University of Amsterdam

Assignment 1: Report

The SIR Model and Its Variants

October 3, 2023

Tutor:
Cillian Hourican

Group:
A

Student:
Boyan Mihaylov
UvAnetID 15067602

Lecturer:
Michael Lees

Course:
Introduction to Computational Science

1 Abstract

This report presents a series of experiments with mathematical models of infectious diseases studied during the course "Introduction to Computational Science". The inquiry in the dynamics of disease spread begins with basic abstractions of the mechanisms behind pathogen transmission fed into a simulation network. The prototypical model is then gradually enhanced to implement more of the actual mechanisms driving such dynamics, providing detailed analysis of the relevant parameters, the convergence points of the emerging systems and the dynamic behaviour around them. Throughout the process, valuable insights are gained about the applicability of such models and of the information they provide concerning real-life scenarios.

2 Introduction

Infectious diseases are phenomena of ecosystemic relevance inextricably affecting aspects of human society as well as the balance of broader biological networks.[1, 2] Their emergence as complex systems that develop on many different scales, driven by a multitude of factors, necessitates an appropriate scientific approach that can capture the intricacies of their dynamics through a reliable and extensive framework of models. As a focus subject of the introductory course in Computational Science the process of modelling infectious diseases presents a set of mathematical methods and a mode of scientific thinking applicable to many other areas of similar complexity.[3]

This report documents the incorporation of newly acquired knowledge about infectious diseases and dynamic systems in the handling of specific problems outlined by an assignment brief. The experiments begin with more simplistic computational models and proceed to gradually expand these, delivering insights about different aspects of the systems in question. The results from these experiments are then analysed and in many cases interpreted with regard to their practical relevance and real-life representation.

The first set of studies constructs a basic simulation framework around the SIR model without demography and conducts several tests to examine the behaviour under different parameter settings. The predictive power of the model is then tested by fitting its parameters to approximate epidemiological data from a historical influenza outbreak at a boys school. The use

of infection models for policy making is then illustrated in the proposal of several vaccination strategies which could prevent the spread of the disease in a hypothetical scenario. The model framework is then extended to include demographic phenomena such as birth and death. Specific parameter settings of this model are observed to produce interesting effects, including an endemic state of infection and an oscillatory behaviour. The conditions and characteristics of these effects are examined empirically through the simulation and are compared to analytical predictions derived from literature. A further addition to the model is a mortality rate induced by the disease itself. The influence of its magnitude is studied an another set of simulations. Finally, a more complex extension of the SIR model is examined, in which the transmission does not occur via the members of the same population but rather through animal vectors such as mosquitoes, which is characteristic for diseases like malaria. An attempt is made to align this multi-host model even further with real-life observations by introducing seasonal variations in the form of temporal forcing.

The book "Modeling Infectious Diseases in Humans and Animals" by Keeling and Rohani[4] serves as a main resource of information about the methods and mathematical derivations used in the modelling process. Furthermore, the book "Nonlinear Dynamics and Chaos" by Steven Strogatz[3] is used as a guide in understanding the dynamics of the systems by studying the trajectories of population variables while quantifying the fixed points and their respective stability.

3 Background

Compared to other diseases, microparasitic, directly transmitted infectious diseases exhibit specific layers of complexity due to the mechanisms of pathogen transmission through interacting hosts, a system closely interlinked with population dynamics. As microparasites have a simple anatomy and rapidly growing numbers upon infection, they can be idealized in mathematical models with substantial predictive power that can capture essential patterns of epidemic spread. Such models are the primary focus of the book by Keeling et al[4] and of the presented experiments.

3.1 The naïve SIR model

The fundamental prototype used for the models in the experements dates back to a paper by Kermack and McKendrick from 1927[5]. In it, a model of the population is constructed which comprises three compartments - susceptible (S), infected (I) and recovered (R). These form the so-called SIR model, in which individuals can move from one compartment to another based on formulated rates of change that are dependent on different parameters. The three categories can be viewed as temporal segments over the timeline of a disease developing in a human:

- Susceptible individuals are in a pre-exposure state, they have no immunity and can get infected by infected ones.
- Infected individuals have been exposed to pathogens and are actively infectious themselves.
- Recovered individuals have healed from the disease and are assumed to have acquired immunity against further infection.

