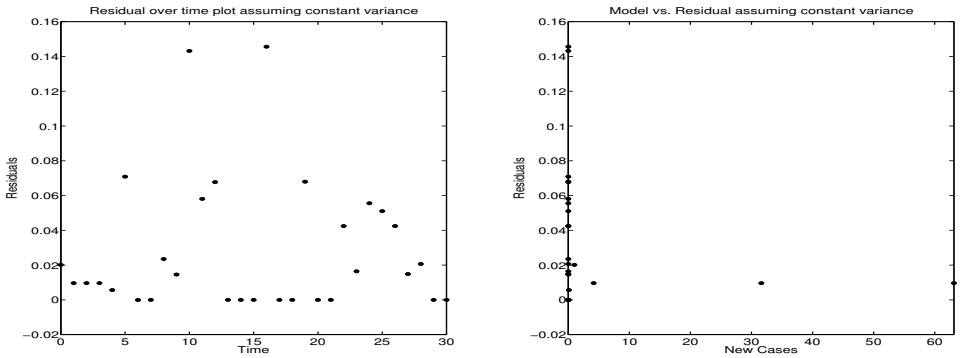
Figure 5.3: On the left graphs, the residuals (ε) over time and on the right graphs, the residuals (ε) versus the estimated incidence model (f) under the Exponentially distributed infectious period assumption.



(a) The case of high transmissibility, $\mathscr{R}_0 = 15$ and Gamma distributed infectious period as: Gamma$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$.



(b) The case of high transmissibility, $\mathscr{R}_0 = 15$ and Exponentially distributed infectious period as: Exp$(\frac{1}{\sqrt{15}}), Var = 15$.



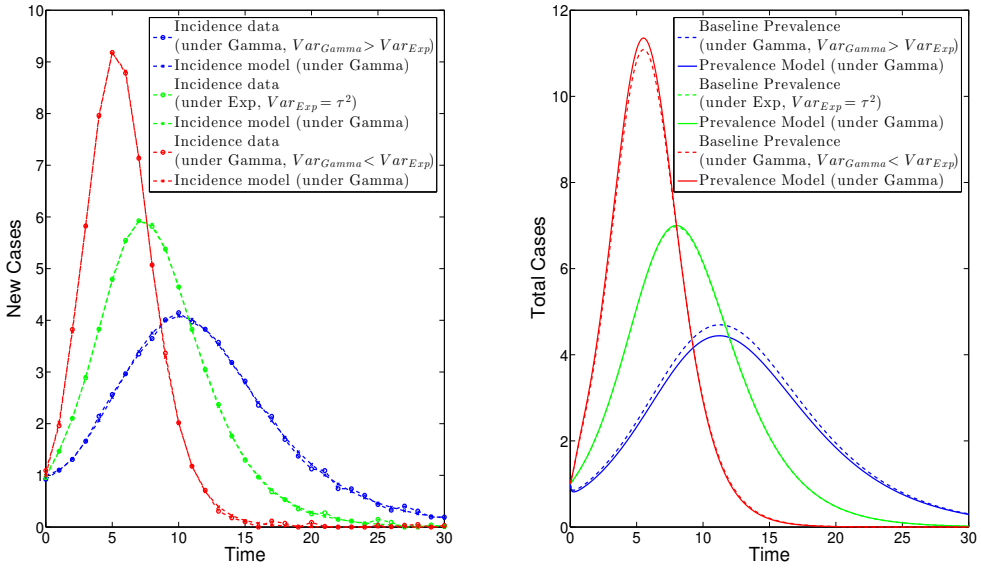
(c) The case of high transmissibility, $\mathscr{R}_0 = 15$ and Gamma distributed infectious period as: Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$.

Figure 5.4: On the left graphs, the residuals (ε) over time and on the right graphs, the residuals (ε) versus the estimated incidence model (f) under the Exponentially distributed infectious period assumption.

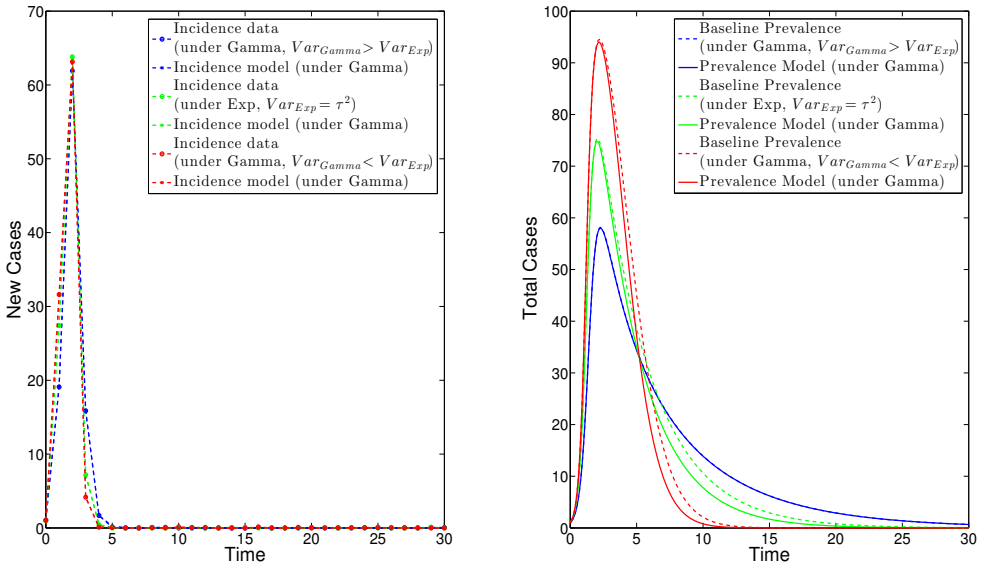
5.3.2 Results using the mathematical model f assuming Gamma distributed infectious period

As expected, the fit to all the incidence data sets considered (see **Figure 5.5**, left panels) and the predicted prevalence (see **Figure 5.5**, right panels) from using the mathematical model f for the incidence under the Gamma distributed infectious period assumption are notably better than in the case previously discussed in the subsection 5.3.1. For low transmissibility, all cases, the parameter estimates for the set $\hat{\vec{\theta}}_{\Gamma} = (\hat{\beta}, \hat{\tau}, \hat{Var})^T$ are in agreement with the "true" values ($\vec{\theta}_{0,\Gamma}$) (see **Table 5.4**). For high transmissibility the parameter estimates for the set $\hat{\vec{\theta}}_{\Gamma} = (\hat{\beta}, \hat{\tau}, \hat{Var})^T$ are in agreement with the "true" values ($\vec{\theta}_{0,\Gamma}$), only when the "true" variance of the Gamma distributed infectious period is greater than the variance of the Exponential distribution (see **Table 5.5**). For high transmissibility, the estimate of the mean τ and variance Var of the infectious period distribution are sensitive parameters for the cases where the "true" variance of the Gamma distributed infectious period is less or equal than the variance of the Exponential distribution (see **Table 5.5**). Nevertheless, in all scenarios and cases, still the predicted infectious period distribution is identifiable and in agreement with the "true" probability density function for the infectious period (see **Figure 5.6**). In both scenarios, all cases, the predicted prevalence of infected individuals capture nicely the transient trend of the "true" prevalence (see **Figure 5.5**, right panels). For all scenarios and cases the prediction of the outcome variables are in close agreement with the "true" values (see **Figures 5.5 and 5.6**). The conclusions of the statistical test are expected, since when the "true" infectious period distribution is assumed to be non-Exponential, then the test suggest the rejection of the null hypothesis H_0 and the acceptance of the H_a, which state that the infectious period is Gamma, but non-exponentially distributed and when the "true" infectious period distribution is assumed to be Exponential, then it fails to reject the null hypothesis H_0 (see **Table 5.6**). There is only one exceptional case where the test

did not provided the correct suggestion and it is for high transmissibility when the "true" variance of the Gamma distributed infectious period is less than the variance of the Exponential distribution. This might be due to numerical roundoff error or inaccuracies in the numerical solver used, since there is a lot of almost zero values in the incidence when the transmissibility is high.



(a) Low transmissibility, $\mathscr{R}_0 = 1.5$.



(b) High transmissibility, $\mathscr{R}_0 = 15$.

Figure 5.5: Incidence data (left panels, in dotted-dash lines) and prevalence baselines (left panels, in dash line) of infectious individuals for low (top panels) and high (bottom panels) transmissibility under different infectious period distributions. The predictions from the SIR model under Gamma distributed infectious period are in solid lines.

Table 5.4: Estimates for $\vec{\theta}_{0,\Gamma}$. The low transmissibility case, $\mathscr{R}_0 = 1.5$.

Case	Parameter	Infectious Period Distribution assumed for the generated data	True Value	Estimate
Low $\mathscr{R}_0(=1.5)$	β	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	$\sqrt{1.5}$	1.303
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	$\sqrt{1.5}$	1.217
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	$\sqrt{1.5}$	1.190
	τ	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	$\sqrt{1.5}$	1.155
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	$\sqrt{1.5}$	1.232
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	$\sqrt{1.5}$	1.254
	Var	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	3.0	2.950
		Gamma$(k=1, \theta=\sqrt{1.5}) =$ **Exp**$(\frac{1}{\sqrt{1.5}}), Var=1.5$	1.5	1.499
		Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	0.375	0.306

Table 5.5: Estimates for $\vec{\theta}_{0,\Gamma}$. The high transmissibility case, $\mathscr{R}_0 = 15$.

Case	Parameter	Infectious Period Distribution assumed for the generated data	True Value	Estimate
High $\mathscr{R}_0 (= 15)$	β	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	$\sqrt{15}$	3.883
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	$\sqrt{15}$	3.866
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	$\sqrt{15}$	3.871
	τ	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	$\sqrt{15}$	3.864
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	$\sqrt{15}$	3.473
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	$\sqrt{15}$	3.483
	Var	**Gamma**$(k = 0.5, \theta = \frac{\sqrt{15}}{0.5}), Var = 30$	30	30.033
		Gamma$(k = 1, \theta = \sqrt{15}) =$ **Exp**$(\frac{1}{\sqrt{15}}), Var = 15$	15	11.216
		Gamma$(k = 4, \theta = \frac{\sqrt{15}}{4}), Var = 3.75$	3.75	2.773