Extensions such as the SEIR model may also include the "exposed" category, which indicates the stage when the pathogen is present in an individual but is not abundant enough to cause spread to other hosts.

The main compartments can either be represented as fractions of the population N, in which case $S + I + R = 1$, or as the actual number of individuals, in which case commonly X denotes susceptible, Y - infected, and Z - recovered, and $X + Y + Z = N$. The population size can play a different role depending on the classification of the model as frequency-dependent or density-dependent. Frequency-dependent models assume that the population size does not affect the number of contacts per person, while density-dependent ones take into consideration systems in which changing density under spatial constraints alters the contact rate. Studies[6] have found that frequency-dependent transmission is well applicable to human societies, while

density-dependent systems are more likely to represent animal diseases within confined ecosystems.

In the most simplistic version of the SIR model, the interactions between the S, I and R compartments are represented by a system of ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\beta SI \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

The two main parameters determining the flux between different compartments are the infectious transmission rate β and the recovery rate γ. β is calculated as a function of the average number of contacts κ and the probability c of infection given contact with an infected individual ($\beta = -\kappa ln(1-c)$). It can therefore incorporate many different aspects that affect the transmission of a disease, such as the biological mechanisms of the pathogen, the spatial and topological structure of the society it is transmitted in and the behavioral patterns of the hosts. γ represents the inverse of the average period in which a host can infect others before moving to the recovered compartment, measured in days.

These two parameters are based on epidemiological data and can be combined together in the disease-specific indicator R_0 called the basic reproductive ratio, calculated as:

$$R_0 = \frac{\beta}{\gamma} \tag{4}$$

This ratio defined by Diekmann et al as the "expected number of secondary cases per primary case"[7] in an entirely susceptible population is of essential importance for determining whether a certain disease would spread in an epidemic fashion or would die out quickly. While different extensions of the SIR model formulate the fraction using a different set of parameters, it can be observed that for values of $R_0 > 1$ the numbers of infected would experience growth, whereas for values smaller than one the disease would dissipate without spreading.

3.2 The SIR model with demography

One aspect that the simplistic SIR model does not account for is demography - it assumes a fixed number of susceptibles within the time frame of the disease and therefore guarantees that the entire population would eventually transition to the recovered compartment. While this might be a good approximation for short-term epidemics, infectious diseases acting on a larger time scale have to account for the influx of new population members, e.g. through birth, as well as the removal of population members by causes like natural death. These phenomena can be implemented in the model by extending the original SIR differential equations with several demography-related terms:

$$\frac{dS}{dt} = \mu - \beta SI - \mu S \tag{5}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I \tag{6}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{7}$$

Assuming a stable population size in which the number of deaths is balanced by an equal amount of births, the parameter μ is added to represent both of these effects. Since the general assumption here is that individuals are not born infectious or with an inherent immunity, μ is only added with a positive magnitude to the susceptible compartment. On the other hand, each compartment loses members to causes unrelated to the disease with the same rate, hence the negative term with factor μ in each equation. The basic reproductive ratio then becomes:

$$R_0 = \frac{\beta}{\gamma + \mu} \tag{8}$$

It has been observed that such an extension may cause the model to exhibit an equilibrium state other than the disease-free condition under specific parameter settings, with the population fractions approaching this state in an oscillating manner.[4] Such observations are aligned with the understanding that diseases may embrace an endemic status, at which there is always a certain amount of infected individuals in the population.

Equilibrium states, also called fixed points, are a common feature of dynamic systems[3] that indicate the values that variables approach as time approaches infinity. They can be analytically quantified by observing that the rate of change at these points is zero, and can exhibit two types of behaviour depending on the stability of small perturbations around them. It is said that the fixed point is stable when all values in its vicinity are attracted to it, and unstable when these tend to diverge from it. Taking the Jacobian matrix of the differential equations at the fixed point one can perform eigenanalysis[8] to determine the specifics of the system's movement around it. The following cases exist:

- Stable nodes - both eigenvalues are negative and real, the trajectories approach the point with a speed determined by the magnitude of the eigenvectors;
- Unstable nodes - both eigenvalues are positive and real, the trajectories diverge from the point with a speed determined by the magnitude of the eigenvectors;
- Stable spirals - both eigenvalues are negative and contain a complex term, the trajectories approach the point in a spiraling motion;
- Unstable spirals - both eigenvalues are positive and contain a complex term, the trajectories diverge from the point in a spiraling motion;
- Saddles - the eigenvalues differ in sign, the trajectories approach one of the eigenvectors asymptotically for $t \leftarrow \infty$ and the other for $t \leftarrow -\infty$
- Centers - both eigenvalues are purely complex, the trajectories orbit around the point periodically without being attracted or repelled by it;

Furthermore, degenerate cases exist for the situations when the eigenvectors are equal, both zero or when there is only one eigenvector. Typically for the SIR model with demography an endemic state appears for $R_0 > 1$ at:

$$(S^*, I^*, R^*) = (\frac{1}{R_0}, \frac{\mu}{\beta}(R_0 - 1), 1 - \frac{1}{R_0} - \frac{\mu}{\beta}(R_0 - 1)) \tag{9}$$

The analysis of the eigenvalues at the fixed point leads to solutions containing complex terms, from which we can conclude that the point is a stable spiral:

$$\lambda \approx -\frac{\mu R_0}{2} \pm \frac{i}{\sqrt{AG}} \tag{10}$$

Here $A = \frac{1}{\mu(R_0-1)}$ represents the mean age at infection and $G = \frac{1}{\mu+\gamma}$ denotes the typical infectious period of a host. It has been shown[3] that the general solution of a dynamical system can be expressed as a function of the eigenvectors $\vec{v}$ and the eigenvalues λ:

$$\vec{x}(t) = c_1 e^{\lambda_1 t}\vec{v}_1 + c_2 e^{\lambda_2 t}\vec{v}_2 \tag{11}$$

With λ being complex, this means that the equation involves terms of the form $e^{(\alpha+i\omega)t}$. Euler's formula states that:

$$e^{i\omega t} = cos(\omega t) + isin(\omega t) \tag{12}$$

which suggests that the factor multiplying the imaginary part of the eigenvalue, in our case $\omega = \frac{1}{\sqrt{AG}}$, determines the frequency of the oscillations. If one full cycle around the fixed point is denoted by an angular period of 2π, then the period of the oscillations is:

$$T = 2\pi\sqrt{AG} \tag{13}$$

its inverse determining the frequency $\frac{1}{2\pi\sqrt{AG}}$. The real part of the eigenvalue can be interpreted as the decay rate of the amplitude. Since $-\frac{\mu R_0}{2}$ is always negative, there would always be a decreasing amplitude and thus stable attraction towards the equilibrium point. The spiraling motion can be observed on a phase plot and essentially means that the numbers of S and I undergo damped oscillation with a phase offset to each other.

While the demography parameters assume that mortality is a phenomenon that occurs independently of the disease and removes individuals from the compartments at an equal rate, in the cases of a deadly pathogen there is an infection-induced mortality rate to be considered. The probability ρ that an infected person dies from illness before recovering or dying of a natural cause can be included in the differential equation for I so that:

$$\frac{dI}{dt} = \beta SI - \frac{(\gamma + \mu)}{1 - \rho} I \tag{14}$$

This additional removal factor has the effect that the population size cannot be balanced to a constant size by a birth rate identical to μ. To avoid extinction of the population in the simulation, a birth rate v van be introduced to the S term, which is decoupled from μ and can have a correcting effect:

$$\frac{dS}{dt} = v - \beta SI - \mu S \tag{15}$$

The new parameters redefine the reproductive ratio as:

$$R_0 = \frac{\beta(1 - \rho)v}{(\mu + \gamma)\mu} \tag{16}$$

A calculation of the endemic state (X^*, Y^*, Z^*) yields:

$$X^* = \frac{v}{\mu R_0} \tag{17}$$

$$Y^* = \frac{\mu}{\beta}(R_0 - 1) \tag{18}$$

$$Z^* = \frac{\gamma}{\beta}(R_0 - 1) \tag{19}$$

With the variation of the population size becoming inevitable, the formulation of the model as frequency-dependent versus density-dependent can have a substantial effect on the equilibrium points, as a density-dependent mechanism implies that the decrease in population reduces infectious contact.