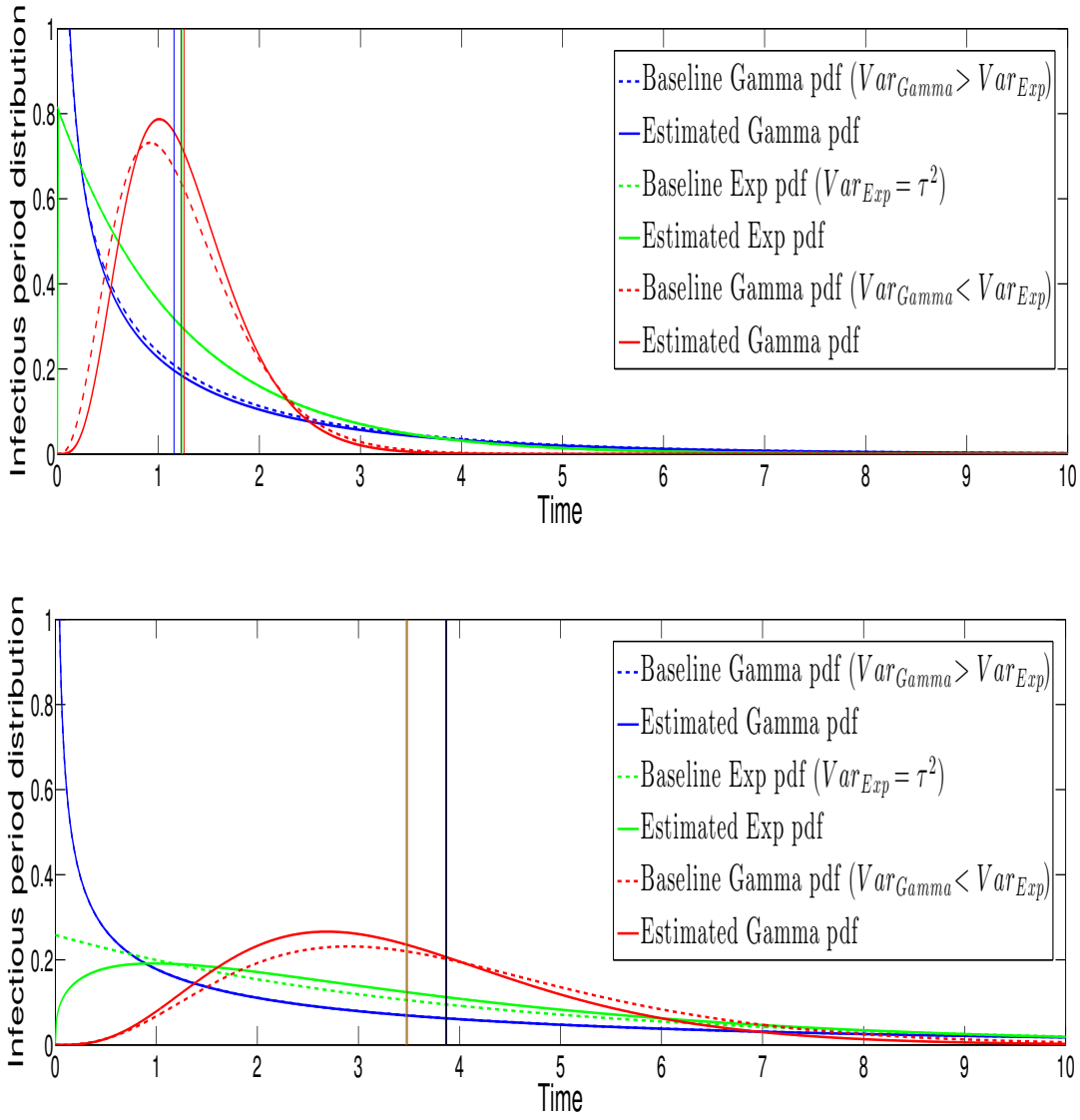


Figure 5.6: Baseline (in dash lines) and estimated (in solid lines) probability density function for the infectious period for low (top panel) and high (bottom panel) transmissibility.

Table 5.6: Objective functional values for $J_n(\vec{y}, \hat{\vec{\theta}}_{LS})$ and $J_n(\vec{y}, \hat{\vec{\theta}}_{H})$, Test statistics values $\hat{U}_n(\vec{y})$, χ^2-statistics values $inv-\chi^2(1-0.05; r=1)$, probability values and conclusion of the statistical test

Case	Infectious Period Distribution assumed for the generated data	Objective functionals		Statistics		P-value	Conclusion
		$J_n(\vec{y}, \hat{\vec{\theta}}_{LS})$	$J_n(\vec{y}, \hat{\vec{\theta}}_{H})$	$\hat{U}_n(\vec{y})$	$inv-\chi^2(0.95;1)$		
Low $\mathscr{R}_0$	**Gamma**$(k=0.5, \theta=\frac{\sqrt{1.5}}{0.5}), Var=3.0$	0.10	0.18	24.21	3.84	8.6×10^{-7}	H_0 is rejected
	Gamma$(k=1, \theta=\sqrt{1.5}), Var=1.5$	0.0296	0.0304	0.79	3.84	0.37	Fails to reject H_0
	Gamma$(k=4, \theta=\frac{\sqrt{1.5}}{4}), Var=0.375$	0.06	0.71	340.20	3.84	0	H_0 is rejected
High $\mathscr{R}_0$	**Gamma**$(k=0.5, \theta=\frac{\sqrt{15}}{0.5}), Var=30$	0.07	0.11	19.05	3.84	1.3×10^{-5}	H_0 is rejected
	Gamma$(k=1, \theta=\sqrt{15}), Var=15$	0.0325	0.0323	-0.174	3.84	1	Fails to reject H_0
	Gamma$(k=4, \theta=\frac{\sqrt{15}}{4}), Var=3.75$	0.0712	0.0710	-0.0732	3.84	1	Fails to reject H_0

5.4 Conclusions

The epidemiological quantities whose prediction are not affected by the standard Exponentially distributed infectious period modeling assumption are: for the low transmissibility scenario the prevalence peak time, final epidemic size, epidemic duration and basic reproduction number and for the high transmissibility scenario the prevalence peak time and final epidemic size. This conclusion is particular to the SIR model considered and regardless if the estimates of the input model parameters are either close or not too far off from the "true" parameters values. However, in comparison with the SIR under Exponentially distributed infectious period, the SIR model governed by a nonlinear system of Volterra integral equations will produce a more accurate fit of the data, estimates for the input parameters and therefore more accurate predictions for the outcome variable. For this reason and to avoid problems of either significant under or overestimation on prediction, we suggest the used of the SIR model governed by a nonlinear system of Volterra integral equations for outcome variables such as the epidemic duration and basic reproduction number for high transmissibility and the prevalence peak size for both scenarios. Lastly, an advantage of the latter model over the standard SIR model is that the probability density function for the infectious period of a specific infectious disease can be identifiable from epidemiological data, this, indirectly by estimating the distribution parameters, which are input parameters in the SIR model governed by a nonlinear system of Volterra integral equations. This last conclusion is particular to the SIR model considered and regardless if the estimates of the input model parameters are either close or not too far off from the "true" parameters values.

5.5 Discussion

The main contributions on this chapter are:

Method one: The development of a new application which test statistically whether or not the infectious period distribution is non-Exponentially distributed.

Method two: In addition, a method for estimating the probability density function for the infectious period of a particular disease from epidemiological data is provided, by considering the parameters of a general infectious period distribution (Gamma in our case) as input model parameters in the SIR model governed by a nonlinear system of Volterra integral equations.

Further work can be done on assessing the robustness and power of **Method one** by applying it to generated incidence data from agent based-like stochastic models. **Method two** can be expanded for generating a uncertainty bound where the "True" probability density function for the infectious period will be contained.

Chapter 6

DISCUSSION AND CONCLUSIONS

6.1 Summary of main conclusions: What we learnt?

From the first core project, chapter three, we learnt that, for relatively low ($\mathscr{R}_0$ close to one) to excessively high (mean of $\mathscr{R}_0$ equals 15) transmissibility, the variance of the Gamma distribution for the infectious period, input parameter of the deterministic age-of-infection SIR model, is key (statistically significant) for the predictability of epidemiological variables such as the epidemic duration and the peak size and timing of the prevalence of infectious individuals. Hence, it is preferable to utilize a nonlinear system of Volterra integral equations, rather than a nonlinear system of ordinary differential equations if the goal is to have better predictions or forecasting. On the other hand, the predictability of epidemiological variables such as the final epidemic size and the basic reproduction number are unaffected by (or independent of) the variance of the Gamma distribution (for the infectious period) and therefore independent on the choice of the type of nonlinear system used for the description of the SIR model (VIE's or ODE's). Although, practical proposes (with the aim of lowering the complexity and number operations in the numerical methods) supports the use of a nonlinear system of ordinary differential equations.

From the second core project, Chapter four, we learned that, for relatively low transmissibility (around $\mathscr{R}_0 = 1.5$), the variance of the Gamma distribution for the infectious period, input parameter of the stochastic age-of-infection SIR model, is key (statistically significant) for the estimation of the probability of a particular event; as defined from the empirical distribution of random epidemiological variables such as the final epidemic size, epidemic duration, prevalence peak size and prevalence peak time. For the case of high

transmissibility (around $\mathscr{R}_0 = 15$), the variance of the Gamma distribution for the infectious period is a key (statistically significant) parameter on the estimation of the probability of a particular event, defined from the empirical distribution of the epidemic duration. By relating the main conclusions from Chapters three and four, it can be concluded that, for relatively low transmissibility (around $\mathscr{R}_0 = 1.5$) and eventhough, in the deterministic sense, the variance of the Gamma distribution for the infectious period does not play any role in the predictability of the final epidemic size, the fact is, that in the stochastic (distributinal) sense, it does. However, for high transmissibility (around $\mathscr{R}_0 = 15$), even though, in the deterministic sense, the variance of the Gamma distribution for the infectious period is a key input parameter in the predictability of variables such as the peak size and timing of the prevalence of infectious individuals, the fact is, that in the stochastic (distributinal) sense, it does not. Generally speaking, these discrepancies are justifiable, attributed or induced by the stochasticity, which was introduced on the time at which infection and recovery events occur.

From the third core project, Chapter five, we learned that, the epidemiological quantities unaffected (in terms of prediction) by the standard Exponentially distributed infectious period modeling assumption are: for the low transmissibility scenario the prevalence peak time, final epidemic size, epidemic duration and basic reproduction number and for the high transmissibility scenario the prevalence peak time and final epidemic size. However, when compared with the SIR under Exponentially distributed infectious period, it turns out that the SIR model governed by a nonlinear system of Volterra integral equations, actually produces more accurate fit to the data, estimates for the input parameters and therefore more accurate predictions for the outcome variables. For this reason and to avoid problems of either significant under or overestimation on prediction, we suggest the used of the SIR models governed by a nonlinear system of Volterra integral equations, in particular, when we are interested in outcome variables that include epidemic duration, basic reproduction

number for high transmissibility, and the prevalence peak size for both high and low scenarios. Lastly, an advantage of the nonlinear system of Volterra integral equations over the standard SIR model, is that the probability density function for the infectious period of a specific infectious disease can be identifiable directly from epidemiological data. By relating the main conclusions from Chapters three and five, it can be concluded that, for relatively low transmissibility (around $\mathscr{R}_0 = 1.5$), even though, in the deterministic sense, the variance of the Gamma distribution for the infectious period is a key input parameter in the predictability of variables such as the peak timing of the prevalence of infectious individuals and the epidemic duration, in the parameter estimation (or inverse problem) context, it does not. However, for high transmissibility (around $\mathscr{R}_0 = 15$), even though, in the deterministic sense, the variance of the Gamma distribution for the infectious period is a key input parameter in the predictability of peak timing of the prevalence of infectious individuals, in the parameter estimation (or inverse problem) context, it does not. Another observation, for high transmissibility (around $\mathscr{R}_0 = 15$), is that while in the deterministic sense, the variance of the Gamma distribution for the infectious period does not play any role in the predictability of the basic reproduction number, in the parameter estimation (or inverse problem) context, it does. Generally speaking, these discrepancies are justifiable, attributed or induced by the ill-posedness nature of the parameter estimation (or inverse) problem.

6.2 Summary of main contributions

In summary, the main contributions of the work included in this dissertation are:

• From Chapter three the main contribution lies in the development of a model based decision-tool that helps determine when Volterra integral equations are equivalent or better suited than ordinary differential equations models in predicting epidemiological outcome variables considered.

• The application in Chapter four is designed to determine whether the non-exponential (Erlang) distribution for the infection period is an important aspect to be considered in the estimation of the probability of an event, defined from the empirical distribution of any of the random variables considered.

• From Chapter five an application designed to determine from incidence data whether there is sufficient statistical evidence to conclude that the infectious period distribution is non-Exponentially distributed is developed. In addition, a method for estimating the explicitly specified non-exponential parametric probability density function for the infectious period from epidemiological data is developed.

6.3 Future work

Further explorations to be considered for the methodologies presented are:

• Moving beyond the homogenous mixing assumption for the contact of individuals so as to include population structure.

• Moving beyond the constant parameters thought time assumption by applying the Optimal Control Theory to epidemiological models; that is, we would like to consider control functions in the formula of the distribution of the infectious period.

• As an effort to bring all the methodologies presented in this dissertation into practice two essential tasks need to be considered: First, *enhancement* of the codes and programs used for the implementation of all the methodologies through high performance computing (or parallel computing) and second, increase the *accessibility* to a general audience through graphical user interfaces and the creation of applications, where users may be capable to build their own SIR-type model by choosing the compartments, distributions associated and input parameter values and which results (or graphs) wish to produce and display.