3.3 Multi-host model

There are many extensions and variants of the standard SIR model - some introduce an additional compartment to the system or operate with one less, others explore more intricate interactions between the fractions of the population. One family of models that exhibits particular complexity is the multi-host model which can represent vector-transmitted diseases like malaria and yellow fever. The special feature of such models is that the disease cannot be transmitted between the members of the same species but relies on cross-species interaction, implying that two distinct populations are involved in the pathogenic spread.

In the case of malaria, an infected mosquito can transmit spread the infection to susceptible humans through blood meals. The opposite is also true - a susceptible mosquito can become infected by biting an infected human host. The typical β parameter dependent on human societal interactions is replaced here with a term involving b, which is the number of mosquito bites per unit time, and the probabilities T_{HM} and T_{MH} that an infected mosquito bite transfers the disease to a human and vice versa. The bite rate can be related to the human population size N_H by setting $r = \frac{b}{N_H}$ as the rate at which a particular human is bitten by a particular mosquito. These parameters form the transmission matrix:

$$\beta = \begin{pmatrix} 0 & rT_{HM} \\ rT_{HM} & 0 \end{pmatrix} \tag{20}$$

The zeros along the diagonals indicate that transmission is possible only between species and not internally. Taking these new expressions, the system of ODEs becomes:

$$\frac{dX_H}{dt} = v_H - rT_{HM}Y_M X_H - \mu_H X_H \tag{21}$$

$$\frac{dY_H}{dt} = rT_{HM}Y_M X_H - \mu_H Y_H - \gamma_H Y_H \tag{22}$$

$$\frac{dX_M}{dt} = v_M - rT_{MH}Y_H X_M - \mu_M X_M \tag{23}$$

$$\frac{dY_M}{dt} = rT_{MH}Y_H X_M - \mu_M Y_M \tag{24}$$

The subscripts indicate the species that the corresponding parameter applies to (H - human, M - mosquito). The $\frac{dZ}{dt}$ terms are analogous to the standard SIR model with demography (Eq. 7). It should be noted that the transmission is greatly dependent on the population sizes, so the actual population numbers X, Y and Z are used instead of their corresponding proportions S, I and R. The relationship between the population sizes and the epidemic reproduction becomes evident in the formula for R_0:

$$R_0 = \frac{b^2 T_{HM}T_{MH}N_M}{\mu_M(\gamma_H + \mu_H)N_H} \tag{25}$$

Seasonal variations and other forms of periodic cycles have been recognized as an important factor affecting the recurrence of infection peaks, most notably by Soper (1929)[9]. The relationship between the frequency of these periodic undulations and the natural frequency of the epidemiological model can have a dramatic effect on the amplification of infection numbers due to harmonic or sub-harmonic resonance. Many of the existing SIR models and their variations can be extended to implement temporal forcing, enabling more sophisticated studies of the resonance effects. In the more standard cases this implies transforming the β parameter into a cosine term with a baseline magnitude and an oscillation amplitude, reflecting periodic variations in human interactions (school periods, holidays). If the amplitude of the oscillating parameter gradually exceeds certain threshold values, the equilibrium point for a certain population compartment can undergo bifurcations, splitting into multiple equilibria until it dissipates into a chaotic behaviour with very low predictability.

In the case of vectored systems the transmission term is more complex but periodic variations can be found in different parameters nonetheless. Such is the birth rate of mosquitoes, which can vary greatly depending on climate fluctuations. The v_M term in the above equations can therefore be expressed as:

$$v_M(t) = v_{M0}(1 + v_{M1}cos(\omega t)) \tag{26}$$

in which v_0 is the average birth rage (which can be set to μN_M for the purpose of stabilizing the population size) and v_1 represents the amplitude of the birth rate variation.

3.4 Vaccination

Seen in the context of an SIR model, vaccination presents a means to automatically remove individuals from the susceptible section, thus making it more difficult for a disease to spread among hosts. The effects of different vaccination strategies can therefore be tested by simulating such a mechanism in the available models. Depending on the complexity of the model and the assumptions about the efficiency and time-scale of the vaccinations, simulated interventions can provide insights about the change in infection patterns. The interference of immunisation with the oscillatory behaviour of epidemic systems has not only been observed in silico but has also been also reinforced by historical data as in the studies of Earn et al (2000)[10].