REFERENCES

[1] J. Ma and D. J. D. Earn, *Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease*, Bulletin of Mathematical Biology, **68** (2006).

[2] F.O. Fasina, S.P. Bisschop, T.M. Joannis, L.H. Lombin, C. Abolnik, *Molecular characterization and epidemiology of the highly pathogenic avian influenza H5N1 in Nigeria*, Epidemiol Infect., **137** (2009), 456–463.

[3] L. Fiebig, T. Smieszek, J. Saurina, J. Hattendorf and J. Zinsstag, *Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness*, Geospatial Health **4**(1), (2009), pp. 79–95.

[4] E. H. Kaplan and E. O′Keefe, *Let the Needles Do the Talking! Evaluating the New Haven Needle Exchange*, INTERFACES **23**: 1 (1993).

[5] R. HEIMER, E. H. KAPLAN, K. KHOSHNOOD, B. JARIWALA and E. C. CADMAN, *Needle Exchange Decreases the Prevalence of HIV-1 Proviral DNA in Returned Syringes in New Haven, Connecticut* The American Journal of Medicine Vol. **95** (1993).

[6] E. H. Kaplan and R. Heimer, *HIV Prevalence Among Intravenous Drug Users: Model-Based Estimates from New Haven's Legal Needle Exchange* Journal of Acquired Immune Deficiency Syndromes Vol. **5** 163–169 (1992).

[7] C. Castillo-Chavez and S. Lee, *Epidemiology Modeling*, In, Bjrn Engquist (Ed.) Encyclopedia of Applied and Computational Mathematics, Springer New York Heidelberg, 2013.

[8] J. L. Beven II *et al.*, *Atlantic Hurricane Season of 2005*, Annual Summary, Tropical Prediction Center, NOAA/NWS/National Hurricane Center, Miami, Florida, March 2008.

[9] D. E. Wildasin, *Disasters: Issues for State and Federal Government Finances*, IFIR Working Paper No. 2006-07.

[10] NBC News, *Katrina forecasters were remarkably accurate: Levee breaks, catastrophic damage predicted, contrary to Bush claims*, Hurricane archives on NBCNEWS.com, updated 9/19/2005 8:14:57 AM ET.

[11] S. K. Schneider, *Administrative Breakdowns in the Governmental Response to Hurricane Katrina*, Public Administration Review, Sept./Oct. 2005, Vol. **65**, No. 5.

[12] R. S. Sobel and P. T. Leeson, *Governments response to Hurricane Katrina: A public choice analysis*, Public Choice **127** (2006).

[13] B. A. Cunha, *Influenza: historical aspects of epidemics and pandemics*, Infect Dis Clin North Am., **18** (2004), 141–55.

[14] C.J. Murray, A.D. Lopez, B. Chin, D. Feehan, K.H. Hill, *Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918-20 pandemic: a quantitative analysis*, Lancet, **368** (2006), 2211–2218.

[15] H. Markel, H.B. Lipman, J.A. Navarro, A. Sloan, J.R. Michalsen, A.M. Stern, M.S. Cetron, *Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic*, JAMA, **298** (2007), 644–654.

[16] E. Sydenstricker, *Variations in case fatality during the influenza epidemic of 1918*, Public Health Reports, **36** (1921), 2201–2211.

[17] A. McLeod, *Economics of avian influenza management and control in a world with competing agendas*, Avian Diseases, **54** (2010), 374–379.

[18] The World Health Organization (WHO), *Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO*, Accessed August 31, 2010.

[19] S. Ross, *Continuous Random Variables*, In: "A First Course in Probability", S. Ross, Pearson Pertice Hall, 2006, Chapter 5, pages 205-257.

[20] Z. Feng, W. Huang, and C. Castillo-Chavez *On the Role of Variable Latent Periods in Mathematical Models for Tuberculosis*, Journal of Dynamics and Differential Equations, **13** (2001).

[21] P. Yan and Z. Feng, *Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness*, Mathematical Biosciences, **224** (2010).

[22] Z. Feng, D. Xub and H. Zhao, *Epidemiological Models with Non-Exponentially Distributed Disease Stages and Applications to Disease Control*, Bulletin of Mathematical Biology, **69** (2007).

[23] V. Volterra, *Biography of Vito Volterra* by E. T. Whittaker, In: "Theory of Functionals: And of Integral and Integro-Differenctial Equations", V. Volterra, Dover Publications, Inc, 1959, Biography of Vito Volterra, pages 5-28.

[24] W. O. Kermack and A. G. McKendrick, *A contribution to the mathematical theory of epidemics*, xxx, **xxx** (1927).

[25] P. E. Sartwell, *The distribution of incubation periods of infectious disease*, American journal of hygiene, **51** (1950).

[26] P. E. Sartwell, *The incubation period of poliomyelitis*, American journal of public health, **42** (1952).

[27] P. E. Sartwell, *The incubation period and the dynamics of infectious disease*, American journal of epidemiology, **83** (1966).

[28] Norman T. J. Bailey *Some Stochastic Models for Small Epidemics in Large Populations*, Journal of the Royal Statistical Society. Series C (Applied Statistics), **13** (1964).

[29] D. Anderson and R. Watson, *On the spread of a disease with gamma distributed latent and infectious periods*, Biometrika, **67** (1980).

[30] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L.-M. Ho, T.-Q. Thach, P. Chau, K.-P. Chan, T.-H. Lam, L.-Y. Tse, T. Tsang, S.-H. Liu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, R. M. Anderson, *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong*, THE LANCET, **xxx** (2003).

[31] H. J. Wearing, P. Rohani and M. J. Keeling, *Appropriate Models for the Management of Infectious Diseases*, PLoS Medicine, **2** (2005).

[32] Z. Feng, *Final and peak epidemic sizes for SEIR models with quarantine and isolation*, Mathematical biosciences and engineering, **4** (2007).

[33] P. Yan, *Separate Roles of the Latent and Infectious Periods in Shaping the Relation Between the Basic Reproduction Number and the Intrinsic Growth Rate of Infectious Disease Outbreaks*, Journal of Theoretical Biology, **251** (2008).

[34] C. K. Yang, *Using age of infection models to derive an explicit expression for R0*, Thesis from The University of British Columbia, **XXX** (2007).

[35] C. K. Yang and F. Brauer *Calculation of R0 for age-of-infection models*, Mathematical biosciences and engineering, **5** (2008).

[36] F. Brauer, *Age-of-infection and the final size relation*, Mathematical biosciences and engineering, **5** (2008).

[37] A.L. Lloyd, *Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions*, In: "Mathematical and Statistical Estimation Approaches in Epidemiology", G. Chowell, M. Hyman, L. Bettencourt and C. Castillo-Chavez (eds.), Springer, 2009, Chapter 6, pages 123-141.

[38] E. Vergu, H. Busson and P. Ezanno, *Impact of the Infection Period Distribution on the Epidemic Spread in a Metapopulation Model*, PLoS one, **5** (2010).

[39] M. D. McKay, R. J. Beckman and W. J. Conover, *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code*, Technometrics, **21**, (1979).

[40] S. M. Blower and H. Dowlatabadi, *Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: An HIV Model, as an Example*, International Statistical Review / Revue Internationale de Statistique, **62** (1994).

[41] H. Brunner and P. J. Van der Houwen, "The Numerical Solution of Volterra Equations", Cwi Monographs, 1986.

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

[43] G. Chowell, N.W. Hengartner, C. Castillo-Chavez, P.W. Fenimore, J.M. Hyman, *The basic reproduction number of Ebola and the effects of public health measures: the cases of Congo and Uganda*, J Theor Biol., **229** (2004), 119–126.

[44] G. Chowell, C.E. Ammon, N.W. Hengartner, J.M. Hyman, *Estimating the reproduction number from the initial phase of the Spanish flu pandemic waves in Geneva, Switzerland*, Math Biosci Eng., **4** (2007), 457–470.

[45] D. Rios-Doria, G. Chowell, *Qualitative analysis of the level of cross-protection between epidemic waves of 1918-1919 influenza pandemic*, J Theor Biol., **261** (2009), 584–592.

[46] G. Chowell, P. Diaz-Dueñas, J.C. Miller, A. Alcazar-Velazco, J.M. Hyman, P.W. Fenimore, C. Castillo-Chavez, *Estimation of the reproduction number of dengue fever from spatial epidemic data*, Mathematical Biosciences, **208** (2007), 571–589.

[47] G. Chowell, P. W. Fenimore, M. A. Castillo-Garsow, C. Castillo-Chavez, *SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism*, Journal of Theoretical Biology, **224** (2003).

[48] K.L. Sutton, H.T. Banks, C.Castillo-Chavez, *Estimation of invasive pneumococcal disease dynamics parameters and the impact of conjugate vaccination in Australia*, Math Biosci Eng., **5** (2008), 175–204.

[49] B. P. Holder and C. AA. Beauchemin, *Exploring the Effect of Biological Delays in Kinetic Models of Influenza Within a Host or Cell Culture*, BioMed Central Public Health, **11** (2011).

[50] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore and J. M. Hyman, *The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda*, Journal of Theoretical Biology, **229** (2004).

[51] P. Piot, P. Bureau, G. D. Breman, D. L. Heymann, et al. *Clinical aspects of Ebola virus infectionin Yambuku are , Zaire, 1976*, Ebola virus haemorrhagic fever: proceedings of an international colloquium on Ebola virus infection and other haemorrhagic fevers, held in Antwerp, Belgium **69** (1977).

[52] J. G. Breman, G. van der Groen, C. J. Peters, D. L. Heymann, *International colloquium on Ebola virus research: summary report*, The Journal of Infectious Diseases, **176** (1997).

[53] H.T. Banks, H.T. Tran, "Mathematical and experimental modeling of physical and biological processes", CRC Press, Boca Raton, FL, 2009.

[54] A. Cintrón-Arias, H.T. Banks, A. Capaldi, A. Lloyd, *A sensitivity matrix based methodology for inverse problem formulation*, Journal of Inverse & Ill-Posed Problems., **17** (2009), 545–564.

[55] H.T. Banks, S. Dediu, S.L. Ernstberger, *Sensitivity functions and their uses in inverse problems*, Journal of Inverse & Ill-Posed Problems., **15** (2007), 683–708.

[56] H. T. Banks, S. L. Ernstberger, S. L. Grove, *Standard errors and condence intervals in inverse problems: sensitivity and associated pitfalls*, Journal of Inverse & Ill-Posed Problems., **15** (2007), 1–18.

[57] F. Brauer, C. Castillo-Chavez, "Mathematical Models in Population Biology and Epidemiology", Springer, New York, Sartwell, P.E., 2000.

[58] G. Chowell, H. Nishiura, L.M.A. Bettencourt, *Comparative estimation of the reproduction number for pandemic influenza from daily case notification data*, J R Soc Interface, **4** (2007), 155–166.

[59] J. A. Rice, "Mathematical statistics and data analysis", 2^{nd} edition, China machine press, 2003.

[60] SARS Expert Committee, *SARS in Hong Kong: from Experience to Action: Chapter 3: The SARS Epidemic*, www.sars-expertcom.gov.hk., (2003).

[61] T. Tsang and T.H. LAM, *SARS: public health measures in Hong Kong*, Respirology, **8** (2003).

[62] S. Riley and et al., *Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions*, Science, **300** (2003).

[63] WHO, *Cumulative number of reported probable cases of SARS*, www.who.int/csr/sars/country/en. (Accessed March 2012).

[64] A. Capaldi, S. Behrend, B. Berman, J. Smith, J. Wright, A. L. Lloyd, *Parameter estimation and uncertainty quantication for an epidemic model (In preparation)*.

[65] H. Brunner and P. J. Van der Houwen, "The Numerical Solution of Volterra Equations", Cwi Monographs, 1986.

[66] R. K. Miller, "Nonlinear Volterra Integral Equations", W. A. Benjamin, Inc., Menlo Park, California, 1971.