4 Methods

4.1 Programming

The framework for the experimental studies documented in this report is programmed in Python due to its ease of use and the abundance of related libraries that support scientific computing[11]. One extensively used package throughout the experiments is NumPy[12], mostly for facilitating the management of data arrays. Furthermore, specific methods from the library SciPy are used, such as the *curve_fit* and *minimize_scalar* algorithms and several methods for Fourier analysis. The library Matplotlib provides all the methods for representing numerical results in different types of plots. The code is executed in a Jupyter Notebook, supported by notations in Markdown and LaTeXas a way to document the experimentation process together with the used programming methods.

As the studied models are heavily reliant on the numerical integration of ODEs, a fourth-order Runge-Kutta method[13] is implemented in a broadly used function as the default integration method. The choice was made due to its increased precision in comparison to faster but simpler methods such as the Euler forward method.[3] An additional function that repeatedly comes in use during the first experiments is a method for graphing the phase plots of S versus I with different initial values.

The Python language is particularly suitable for object-oriented programming (OOP)[14]. It is therefore attempted to use this programming paradigm as much as possible to provide clearly structured code and a hierarchical relationship between objects involved in the simulations. Typically, for each iteration of the epidemiological model a class is created that contains all the features and mechanisms of the corresponding model. Whenever the differences between two model iterations are small enough, the new model class is created as a subclass of the more general model, integrating the additional parameters and methods. A typical infection model class contains:

- an initialisation function *__init__*;
- a function containing the fundamental ODEs;
- a function performing numerical integration of the system over a specified time-frame;
- other functions specific to the current model, including special updates of parameters in the case of temporal forcing.

The initialisation arguments of the class present the option to either express parameters such as β, γ and μ through their respective lower-level components or supply the pre-calculated rates directly.

In the case of the multi-host system it could be formalised that we have two interacting systems embedded in the system of the model. This is represented by a hierarchical structure of the model class, containing in itself a nested class representing species. This class can be instantiated for each of the species involved in the transmission and store their corresponding parameters.

For special cases when the integration function of a class instance needs to be run only with specific parameters (for instance during the fitting of β and γ against the historical data from the influenza case) a "tester" class is created that can store the constant model parameters, instantiate a model with only the variable ones provided, run it over a given time-frame and sample the I results at time points corresponding to the times of the recorded influenza cases. The *curve_fit* method used to perform the parameter fitting of the prototypical SIR model uses the Trust Region Reflective algorithm (trf) in order to minimize the error between the curve of traced I values and the data points. In this process it searches for the optimal β and γ within an interval of $[0, 100]$. The data used for the fitting procedure of the SIR model is taken from the table in the assignment brief.

To support the analytical calculations in the analysis of the oscillation frequencies of the SIR model with demography, a Fourier analysis is implemented using Fast Fourier Transform[15] methods of the SciPy library.

4.2 Experiment Setting

The general procedure for the experiments is as follows:

1. Follow the sequential order of the problems as listed in the assignment;
2. Interpret each problem by sourcing information from the book by Keeling et al[4] or from the course curriculum;
3. Construct a model or a procedure that can address the task;
4. Test the model on the specific task and plot graphs or print strings containing the results;
5. Interpret the results with reference to real-life observations or validate the results through comparison with similar outputs from literature.

The parameters used in the different models are chosen following different modes of reasoning. In most of the cases the goal is to express β and γ through their constituent sub-parameters which are closer to concrete real-life measurements. For the initial tests of the naïve SIR model these are the number of contacts per person κ, the probability of infection via infectious contact c and the infectious period $1/\gamma$. The values for these are arbitrarily set to represent a fictitious medium-contagious, fast-acting disease. For the initial SIR model with demography the infection probability c is increased for visualisation purposes in order to make the oscillations more distinct. The demographic parameter μ is expressed as the inverse of an expected human lifespan. This life expectancy is set to the exemplary value of 44.1 years, which is the median age in the EU for 2022 as reported by Eurostat. Some of the initial parameters are tuned as part the experiments and change in magnitude for the subsequent iterations. The heuristically determined starting parameter values are listed in Table 1. When the decoupled birth rate v is introduced in the model with infection-induced mortality, it is kept similar to the death rate μ in order to keep the size of the population relatively stable.

Model	κ	c	P_{inf}	μ
SIR without demography	8	0.1	7	16107.525
SIR with demography	8	0.5	7	16107.525

Table 1: Initial parameters for the model class.

For the multi-host system most of the characteristic parameters were taken from Figure 4.13. of the book by Keeling et al (Fig. 2). It is acknowledged that a more precise study of the parameters of malaria models exists in the work of Chitnis et al (2008)[16] but they are used in the context of a differently formulated model, so the former values are kept for better comparison with the graphic results by Keeling.

Parameter	Value for Humans	Value for Mosquitoes
$T_{SelfOther}$	0.5	0.8
Population Size	10e4	10e5
Lifespan (Days)	16107.525	7
γ	0.033	0
Y_{start}	0	1

Table 2: Initial parameters for the multi-host model.

For all models except for the multi-host system the starting population size is set to 1000. The human population in the vectored transmission model is also set to 1000 but the starting population of mosquitoes is set to be ten times larger. The general procedure for creating phase space plots involves varying the initial numbers of two of the population compartments, usually increasing one and decreasing the other while keeping their sum constantly equal to the initial population. Due to the presence of many different dimensions in the multi-host systems, only

selected layers can be represented through phase plots. Some of these plots are laid out on a logarithmic scale as it is found that oscillations may occur on a minuscule level compared to the overall population size.

As the experiments begin with short-term observations, the time scale of the temporal values determining γ, μ and v kept throughout the studies is set to days, meaning that yearly lifespan measures are converted. Whenever longer periods of development are observed, the time scale in the plots is presented in years. The observed time spans for the separate experiments are documented in Table 3.

Model	Time-span
SIR without demography (basic, epidemic)	90 days
SIR without demography (basic, no epidemic)	90 days
SIR without demography (fitting)	14 days
SIR without demography (vaccination)	90 days
SIR with demography (basic)	5 years
SIR with demography (tuned for oscillations)	70 years
SIR with demography (oscillation analysis)	150 years
SIR with demography (infection-induced mortality)	50 years
Multi-host model	900 days

Table 3: Simulated time spans of the different experiments.

For the first part of the experiments up to the infection-induced mortality model, a frequency-dependent transmission is assumed and the variables of interest are the proportions S, I and R. As soon as larger fluctuations of the population size are made possible due to decoupled birth and death rates, the focus is redirected to the actual number of individuals X, Y and Z.

When faced with the task of proposing vaccination policies, several assumptions were made. First, it was decided to interpret the problem as a short-term intervention taking place either at the beginning of the disease spread or throughout its course. Hence the proposals do not consider the vaccination of young children prior to the outbreak or of other targeted sub-compartments of the susceptible class. Second, it is assumed that the vaccine starts acting immediately and is 100% efficient. The first two vaccination strategies following the fitted naïve SIR model are based on simple observations of the existing model as it is. The third strategy proposes an adjustment of the model to incorporate a vaccination term α to simulate a vaccination process happening parallel to the disease spread and transferring individuals from the S to the R compartment at every time step, the ODEs thus becoming:

$$\frac{dS}{dt} = -\beta SI - \alpha \tag{27}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{28}$$

$$\frac{dR}{dt} = \gamma I + \alpha \tag{29}$$

The formulation of this method has not been found to have a direct precedence in literature but was considered analogical to other commonly used rates such as μ. The criterion for choosing it was that it should realistically represent a constant number of the entire population receiving immunity per time frame rather than being just a proportion of the susceptibles. Furthermore, its introduction in the equations ensured that the rule $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt}$ is respected. Due to the limited time-frame for the assignment a rigorous justification of this method could not be conducted, so it remains subject to further discussion.