[67] P. van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Mathematical Biosciences, **180** (2002).

[68] F. Brauer and C. Castillo-Chavez, "Mathematical Models in Population Biology and Epidemiology", Springer, New York, Sartwell, P.E., 2000.

[69] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*, Wiley, Chichester, (2000).

[70] F. Brauer and J. A. Nohel, "The Qualitative Yheory of Ordinary Differential Equations An Introduction", Dover publications inc., New York, 1969.

[71] H. W. Hethcote, *Three Basic Epidemiological Models*, Biomathematics , **18** (1989).

[72] W. H. Press, W. T. Vetterling, S. A. Teukolsky, B. P. Flannery, "Numerical recipes in $C++$: The art of scientific computing", Cambridge University Press, second edition, 2002.

APPENDIX A

REWRITING A NONLINEAR SYSTEM OF VOLTERRA INTEGRO-ORDINARY DIFFERENTIAL EQUATIONS INTO A NONLINEAR SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS: THE CASE OF EXPONENTIALLY DISTRIBUTED INFECTIOUS PERIOD

Lets assume that the survivor function is of the form:

$$\bar{F}(t) = \begin{cases} e^{-\frac{1}{\tau}t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases}$$

or equivalently, lets assume that the probability density function for the infectious period has the following form:

$$[-\dot{\bar{F}}(t)] = \begin{cases} \frac{1}{\tau}e^{-\frac{1}{\tau}t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases}$$

then

$$\begin{aligned}
\frac{dI}{dt}(t) &= \frac{\beta}{N}I(t)S(t) - \left(I_0[-\dot{\bar{F}}(t)] + \int_0^t \frac{\beta}{N}I(s)S(s)[-\dot{\bar{F}}(t-s)]ds \right) \\
&= \frac{\beta}{N}I(t)S(t) - \left(I(0)\frac{1}{\tau}e^{-\frac{1}{\tau}t} + \int_0^t \frac{\beta}{N}I(s)S(s)\frac{1}{\tau}e^{-\frac{1}{\tau}(t-s)}ds \right) \\
&= \frac{\beta}{N}I(t)S(t) - \frac{1}{\tau}\left(I(0)e^{-\frac{1}{\tau}t} + \int_0^t \frac{\beta}{N}I(s)S(s)e^{-\frac{1}{\tau}(t-s)}ds \right) \\
\frac{dI}{dt}(t) &= \frac{\beta}{N}I(t)S(t) - \frac{1}{\tau}\underbrace{\left(I_0\bar{F}(t) + \int_0^t \frac{\beta}{N}I(s)S(s)\bar{F}(t-s)ds, \right)}_{\text{By definition this expression is } I(t).}, \\
\frac{dS}{dt}(t) &= -\frac{\beta}{N}I(t)S(t), \quad \frac{dI}{dt}(t) = \frac{\beta}{N}I(t)S(t) - \frac{1}{\tau}I(t).
\end{aligned}$$

APPENDIX B

REWRITING A NONLINEAR SYSTEM OF VOLTERRA INTEGRO-ORDINARY DIFFERENTIAL EQUATIONS INTO A LARGER DIMENSIONAL NONLINEAR SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS VIA THE LINEAR CHAIN TRICKERY: THE CASE OF ERLANG DISTRIBUTED INFECTIOUS PERIOD

Lets assume that the probability density function for the infectious period has the following form:

$$[-\dot{\bar{F}}(t)] = \begin{cases} \frac{t^{k-1}e^{-\frac{t}{(\frac{\tau}{k})}}}{(k-1)!(\frac{\tau}{k})^k} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases} = \begin{cases} \frac{(k\frac{1}{\tau})^k t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases}$$

where k denotes the shape parameter and $\frac{\tau}{k}$ the scale parameter, then

$$\begin{aligned} \frac{dI}{dt}(t) &= \frac{\beta}{N}I(t)S(t) - \left(I_0[-\dot{\bar{F}}(t)] + \int_0^t \frac{\beta}{N}I(s)S(s)[-\dot{\bar{F}}(t-s)]ds \right) \\ &= \frac{\beta}{N}I(t)S(t) \\ &- \left(I_0\left(\frac{(k\frac{1}{\tau})^k t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!} \right) + \int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^k (t-s)^{k-1}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!} \right)ds \right) \\ &= \frac{\beta}{N}I(t)S(t) \\ &- (k\frac{1}{\tau}) \underbrace{\left(I_0\left(\frac{(k\frac{1}{\tau})^{k-1} t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!} \right) + \int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-1} (t-s)^{k-1}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!} \right)ds \right)}_{\text{Define this term as } I_k(t).} \\ \frac{d}{dt}I(t) &= \frac{\beta}{N}I(t)S(t) - (k\frac{1}{\tau})I_k(t), \end{aligned}$$

Computing the time derivative of $I_k(t)$ we obtain:

$$
\begin{aligned}
I_k(t) &= I_0\left(\frac{(k\frac{1}{\tau})^{k-1}t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!}\right)+\int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-1}(t-s)^{k-1}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!}\right)ds,\\
\frac{d}{dt}I_k(t) &= I_0\left(\frac{(k\frac{1}{\tau})^{k-1}(k-1)t^{k-2}e^{-(k\frac{1}{\tau})t}}{(k-1)!}+\frac{(k\frac{1}{\tau})^{k-1}t^{k-1}(-k\frac{1}{\tau})e^{-(k\frac{1}{\tau})t}}{(k-1)!}\right)\\
&+ \frac{\beta}{N}I(t)S(t)\underbrace{\left(\frac{(k\frac{1}{\tau})^{k-1}(t-t)^{k-1}e^{-(k\frac{1}{\tau})(t-t)}}{(k-1)!}\right)\frac{dt}{dt}}_{\text{This term is } \delta_{1,k}, \text{ since } 0^0=1}+\underbrace{\frac{\beta}{N}I(t)S(t)\left(\frac{(k\frac{1}{\tau})^{k-1}(t-0)^{k-1}e^{-(k\frac{1}{\tau})(t-0)}}{(k-1)!}\right)\frac{d0}{dt}}_{0}\\
&+ \int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-1}(k-1)(t-s)^{k-2}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!}+\frac{(k\frac{1}{\tau})^{k-1}(t-s)^{k-1}(-k\frac{1}{\tau})e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!}\right)ds\\
&= \frac{\beta}{N}I(t)S(t)\delta_{1,k}+(k\frac{1}{\tau})I_0\left(\frac{(k\frac{1}{\tau})^{k-2}t^{k-2}e^{-(k\frac{1}{\tau})t}}{(k-2)!}\right)-(k\frac{1}{\tau})I_0\left(\frac{(k\frac{1}{\tau})^{k-1}t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!}\right)\\
&+ (k\frac{1}{\tau})\int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-2}(t-s)^{k-2}e^{-(k\frac{1}{\tau})(t-s)}}{(k-2)!}\right)ds\\
&- (k\frac{1}{\tau})\int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-1}(t-s)^{k-1}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!}\right)ds\\
&= \frac{\beta}{N}I(t)S(t)\delta_{1,k}\\
&+ (k\frac{1}{\tau})\underbrace{\left[I_0\left(\frac{(k\frac{1}{\tau})^{k-2}t^{k-2}e^{-(k\frac{1}{\tau})t}}{(k-2)!}\right)+\int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-2}(t-s)^{k-2}e^{-(k\frac{1}{\tau})(t-s)}}{(k-2)!}\right)ds\right]}_{\text{This term is } I_{k-1}. \text{ This is possible since the term } k\frac{1}{\tau} \text{ as a whole is considered a symbol (the rate parameter).}}\\
&- (k\frac{1}{\tau})\underbrace{\left[I_0\left(\frac{(k\frac{1}{\tau})^{k-1}t^{k-1}e^{-(k\frac{1}{\tau})t}}{(k-1)!}\right)+\int_0^t \frac{\beta}{N}I(s)S(s)\left(\frac{(k\frac{1}{\tau})^{k-1}(t-s)^{k-1}e^{-(k\frac{1}{\tau})(t-s)}}{(k-1)!}\right)ds\right]}_{\text{This term is } I_k}\\
\frac{d}{dt}I_k(t) &= \frac{\beta}{N}I(t)S(t)\delta_{1,k}+(k\frac{1}{\tau})I_{k-1}-(k\frac{1}{\tau})I_k.
\end{aligned}
$$

As the results from the last equation above we obtained the following reduction to a system of non-linear ordinary differential equations:

$$\begin{aligned}
\frac{d}{dt}S(t) &= -\frac{\beta}{N}I(t)S(t), \\
\frac{d}{dt}I(t) &= \frac{\beta}{N}I(t)S(t) - (k\frac{1}{\tau})I_k(t), \\
\frac{d}{dt}I_k(t) &= (k\frac{1}{\tau})I_{k-1} - (k\frac{1}{\tau})I_k, \\
\frac{d}{dt}I_{k-1}(t) &= (k\frac{1}{\tau})I_{k-2} - (k\frac{1}{\tau})I_{k-1}, \\
&\vdots \\
\frac{d}{dt}I_2(t) &= (k\frac{1}{\tau})I_1 - (k\frac{1}{\tau})I_2, \\
\frac{d}{dt}I_1(t) &= \frac{\beta}{N}I(t)S(t) - (k\frac{1}{\tau})I_1. \text{ (from the second equation in previous page)}
\end{aligned}$$

Given that $I(t) = I_1(t) + I_2(t) + \cdots + I_k(t)$, the system above can be solved without including the $\frac{d}{dt}I(t)$ equation. Therefore the system can be rewritten as follows:

$$\begin{aligned}
\frac{d}{dt}S(t) &= -\frac{\beta}{N}I(t)S(t), \\
\frac{d}{dt}I_1(t) &= \frac{\beta}{N}I(t)S(t) - (k\frac{1}{\tau})I_1, \\
\frac{d}{dt}I_2(t) &= (k\frac{1}{\tau})I_1 - (k\frac{1}{\tau})I_2, \\
&\vdots \\
\frac{d}{dt}I_{k-1}(t) &= (k\frac{1}{\tau})I_{k-2} - (k\frac{1}{\tau})I_{k-1}, \\
\frac{d}{dt}I_k(t) &= (k\frac{1}{\tau})I_{k-1} - (k\frac{1}{\tau})I_k.
\end{aligned}$$

APPENDIX C

REWRITING A NONLINEAR SYSTEM OF VOLTERRA INTEGRO-ORDINARY DIFFERENTIAL EQUATIONS INTO A NONLINEAR SYSTEM OF DISCRETE DELAY DIFFERENTIAL EQUATIONS: THE CASE OF FIXED INFECTIOUS PERIOD

Lets assume that the survivor function is of the form:

$$\bar{F}(t) = \begin{cases} 1 & \text{for } 0 \le t < \tau, \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\begin{aligned}
I(t) &= I_0\bar{F}(t) + \int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)ds \\
&= I_0\bar{F}(t) + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)\bar{F}(s)ds \\
&= \begin{cases} I_0 + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)ds & \text{for } 0 \le t < \tau, \\ \int_0^\tau \frac{\beta}{N} I(t-s)S(t-s)ds & \text{otherwise} \end{cases} \\
&= \begin{cases} I_0 + \int_0^t [-\dot{S}(t-s)]ds & \text{for } 0 \le t < \tau, \\ \int_0^\tau [-\dot{S}(t-s)]ds & \text{otherwise} \end{cases} \\
&= \begin{cases} I_0 + \int_t^0 \dot{S}(s)ds & \text{for } 0 \le t < \tau, \\ \int_t^{t-\tau} \dot{S}(s)ds & \text{otherwise} \end{cases} \\
I(t) &= \begin{cases} I_0 + S_0 - S(t) & \text{for } 0 \le t < \tau, \\ S(t-\tau) - S(t) & \text{otherwise,} \end{cases} \\
\frac{dI}{dt}(t) &= \begin{cases} -\dot{S}(t) & \text{for } 0 \le t < \tau, \\ \dot{S}(t-\tau) - \dot{S}(t) & \text{otherwise} \end{cases} \\
\frac{dI}{dt}(t) &= \begin{cases} \frac{\beta}{N} I(t)S(t) & \text{for } 0 \le t < \tau, \\ \frac{\beta}{N} I(t)S(t) - \frac{\beta}{N} I(t-\tau)S(t-\tau) & \text{otherwise,} \end{cases} \\
\frac{dS}{dt}(t) &= -\frac{\beta}{N} I(t)S(t), \quad \frac{dI}{dt}(t) = \begin{cases} \frac{\beta}{N} I(t)S(t) & \text{for } 0 \le t < \tau, \\ \frac{\beta}{N} I(t)S(t) - \frac{\beta}{N} I(t-\tau)S(t-\tau) & \text{otherwise.} \end{cases}
\end{aligned}$$

APPENDIX D

A 4^{TH}-STAGE VOLTERRA-RUNGE-KUTTA FORMULA OF POUZET TYPE

• 4^{th}-stage PVRK formula:

$$Y_{n,j} = \tilde{F}_n(t_n + c_j h) + h \sum_{i=1}^{4} a_{j,i} k(t_n + c_j h, t_n + c_i h, Y_{n,i}), \quad j = 1, \ldots, 4$$

$$y_{n+1} = \tilde{F}_n(t_n + h) + h\tilde{\Phi}_n(t_n + h), \quad n = 0, \ldots, N-1$$

• Lag term formula:

$$\tilde{F}_n(t) := g(t) + h \sum_{l=0}^{n-1} \sum_{j=1}^{4} b_j k(t, t_l + c_j h, Y_{l,j}), \quad n = 0, \ldots, N-1$$

• PVRK formula:

$$\tilde{\Phi}_n(t) := \sum_{j=1}^{4} b_j k(t, t_n + c_j h, Y_{n,j})$$

• Butcher array:

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|ccc} c_1 & a_{1,1} & \cdots & a_{1,4} \\ \vdots & \vdots & & \vdots \\ c_4 & a_{4,1} & \cdots & a_{4,4} \\ \hline & b_1 & \cdots & b_4 \end{array} = \begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

• The 4^{th}-stage PVRK formula is explicit since A is strictly lower triangular matrix
Adapted from the 1986 book by H. Brunner and P. J. Van der Houwen on "The Numerical Solution of Volterra Equations".

APPENDIX E

A 4^{TH}-STAGE VIODE-RUNGE-KUTTA FORMULA OF POUZET TYPE

- 4^{th}-stage PVDRK formula:

$$Y_{n,j} = y_n + h\sum_{i=1}^{4} a_{j,i} f(t_n + c_i h, Y_{n,i}, \tilde{F}_n(t_n + c_i h) + h\tilde{\Phi}_n(t_n + c_i h)), \quad j = 1,\ldots,4$$

with $y_0 = y(0)$

$$y_{n+1} = y_n + h\sum_{j=1}^{4} b_j f(t_n + c_j h, Y_{n,j}, \tilde{F}_n(t_n + c_j h) + h\tilde{\Phi}_n(t_n + c_j h)), \quad n = 0,\ldots,N-1$$

- Lag term formula:

$$\tilde{F}_n(t) := h\sum_{l=0}^{n-1}\sum_{j=1}^{4} b_j k(t, t_l + c_j h, Y_{l,j}), \quad n = 0,\ldots,N-1$$

- PVDRK formula:

$$\tilde{\Phi}_n(t) := \sum_{l=1}^{4} a_{i,l} k(t, t_n + c_l h, Y_{n,l})$$

- Butcher array:

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|ccc} c_1 & a_{1,1} & \cdots & a_{1,4} \\ \vdots & \vdots & & \vdots \\ c_4 & a_{4,1} & \cdots & a_{4,4} \\ \hline & b_1 & \cdots & b_4 \end{array} = \begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

- This 4^{th}-stage PVDRK formula is explicit since A is strictly lower triangular matrix **Adapted from the 1986 book by H. Brunner and P. J. Van der Houwen on "The Numerical Solution of Volterra Equations".**

APPENDIX F

DERIVATION OF THE BASIC REPRODUCTION NUMBER $\mathscr{R}_0$ FOR THE AGE-OF-INFECTION SIR MODEL

When the epidemiological model is described by a system of ordinary differential equations, the basic reproduction number is traditionally calculated via the next generator operator (see [69]). For the age-of-infection SIR model in (2.4)-(2.6) the same method does not apply. Here we derive the basic reproduction number in two different ways.

The first and easy way is a heuristic approach. Strictly by definition we have that β is the transmission rate, which by simplicity can be composed by two main terms, the average number of contacts that a single individual have, times the probability of the individual to infect others. The expected infectious period of an individual τ by definition is given by $\int_{-\infty}^{\infty} t f(t)dt$ or by a simple integration by parts $\int_0^{\infty} \bar{F}(t)dt$. Then heuristically, by definition of the basic reproduction number, it is given by:

$$\mathscr{R}_0 = \beta\Big(\int_{-\infty}^{\infty} t f(t)dt\Big) = \beta\Big(\int_0^{\infty} \bar{F}(t)dt\Big) = \beta\tau.$$

The second approach is via the stability analysis of the endemic equilibrium of a similar model as the age-of-infection SIR model in (2.4)-(2.6), but with demographic terms, constant and identical births and deaths rates. Consider the following model:

$$\begin{aligned}
S(t) &= \int_0^t N(\mu e^{-\mu(t-s)})ds - \int_0^t \frac{\beta}{N} I(s)S(s)e^{-\mu(t-s)}ds + S_0 e^{-\mu t},\\
I(t) &= I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)e^{-\mu(t-s)}ds,\\
R(t) &= R_0 e^{-\mu t} + \int_0^t \Big[I_0 f(s)e^{-\mu s} + \int_0^s \frac{\beta}{N} I(\tau)S(\tau)f(s-\tau)e^{-\mu(s-\tau)}d\tau\Big]e^{-\mu(t-s)}ds
\end{aligned}$$

By integrating the first term in the $S(t)$ equation, a change of variable ($u = t-s$) in the second term of the $S(t)$ and $I(t)$ equations and changing the order of integration in the $R(t)$ equation we have:

$$\begin{aligned}
S(t) &= N(1-e^{-\mu t}) - \int_0^t \frac{\beta}{N} I(t-s)S(t-s)e^{-\mu(s)}ds + S_0 e^{-\mu t},\\
I(t) &= I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)\bar{F}(s)e^{-\mu s}ds,\\
R(t) &= R_0 e^{-\mu t} + \int_0^t I_0 f(s)e^{-\mu s}e^{-\mu(t-s)}ds + \int_0^t\int_\tau^t \frac{\beta}{N} I(\tau)S(\tau)f(s-\tau)e^{-\mu(s-\tau)}e^{-\mu(t-s)}dsd\tau
\end{aligned}$$

$$\begin{aligned}
S(t) &= N(1-e^{-\mu t}) - \int_0^t \frac{\beta}{N} I(t-s)S(t-s)e^{-\mu(s)}ds + S_0 e^{-\mu t},\\
I(t) &= I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)\bar{F}(s)e^{-\mu s}ds,\\
R(t) &= R_0 e^{-\mu t} + I_0 e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} I(\tau)S(\tau)\left(\int_\tau^t [-\dot{P}_I(s-\tau)]ds\right)e^{-\mu(t-\tau)}d\tau
\end{aligned}$$

$$
\begin{aligned}
S(t) &= N(1-e^{-\mu t}) - \int_0^t \frac{\beta}{N} I(t-s)S(t-s)e^{-\mu(s)}ds + S_0 e^{-\mu t}, \\
I(t) &= I_0 \bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)\bar{F}(s)e^{-\mu s}ds, \\
R(t) &= R_0 e^{-\mu t} + I_0 e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} I(s)S(s)(1-\bar{F}(t-s))e^{-\mu(t-s)}ds
\end{aligned}
$$

By changing the order of integration in the $R(t)$ equation we have:

$$
\begin{aligned}
S(t) &= N(1-e^{-\mu t}) - \int_0^t \frac{\beta}{N} I(t-s)S(t-s)e^{-\mu(s)}ds + S_0 e^{-\mu t}, \\
I(t) &= I_0 \bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)\bar{F}(s)e^{-\mu s}ds, \\
R(t) &= R_0 e^{-\mu t} + I_0 e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} I(t-s)S(t-s)(1-\bar{F}(s))e^{-\mu s}ds
\end{aligned}
$$

The endemic equilibrium of the model is obtained as follows:

$$
\begin{aligned}
S_\infty = \lim_{t\to\infty} S(t) &= \lim_{t\to\infty} \left(N(1-e^{-\mu t}) + S_0 e^{-\mu t}\right) - \int_0^\infty \frac{\beta}{N} I_\infty S_\infty e^{-\mu(s)}ds, \\
I_\infty = \lim_{t\to\infty} I(t) &= \lim_{t\to\infty} I_0 \bar{F}(t)e^{-\mu t} + \int_0^\infty \frac{\beta}{N} I_\infty S_\infty \bar{F}(s)e^{-\mu(s)}ds, \\
R_\infty = \lim_{t\to\infty} R(t) &= \lim_{t\to\infty} \left(R_0 e^{-\mu t} + I_0 e^{-\mu t}(1-\bar{F}(t))\right) + \int_0^\infty \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(s))e^{-\mu s}ds
\end{aligned}
$$

$$
\begin{aligned}
S_\infty &= N - \int_0^\infty \frac{\beta}{N} I_\infty S_\infty e^{-\mu s}ds, \\
I_\infty &= \int_0^\infty \frac{\beta}{N} I_\infty S_\infty \bar{F}(s)e^{-\mu s}ds, \\
R_\infty &= \int_0^\infty \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(s))e^{-\mu s}ds
\end{aligned}
$$

$$
\begin{aligned}
S_\infty &= N - \frac{\beta}{N} I_\infty S_\infty \int_0^\infty e^{-\mu s}ds, \\
I_\infty &= \frac{\beta}{N} I_\infty S_\infty \int_0^\infty \bar{F}(s)e^{-\mu s}ds, \\
R_\infty &= \frac{\beta}{N} I_\infty S_\infty \left(\int_0^\infty e^{-\mu s}ds - \int_0^\infty \bar{F}(s)e^{-\mu s}ds\right)
\end{aligned}
$$

$$
\begin{aligned}
S_\infty &= N - \frac{\beta}{N} I_\infty S_\infty \frac{1}{\mu}, \\
I_\infty &= \frac{\beta}{N} I_\infty S_\infty \int_0^\infty \bar{F}(s) e^{-\mu s} ds, \\
R_\infty &= \frac{\beta}{N} I_\infty S_\infty \left(\frac{1}{\mu} - \int_0^\infty \bar{F}(s) e^{-\mu s} ds \right)
\end{aligned}
$$

Assuming that $I_\infty \neq 0$, we solve for S_∞ in the equation for I.

$$
\begin{aligned}
1 &= \frac{\beta}{N} S_\infty \int_0^\infty \bar{F}(s) e^{-\mu s} ds, \\
S_\infty &= \frac{N}{\beta \int_0^\infty \bar{F}(s) e^{-\mu(s)} ds} = \frac{N}{\tilde{\mathscr{R}}_0},
\end{aligned}
$$

We solve for I_∞ in the equation for S.

$$
\begin{aligned}
S_\infty &= N - \frac{\beta}{N} I_\infty S_\infty \frac{1}{\mu}, \\
1 &= \frac{N}{S_\infty} - \frac{\beta}{N} I_\infty \frac{1}{\mu}, \\
I_\infty &= \left(\frac{N}{S_\infty} - 1 \right) \frac{N\mu}{\beta}, \\
I_\infty &= \left(\frac{N}{\frac{N}{\tilde{\mathscr{R}}_0}} - 1 \right) \frac{N\mu}{\beta} = \frac{N\mu}{\beta} \left(\tilde{\mathscr{R}}_0 - 1 \right),
\end{aligned}
$$

We solve for R_∞ in the equation for R.

$$
\begin{aligned}
R_\infty &= \frac{\beta}{N} I_\infty S_\infty \left(\frac{1}{\mu} - \int_0^\infty \bar{F}(s) e^{-\mu s} ds \right) \\
R_\infty &= \frac{1}{N} I_\infty S_\infty \left(\frac{\beta}{\mu} - \beta \int_0^\infty \bar{F}(s) e^{-\mu s} ds \right) \\
R_\infty &= \frac{1}{N} \left(\frac{N\mu}{\beta} \left(\tilde{\mathscr{R}}_0 - 1 \right) \right) \left(\frac{N}{\tilde{\mathscr{R}}_0} \right) \left(\frac{\beta}{\mu} - \tilde{\mathscr{R}}_0 \right)
\end{aligned}
$$

Endemic equilibrium:

$$
(S_\infty, I_\infty, R_\infty) = \left(\frac{N}{\tilde{\mathscr{R}}_0}, \frac{N\mu}{\beta} \left(\tilde{\mathscr{R}}_0 - 1 \right), \frac{1}{N} \left(\frac{N\mu}{\beta} \left(\tilde{\mathscr{R}}_0 - 1 \right) \right) \left(\frac{N}{\tilde{\mathscr{R}}_0} \right) \left(\frac{\beta}{\mu} - \tilde{\mathscr{R}}_0 \right) \right)
$$

The stability analysis of the endemic equilibria of the model is as follows. Linearization: Since the total population is constant, we only focused on the $I(t)$ and $R(t)$ equations. First we translate the endemic equilibria to the origin by re-writing the model with $I(t) = I_\infty + V(t)$ and $R(t) = R_\infty + W(t)$:

$$
\begin{aligned}
\begin{bmatrix} I(t) \\ R(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} I(s)S(s)\bar{F}(t-s)e^{-\mu(t-s)}ds \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} I(s)S(s)(1-\bar{F}(t-s))e^{-\mu(t-s)}ds \end{bmatrix} \\
\begin{bmatrix} I_\infty + V(t) \\ R_\infty + W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} (I_\infty + V(s))(N - (I_\infty + V(s)) - (R_\infty + W(s)))\bar{F}(t-s)e^{-\mu(t-s)}ds \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} (I_\infty + V(s))(I_\infty + V(s))(N - (I_\infty + V(s)) - (R_\infty + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)}ds \end{bmatrix} \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} -I_\infty + I_0\bar{F}(t)e^{-\mu t} + \int_0^t \frac{\beta}{N} (I_\infty + V(s))(S_\infty - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)}ds \\ -R_\infty + R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) + \int_0^t \frac{\beta}{N} (I_\infty + V(s))(S_\infty - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)}ds \end{bmatrix} \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} -I_\infty + I_0\bar{F}(t)e^{-\mu t} \\ -R_\infty + R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) \end{bmatrix} + \int_0^t \begin{bmatrix} \frac{\beta}{N} (I_\infty S_\infty + S_\infty V(s) - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (I_\infty S_\infty + S_\infty V(s) - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) \end{bmatrix} - \begin{bmatrix} I_\infty \\ R_\infty \end{bmatrix} + \int_0^t \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (I_\infty S_\infty (1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) \end{bmatrix} - \int_0^\infty \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(s)e^{-\mu s}, \\ \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(s))e^{-\mu s} \end{bmatrix} ds + \int_0^t \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (I_\infty S_\infty (1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) \end{bmatrix} - \int_{-\infty}^0 \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(-s)e^{\mu s}, \\ \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(-s))e^{\mu s} \end{bmatrix} ds - \int_{-t}^0 \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(-s)e^{\mu s} \\ \frac{\beta}{N} (I_\infty S_\infty (1-\bar{F}(-s))e^{\mu s} \end{bmatrix} ds \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) \end{bmatrix} - \int_{-\infty}^{-t} \begin{bmatrix} \frac{\beta}{N} I_\infty S_\infty \bar{F}(-s)e^{\mu s}, \\ \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(-s))e^{\mu s} \end{bmatrix} ds \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))\bar{F}(t-s)e^{-\mu(t-s)} \\ \frac{\beta}{N} (S_\infty V(s) - (V(s) + W(s)))(1-\bar{F}(t-s))e^{-\mu(t-s)} \end{bmatrix} ds \\
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} - \int_{-\infty}^{-t} \frac{\beta}{N} I_\infty S_\infty \bar{F}(-s)e^{\mu s}ds \\ R_0e^{-\mu t} + I_0e^{-\mu t}(1-\bar{F}(t)) - \int_{-\infty}^{-t} \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(-s))e^{\mu s}ds \end{bmatrix} \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N} \bar{F}(t-s)e^{-\mu(t-s)} & 0 \\ \frac{\beta}{N} (1-\bar{F}(t-s))e^{-\mu(t-s)} & 0 \end{bmatrix} \times \begin{bmatrix} S_\infty V(s) - (V(s) + W(s)) \\ 0 \end{bmatrix} ds \\
X(t) &= H(t) + \int_0^t A(t-s)G(X(s))ds
\end{aligned}
$$

The characteristic equation is obtained as follows:
Consider the model:

$$\begin{aligned}
\begin{bmatrix} V(t) \\ W(t) \end{bmatrix} &= \begin{bmatrix} I_0\bar{F}(t)e^{-\mu t} - \int_{-\infty}^{-t} \frac{\beta}{N} I_\infty S_\infty \bar{F}(-s)e^{\mu s} ds \\ R_0 e^{-\mu t} + I_0 e^{-\mu t}(1-\bar{F}(t)) - \int_{-\infty}^{-t} \frac{\beta}{N} I_\infty S_\infty (1-\bar{F}(-s))e^{\mu s} ds \end{bmatrix} \\
&+ \int_0^t \begin{bmatrix} \frac{\beta}{N}\bar{F}(t-s)e^{-\mu(t-s)} & 0 \\ \frac{\beta}{N}(1-\bar{F}(t-s))e^{-\mu(t-s)} & 0 \end{bmatrix} \times \begin{bmatrix} S_\infty V(s) - (V(s)+W(s)) \\ 0 \end{bmatrix} ds \\
X(t) &= H(t) + \int_0^t A(t-s)G(X(s))ds
\end{aligned}$$

The characteristic equation of the linearization of the model above is given by:

$$\det\left(\text{Identity} - \int_0^\infty e^{-\lambda t} A(t) J_G(0,0) dt\right) = 0$$

where $J_G(0,0)$ is the Jacobian of G evaluated at the origin.

$$
\begin{aligned}
0 &= \det\left(\text{Identity} - \int_0^\infty e^{-\lambda t} A(t) J_G(0,0) dt\right) \\
&= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \int_0^\infty e^{-\lambda t} \begin{bmatrix} \frac{\beta}{N}\bar{F}(t)e^{-\mu(t)} & 0 \\ \frac{\beta}{N}(1-\bar{F}(t))e^{-\mu(t)} & 0 \end{bmatrix} \times \begin{bmatrix} S_\infty - I_\infty & I_\infty \\ 0 & 0 \end{bmatrix} dt \right| \\
&= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \int_0^\infty e^{-\lambda t} \begin{bmatrix} \frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)} & -\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)} \\ \frac{\beta}{N}(S_\infty - I_\infty)(1-\bar{F}(t))e^{-\mu(t)} & -\frac{\beta}{N}I_\infty(1-\bar{F}(t))e^{-\mu(t)} \end{bmatrix} dt \right| \\
&= \begin{vmatrix} 1 - \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt & \int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt \\ -\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)(1-\bar{F}(t))e^{-\mu(t)}dt & 1 + \int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty(1-\bar{F}(t))e^{-\mu(t)}dt \end{vmatrix} \\
&= \left(1 - \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt\right)\left(1 + \int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty(1-\bar{F}(t))e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)(1-\bar{F}(t))e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt\right) \\
&= 1 + \int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty(1-\bar{F}(t))e^{-\mu(t)}dt \\
&- \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt \\
&- \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty(1-\bar{F}(t))e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)(1-\bar{F}(t))e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt\right) \\
&= 1 + \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(I_\infty - S_\infty\bar{F}(t))e^{-\mu(t)}dt \\
&- \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt - \int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)e^{-\mu(t)}dt - \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt\right) \\
&\times \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt\right) \\
&= 1 + \int_0^\infty e^{-\lambda t}\frac{\beta}{N}(I_\infty - S_\infty\bar{F}(t))e^{-\mu(t)}dt \\
&- \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)\bar{F}(t)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}(S_\infty - I_\infty)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty\bar{F}(t)e^{-\mu(t)}dt\right)
\end{aligned}
$$

$$
\begin{aligned}
0 &= 1+\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt-\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty \bar{F}(t)e^{-\mu(t)}dt \\
&- \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty \bar{F}(t)e^{-\mu(t)}dt-\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty \bar{F}(t)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty e^{-\mu(t)}dt-\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty \bar{F}(t)e^{-\mu(t)}dt\right) \\
&= 1+\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt-\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty \bar{F}(t)e^{-\mu(t)}dt \\
&- \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty \bar{F}(t)e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty e^{-\mu(t)}dt\right) \\
&+ \left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}S_\infty e^{-\mu(t)}dt\right)\left(\int_0^\infty e^{-\lambda t}\frac{\beta}{N}I_\infty \bar{F}(t)e^{-\mu(t)}dt\right) \\
&= 1+\frac{\beta}{N}I_\infty\frac{1}{\lambda+\mu}-\frac{\beta}{N}S_\infty\int_0^\infty \bar{F}(t)e^{-(\lambda+\mu)t}dt \\
&- \left(\frac{\beta}{N}S_\infty\int_0^\infty \bar{F}(t)e^{-(\lambda+\mu)t}dt\right)\left(\frac{\beta}{N}I_\infty\frac{1}{\lambda+\mu}\right) \\
&+ \left(\frac{\beta}{N}S_\infty\frac{1}{\lambda+\mu}\right)\left(\frac{\beta}{N}I_\infty\int_0^\infty \bar{F}(t)e^{-(\lambda+\mu)t}dt\right) \\
&= 1+\frac{\beta}{N}I_\infty\frac{1}{\lambda+\mu}-\frac{\beta}{N}S_\infty\int_0^\infty \bar{F}(t)e^{-(\lambda+\mu)t}dt
\end{aligned}
$$

Let $L(\lambda)=\beta\int_0^\infty \bar{F}(t)e^{-(\lambda+\mu)t}dt$, then the characteristic equation is rewritten as:

$$0=1+\frac{\frac{\beta}{N}I_\infty}{\lambda+\mu}-\frac{S_\infty}{N}L(\lambda)$$

At the endemic equilibrium $(S_\infty,I_\infty,R_\infty)$ the characteristic equation can be rewritten as:

$$\frac{\lambda+\mu\tilde{\mathscr{R}}_0}{\lambda+\mu}=\frac{L(\lambda)}{\tilde{\mathscr{R}}_0}$$

Assume that λ is of the form $a+ib$, with $a\geq 0$. Then

$$\frac{\lambda+\mu\tilde{\mathscr{R}}_0}{\lambda+\mu}=\frac{(a+\mu)(a+\mu\tilde{\mathscr{R}}_0)+b^2}{(a+\mu)^2+b^2}+i\frac{\mu b(1-\tilde{\mathscr{R}}_0)}{(a+\mu)^2+b^2}$$

Where if $\tilde{\mathscr{R}}_0>1$, then

$$\mathrm{Re}\Big(\frac{\lambda+\mu\tilde{\mathscr{R}}_0}{\lambda+\mu}\Big)=\frac{(a+\mu)(a+\mu\tilde{\mathscr{R}}_0)+b^2}{(a+\mu)^2+b^2}>1.$$

On the right hand side of the equation we have that

$$\begin{aligned}\frac{L(\lambda)}{\tilde{\mathscr{R}}_0} &= \frac{\beta\int_0^\infty \bar{F}(t)e^{-(a+ib+\mu)t}dt}{\tilde{\mathscr{R}}_0} \\ &= \frac{\beta\int_0^\infty \bar{F}(t)e^{-\mu t}e^{-at}e^{-ibt}dt}{\tilde{\mathscr{R}}_0} \\ &= \frac{\beta\int_0^\infty \bar{F}(t)e^{-\mu t}e^{-at}(\cos(bt)+i\sin(bt))dt}{\tilde{\mathscr{R}}_0}\end{aligned}$$

Where

$$\mathrm{Re}\Big(\frac{L(\lambda)}{\tilde{\mathscr{R}}_0}\Big) = \frac{\beta\int_0^\infty \bar{F}(t)e^{-\mu t}e^{-at}\cos(bt)dt}{\tilde{\mathscr{R}}_0} < \frac{\beta\int_0^\infty \bar{F}(t)e^{-\mu t}dt}{\tilde{\mathscr{R}}_0} = 1$$

Therefore λ is not a root of the characteristic equation at the endemic equilibrium if $\mathrm{Re}(\lambda) = a \geq 0$, which implies that the endemic equilibrium if locally asymptotecally stable if the epidemic threshold $\tilde{\mathscr{R}}_0$ is greater than 1.

Lastly, the basic reproduction number for the original age-of-infection SIR model in (2.4)-(2.6), without demographic terms is given by:

$$\mathscr{R}_0 = \lim_{\mu\to 0}\tilde{\mathscr{R}}_0 = \lim_{\mu\to 0}\beta\int_0^\infty \bar{F}(t)e^{-\mu t}dt = \beta\int_0^\infty \bar{F}(t)dt = \beta\int_{-\infty}^\infty t f(t)dt = \beta\tau.$$

APPENDIX G

DERIVATION OF THE FINAL SIZE RELATION AND FINAL SIZE FORMULA

Consider the age-of-infection SIR model in (2.4)-(2.6). The final size relation is derived as follows:

$$
\begin{aligned}
-\frac{\dot{S}(t)}{S(t)} &= \frac{\beta}{N} I(t) \\
-\dot{\ln}(S(t)) &= \frac{\beta}{N}\left(I_0 \bar{F}(t) + \int_0^t \frac{\beta}{N} I(s) S(s) \bar{F}(t-s) ds \right) \\
\int_0^\infty -\dot{\ln}(S(t)) dt &= \int_0^\infty \frac{\beta}{N}\left(I_0 \bar{F}(t) + \int_0^t \frac{\beta}{N} I(t-s) S(t-s) \bar{F}(s) ds \right) dt \\
-(\ln(S_\infty) - \ln(S_0)) &= \frac{I_0}{N}\left(\beta \int_0^\infty \bar{F}(t) dt \right) + \int_0^\infty \frac{\beta}{N} \int_0^t [-\dot{S}(t-s)] \bar{F}(s) ds dt \\
\ln\left(\frac{S_0}{S_\infty}\right) &= \frac{I_0}{N}\mathscr{R}_0 + \int_0^\infty \frac{\beta}{N} \bar{F}(s) \left(\int_s^\infty [-\dot{S}(t-s)] dt \right) ds \\
\ln\left(\frac{S_0}{S_\infty}\right) &= \frac{I_0}{N}\mathscr{R}_0 + \frac{(S_0 - S_\infty)}{N}\left(\beta \int_0^\infty \bar{F}(s) ds \right) \\
\ln\left(\frac{S_0}{S_\infty}\right) &= \mathscr{R}_0 \left(\frac{(S_0 + I_0) - S_\infty}{N} \right) \\
\ln\left(\frac{S_0}{S_\infty}\right) &= \mathscr{R}_0 \left(1 - \frac{S_\infty}{N} \right) \quad \text{(final size relation)}
\end{aligned}
$$

Let z be the cumulative incidence at the end of an epidemic, namely the final epidemic size. For the SIR model in (2.4)-(2.6) the final epidemic size can be defined as $z = S_0 - S_\infty$, then the final size formula can be derived in the following way:

$$\begin{aligned}
\ln\left(\frac{S_0}{S_\infty}\right) &= \mathscr{R}_0\left(1-\frac{S_\infty}{N}\right)\\
\ln\left(\frac{S_0}{S_0-z}\right) &= \mathscr{R}_0\left(1-\frac{S_0-z}{N}\right)\\
-\ln\left(\frac{S_0-z}{S_0}\right) &= \mathscr{R}_0\left(\frac{S_0+I_0-(S_0-z)}{N}\right)\\
\ln\left(1-\frac{z}{S_0}\right) &= -\mathscr{R}_0\left(\frac{I_0+z}{N}\right)\\
1-\frac{z}{S_0} &= exp\left(-\mathscr{R}_0\left(\frac{I_0+z}{N}\right)\right)\\
1-\frac{z}{S_0} &= exp\left(-\mathscr{R}_0\frac{I_0}{N}+\left(-\frac{\mathscr{R}_0}{N}\right)z\right)
\end{aligned}$$

Define the transformation:

$$z = -\frac{\left(t+\frac{\left(-\frac{\mathscr{R}_0}{N}\right)1}{\left(-\frac{1}{S_0}\right)}\right)}{\left(-\frac{\mathscr{R}_0}{N}\right)} = \frac{tN}{\mathscr{R}_0}+S_0$$

and substitute this into the final size formula to obtain an expression of the final epidemic size explicitly as a function of the Lambert W function:

$$\begin{aligned}
1-\frac{z}{S_0} &= exp\left(-\mathscr{R}_0\frac{I_0}{N}+\left(-\frac{\mathscr{R}_0}{N}\right)z\right)\\
1-\frac{\left(\frac{tN}{\mathscr{R}_0}+S_0\right)}{S_0} &= exp\left(-\mathscr{R}_0\frac{I_0}{N}+\left(-\frac{\mathscr{R}_0}{N}\right)\left(\frac{tN}{\mathscr{R}_0}+S_0\right)\right)\\
1-\frac{tN}{\mathscr{R}_0S_0}-1 &= \exp\left(-\mathscr{R}_0\frac{I_0}{N}-t-\left(\mathscr{R}_0\frac{S_0}{N}\right)\right)\\
-\frac{tN}{\mathscr{R}_0S_0} &= e^{(-\mathscr{R}_0-t)}\\
te^t &= -\frac{\mathscr{R}_0S_0}{N}e^{-\mathscr{R}_0} \equiv t = W\left(-\frac{\mathscr{R}_0S_0}{N}e^{-\mathscr{R}_0}\right)
\end{aligned}$$

$$
\begin{aligned}
t &= W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\right) \\
-\left(-\frac{\mathscr{R}_0}{N}\right) z - \left(\frac{\left(-\frac{\mathscr{R}_0}{N}\right) 1}{\left(-\frac{1}{S_0}\right)}\right) &= W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\right) \\
\left(\frac{\mathscr{R}_0}{N}\right) z - \left(\frac{\mathscr{R}_0}{N}\right) S_0 &= W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\right) \\
z &= S_0 + \left(\frac{N}{\mathscr{R}_0}\right) W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\right) \quad \text{(final size formula)}
\end{aligned}
$$

APPENDIX H

DERIVATION FOR THE SENSITIVITY EQUATIONS

Original system of VIE's:

$$\begin{aligned}
S(t) &= S_0 - \int_0^t \lambda(s)S(s)ds, \\
I(t) &= I_0\bar{F}(t) + \int_0^t \lambda(s)S(s)\bar{F}(t-s)ds, \\
R(t) &= R_0 + I_0(1-\bar{F}(t)) + \int_0^t \lambda(s)S(s)(1-\bar{F}(t-s))ds, \\
N(t) &= S(t)+I(t)+R(t) = S_0+I_0+R_0 = N_0, \\
&\quad S(0)=S_0, \quad I(0)=I_0, \quad R(0)=R_0,
\end{aligned}$$

Sensitivity equations as VIE's:

$$\begin{aligned}
\frac{\partial S}{\partial \beta}(t) &= -\int_0^t \Big(\frac{1}{N}I(s)S(s) + \frac{\beta}{N}I(s)\frac{\partial S}{\partial \beta}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \beta}(s)S(s)\Big)ds, \\
\frac{\partial I}{\partial \beta}(t) &= \int_0^t \Big(\frac{1}{N}I(s)S(s) + \frac{\beta}{N}I(s)\frac{\partial S}{\partial \beta}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \beta}(s)S(s)\Big)\bar{F}(t-s)ds, \\
\frac{\partial R}{\partial \beta}(t) &= -\Big(\frac{\partial S}{\partial \beta}(t) + \frac{\partial I}{\partial \beta}(t)\Big) \\
\frac{\partial S}{\partial \tau}(t) &= -\int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial \tau}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \tau}(s)S(s)\Big)ds, \\
\frac{\partial I}{\partial \tau}(t) &= I_0\frac{\partial \bar{F}}{\partial \tau}(t) + \int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial \tau}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \tau}(s)S(s)\Big)\bar{F}(t-s)ds, \\
&\qquad + \int_0^t \frac{\beta}{N}I(s)S(s)\frac{\partial \bar{F}}{\partial \tau}(t-s)ds \\
\frac{\partial R}{\partial \tau}(t) &= -\Big(\frac{\partial S}{\partial \tau}(t) + \frac{\partial I}{\partial \tau}(t)\Big) \\
\frac{\partial S}{\partial Var}(t) &= -\int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial Var}(s) + \frac{\beta}{N}\frac{\partial I}{\partial Var}(s)S(s)\Big)ds, \\
\frac{\partial I}{\partial Var}(t) &= I_0\frac{\partial \bar{F}}{\partial Var}(t) + \int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial Var}(s) + \frac{\beta}{N}\frac{\partial I}{\partial Var}(s)S(s)\Big)\bar{F}(t-s)ds, \\
&\qquad + \int_0^t \frac{\beta}{N}I(s)S(s)\frac{\partial \bar{F}}{\partial Var}(t-s)ds \\
\frac{\partial R}{\partial Var}(t) &= -\Big(\frac{\partial S}{\partial Var}(t) + \frac{\partial I}{\partial Var}(t)\Big)
\end{aligned}$$

$$\frac{\partial S}{\partial \beta}(0) = \frac{\partial I}{\partial \beta}(0) = \frac{\partial R}{\partial \beta}(0) = \frac{\partial S}{\partial \tau}(0) = \frac{\partial I}{\partial \tau}(0) = \frac{\partial R}{\partial \tau}(0) = \frac{\partial S}{\partial Var}(0) = \frac{\partial I}{\partial Var}(0) = \frac{\partial R}{\partial Var}(0) = 0$$

Sensitivity equations as VIODE's: Lets use the notation: $X_y = \frac{\partial X}{\partial y}$

$$
\begin{aligned}
\frac{dS_\beta}{dt}(t) &= -\Big(\frac{1}{N}I(t)S(t) + \frac{\beta}{N}I(t)\frac{\partial S}{\partial \beta}(t) + \frac{\beta}{N}\frac{\partial I}{\partial \beta}(t)S(t)\Big), \\
\frac{dI_\beta}{dt}(t) &= \frac{1}{N}I(t)S(t) + \frac{\beta}{N}I(t)\frac{\partial S}{\partial \beta}(t) + \frac{\beta}{N}\frac{\partial I}{\partial \beta}(t)S(t) \\
&\quad - \int_0^t \Big(\frac{1}{N}I(s)S(s) + \frac{\beta}{N}I(s)\frac{\partial S}{\partial \beta}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \beta}(s)S(s)\Big)[-\dot{\bar{F}}(t-s)]ds, \\
\frac{dR_\beta}{dt}(t) &= -\Big(\frac{dS_\beta}{dt} + \frac{dI_\beta}{dt}\Big) \\
\frac{dS_\tau}{dt}(t) &= -\Big(\frac{\beta}{N}I(t)\frac{\partial S}{\partial \tau}(t) + \frac{\beta}{N}\frac{\partial I}{\partial \tau}(t)S(t)\Big), \\
\frac{dI_\tau}{dt}(t) &= \frac{\beta}{N}I(t)\frac{\partial S}{\partial \tau}(t) + \frac{\beta}{N}\frac{\partial I}{\partial \tau}(t)S(t) + \frac{\beta}{N}I(t)S(t)\frac{\partial \bar{F}}{\partial \tau}(0) - \Bigg(I_0\frac{\partial[-\dot{\bar{F}}]}{\partial \tau}(t) \\
&\quad + \int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial \tau}(s) + \frac{\beta}{N}\frac{\partial I}{\partial \tau}(s)S(s)\Big)[-\dot{\bar{F}}(t-s)]ds + \int_0^t \frac{\beta}{N}I(s)S(s)\frac{\partial[-\dot{\bar{F}}]}{\partial \tau}(t-s)ds\Bigg) \\
\frac{dR_\tau}{dt}(t) &= -\Big(\frac{dS_\tau}{dt} + \frac{dI_\tau}{dt}\Big) \\
\frac{dS_{Var}}{dt}(t) &= -\Big(\frac{\beta}{N}I(t)\frac{\partial S}{\partial Var}(t) + \frac{\beta}{N}\frac{\partial I}{\partial Var}(t)S(t)\Big), \\
\frac{dI_{Var}}{dt}(t) &= \frac{\beta}{N}I(t)\frac{\partial S}{\partial Var}(t) + \frac{\beta}{N}\frac{\partial I}{\partial Var}(t)S(t) + \frac{\beta}{N}I(t)S(t)\frac{\partial \bar{F}}{\partial Var}(0) \\
&\quad - \Bigg(I_0\frac{\partial[-\dot{\bar{F}}]}{\partial Var}(t) + \int_0^t \Big(\frac{\beta}{N}I(s)\frac{\partial S}{\partial Var}(s) + \frac{\beta}{N}\frac{\partial I}{\partial Var}(s)S(s)\Big)[-\dot{\bar{F}}(t-s)]ds \\
&\quad + \int_0^t \frac{\beta}{N}I(s)S(s)\frac{\partial[-\dot{\bar{F}}]}{\partial Var}(t-s)ds\Bigg) \\
\frac{dR_{Var}}{dt}(t) &= -\Big(\frac{dS_{Var}}{dt} + \frac{dI_{Var}}{dt}\Big) \\
\frac{\partial S}{\partial \beta}(0) &= \frac{\partial I}{\partial \beta}(0) = \frac{\partial R}{\partial \beta}(0) = \frac{\partial S}{\partial \tau}(0) = \frac{\partial I}{\partial \tau}(0) = \frac{\partial R}{\partial \tau}(0) = \frac{\partial S}{\partial Var}(0) = \frac{\partial I}{\partial Var}(0) = \frac{\partial R}{\partial Var}(0) = 0
\end{aligned}
$$

Lets consider the case when the probability density function has the form of a gamma distribution function $[-\dot{\bar{F}}(t;k,\theta)]$, where $k = \frac{\tau^2}{Var}$, the shape parameter and the scale parameter is $\theta = \frac{Var}{\tau}$, where *tau* and *Var* are the mean and variance of the gamma distribution for the infectious period: then

$$[-\dot{\bar{F}}(t;k,\theta)] = \Big[-\dot{\bar{F}}\Big(t;\frac{\tau^2}{Var},\frac{Var}{\tau}\Big)\Big] = f\Big(t;\frac{\tau^2}{Var},\frac{Var}{\tau}\Big) = \begin{cases} \frac{1}{\Gamma(k)\theta^k}t^{k-1}e^{-\frac{t}{\theta}} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0, \end{cases}$$

$$\begin{aligned}
\frac{\partial[-\dot{\bar{F}}(t;k(\tau,Var),\theta(\tau,Var))]}{\partial\tau} &= \frac{\partial f\Big(t;k(\tau,Var),\theta(\tau,Var)\Big)}{\partial\tau} \\
&= \frac{\partial f}{\partial k}\frac{\partial k}{\partial\tau} + \frac{\partial f}{\partial\theta}\frac{\partial\theta}{\partial\tau} \\
&= \frac{\partial f}{\partial k}\Big(\frac{2\tau}{Var}\Big) + \frac{\partial f}{\partial\theta}\Big(-\frac{Var}{\tau^2}\Big) \\
\frac{\partial[-\dot{\bar{F}}(t;k(\tau,Var),\theta(\tau,Var))]}{\partial Var} &= \frac{\partial f\Big(t;k(\tau,Var),\theta(\tau,Var)\Big)}{\partial Var} \\
&= \frac{\partial f}{\partial k}\frac{\partial k}{\partial Var} + \frac{\partial f}{\partial\theta}\frac{\partial\theta}{\partial Var} \\
&= \frac{\partial f}{\partial k}\Big(-\frac{\tau^2}{Var^2}\Big) + \frac{\partial f}{\partial\theta}\Big(\frac{1}{\tau}\Big)
\end{aligned}$$

$$
\begin{aligned}
\text{Where, } \frac{\partial f}{\partial k} &= \frac{\partial}{\partial k}\left(\frac{1}{\theta^k}\frac{1}{\Gamma(k)}t^{k-1}e^{-\frac{t}{\theta}}\right) = \frac{\partial}{\partial k}\left(e^{-\frac{t}{\theta}}t^{-1}\left(\frac{t}{\theta}\right)^k\frac{1}{\Gamma(k)}\right) = e^{-\frac{t}{\theta}}t^{-1}\frac{\partial}{\partial k}\left(\left(\frac{t}{\theta}\right)^k\frac{1}{\Gamma(k)}\right) \\
&= e^{-\frac{t}{\theta}}t^{-1}\left(\frac{\partial}{\partial k}\left(\left(\frac{t}{\theta}\right)^k\right)\frac{1}{\Gamma(k)} + \left(\frac{t}{\theta}\right)^k\frac{d}{dk}\left(\Gamma(k)\right)^{-1}\right) \\
&= e^{-\frac{t}{\theta}}t^{-1}\left(\left(\ln\left(\frac{t}{\theta}\right)\left(\frac{t}{\theta}\right)^k\right)\frac{1}{\Gamma(k)} + \left(\frac{t}{\theta}\right)^k\left(-\Gamma(k)\right)^{-2}\frac{d}{dk}\Gamma(k)\right) \\
&= \left(\frac{1}{\theta^k}\frac{1}{\Gamma(k)}t^{k-1}e^{-\frac{t}{\theta}}\right)\left(\ln\left(\frac{t}{\theta}\right) - \frac{\Gamma(k)\psi_0(k)}{\Gamma(k)}\right) = f(t;k,\theta)\left(\ln\left(\frac{t}{\theta}\right) - \psi_0(k)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(\frac{1}{\theta^n}\frac{1}{\Gamma(k)}t^{k-1}e^{-\frac{t}{\theta}}\right) = \frac{t^{k-1}}{\Gamma(k)}\frac{\partial}{\partial \theta}\left(\frac{e^{-\frac{t}{\theta}}}{\theta^k}\right) = \frac{t^{k-1}}{\Gamma(k)}\frac{\partial}{\partial \theta}\left(e^{-\frac{t}{\theta}}\theta^{-k}\right) \\
&= \frac{t^{k-1}}{\Gamma(k)}\left(e^{-\frac{t}{\theta}}\left(-k\theta^{-k-1}\right) + \left(\frac{te^{-\frac{t}{\theta}}}{\theta^2}\right)\theta^k\right) = \left(\frac{1}{\theta^k}\frac{1}{\Gamma(k)}t^{k-1}e^{-\frac{t}{\theta}}\right)\left(\frac{t}{\theta^2} - \frac{k}{\theta}\right) \\
&= f(t;k,\theta)\left(\frac{t}{\theta^2} - \frac{k}{\theta}\right)
\end{aligned}
$$

lastly we have:

$$
\begin{aligned}
\frac{\partial f\left(t;k(\tau,Var),\theta(\tau,Var)\right)}{\partial \tau} &= f(t;k,\theta)\left(\ln\left(\frac{t}{\theta}\right) - \psi_0(k)\right)\left(\frac{2\tau}{Var}\right) \\
&\quad + f(t;k,\theta)\left(\frac{t}{\theta^2} - \frac{k}{\theta}\right)\left(-\frac{Var}{\tau^2}\right) \\
&= \frac{1}{Var}f(t;k,\theta)\left[\left(\ln\left(\frac{t\tau}{Var}\right) - \psi_0(\frac{\tau^2}{Var}) - \frac{1}{2}\right)2\tau + t\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f\left(t;k(\tau,Var),\theta(\tau,Var)\right)}{\partial Var} &= f(t;k,\theta)\left(\ln\left(\frac{t}{\theta}\right) - \psi_0(k)\right)\left(-\frac{\tau^2}{Var^2}\right) \\
&\quad + f(t;k,\theta)\left(\frac{t}{\theta^2} - \frac{k}{\theta}\right)\left(\frac{1}{\tau}\right) \\
&= \frac{\tau}{Var^2}f(t;k,\theta)\left[\left(\psi_0(\frac{\tau^2}{Var}) - \ln\left(\frac{t\tau}{Var}\right) - 1\right)\tau + t\right]
\end{aligned}
$$

APPENDIX I

DERIVATION OF SENSITIVITY EQUATION FOR THE FINAL EPIDEMIC SIZE

From **appendix N** the final epidemic size is given by

$$z = S_0 + \Big(\frac{N}{\mathscr{R}_0}\Big) W\left(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0} \right).$$

Derivation of the sensitivity equation of the final epidemic size with respect to the basic reproduction number:

$$\begin{aligned}
z &= S_0 + \Big(\frac{N}{\mathscr{R}_0}\Big) W\left(\underbrace{-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}}_{=Y} \right) && \text{(I.1)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= -\Big(\frac{N}{R_0^2}\Big) W(Y) + \Big(\frac{N}{\mathscr{R}_0}\Big) \frac{dW}{dY} \frac{\partial Y}{\partial \mathscr{R}_0} && \text{(I.2)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= -\Big(\frac{N}{R_0^2}\Big) W(Y) + \Big(\frac{N}{\mathscr{R}_0}\Big) \Big(\frac{W(Y)}{Y(1+W(Y))}\Big) \Big(-\frac{S_0}{N} e^{-\mathscr{R}_0} + \frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0} \Big) && \text{(I.3)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= S_0 e^{-\mathscr{R}_0} \Big(1 - \frac{1}{\mathscr{R}_0}\Big) \Big(\frac{W(Y)}{Y(1+W(Y))}\Big) - \Big(\frac{N}{R_0^2}\Big) W(Y) && \text{(I.4)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= S_0 e^{-\mathscr{R}_0} \Big(1 - \frac{1}{\mathscr{R}_0}\Big) \left(\frac{W(Y)}{\big(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}\big)(1+W(Y))} \right) - \Big(\frac{N}{R_0^2}\Big) W(Y) && \text{(I.5)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= \frac{N}{\mathscr{R}_0} \Big(\frac{1}{\mathscr{R}_0} - 1\Big) \Big(\frac{W(Y)}{1+W(Y)}\Big) - \Big(\frac{N}{R_0^2}\Big) W(Y) && \text{(I.6)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= -\frac{N}{\mathscr{R}_0} \left(\frac{W(Y)(W(Y)+\mathscr{R}_0)}{\mathscr{R}_0(1+W(Y))} \right) && \text{(I.7)}\\
\frac{\partial z}{\partial \mathscr{R}_0} &= -\frac{N}{\mathscr{R}_0} \left(\frac{W(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0})(W(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}) + \mathscr{R}_0)}{\mathscr{R}_0(1+W(-\frac{\mathscr{R}_0 S_0}{N} e^{-\mathscr{R}_0}))} \right) && \text{(I.8)}
\end{aligned}$$