5 Experiments

The first set of experiments set up the basic SIR model without demography and test its compliance with the predicted behaviour. The model class is instantiated with the preset parameters

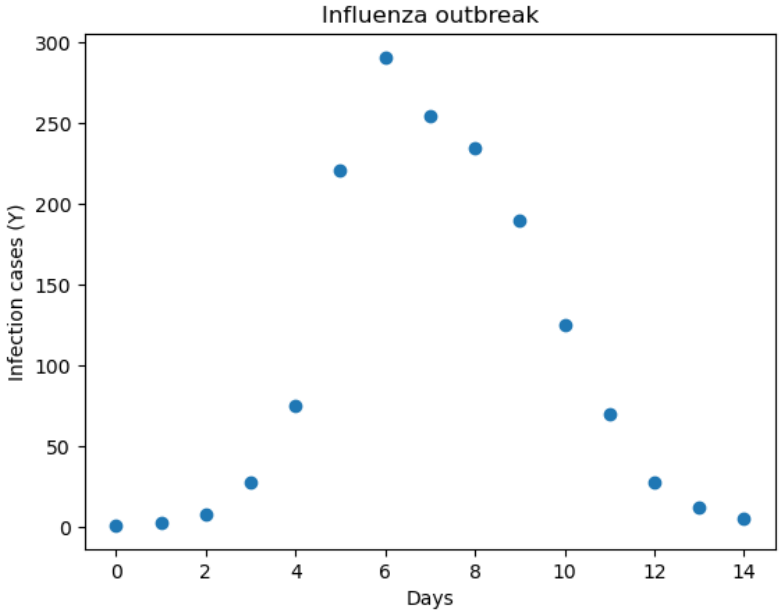

Figure 1: Historical data from an influenza outbreak at a boys school.

which are deduced to result in an infectious growth and the shifts between the three different compartments are traced over time. To examine the relationship between the susceptible and the infected fraction, the two variables are plotted over time, forming indicative trajectories in the phase plot. The outcome is then analysed for the presence of an epidemic and, if present, a new set of parameters is calculated to represent the opposite state. This is done by fixing all parameters except for the infection probability c and asking the question: for what values of c does the infection never experience growth (i.e. $R_0 < 1$)? This can be directly calculated by rewriting the expression for $R_0 T$:

$$R_0 = \frac{\beta}{\gamma} = \frac{-\kappa ln(1-C)}{\gamma} \tag{30}$$

$$\gamma R_0 = -\kappa ln(1-C) \tag{31}$$

$$ln(1-C) = -\frac{\gamma R_0}{\kappa} \tag{32}$$

$$C = 1 - e^{-\frac{\gamma R_0}{\kappa}} \tag{33}$$

The results of the new parameter configuration are plotted for comparison.

The prototype for the naïve SIR model is then applied to a fitting procedure that finds the optimal values of β and γ which bring the curve of I closest to the recorded cases (Figure 1). With these parameters still applied to the model, an additional study is conducted to determine three vaccination strategies:

1. Studying the phase plot of the fitted model, qualitative and quantitative observations are made about the percentage of susceptibles that need to be transferred to the recovered compartment through the means of vaccination on day zero of the outbreak;

2. If the necessary number of removed susceptibles at the start of the outbreak cannot be achieved, a compromise is sought with the percent of day-zero vaccinations which can still keep the epidemic peak below a certain threshold (in this case set to 0.1 or 10% of the population). This is done by approximating the initial value of R that minimizes the difference between the threshold value and the actual peak of the I compartment.

3. If mass vaccination is not possible immediately for logistical reasons (for instance because of a large population size) an extended model is developed in which the vaccination can

take place parallel to the development of the epidemic. By plotting the results from different vaccination rates, it is deduced for what rates of vaccination the effect of the spread can be minimized as much as possible.

For the next set of experiments the infection model is extended to incorporate demography. Using the predefined parameter values it is observed at which point the model starts exhibiting oscillatory behaviour. Subsequently, a range of β and γ parameters is explored to identify when the oscillations happen at a significantly high equilibrium of I. This is made by making use of the knowledge from Equation 9 that the I^* coordinate of the fixed point is located at:

$$I^* = \frac{\mu}{\beta}(R_0 - 1) \tag{34}$$

Breaking down R_0 into its constituent parameters β and γ enables the expression of I^* solely as a function of the two:

$$I^* = \frac{\mu}{\beta}(\frac{\beta}{\gamma + \mu} - 1) = \mu(\frac{1}{\gamma + \mu} + \frac{1}{\beta}) \tag{35}$$

The relationship between the three variables can now be represented in a three-dimensional plot, from which the suitable values can be deduced. These values are then applied in a new instance of the model, from which a phase plot is created to illustrate the periodic motion around the fixed point. A simple calculation of the coordinates through the formula in Equation 9 lets us verify the correctness of the fixed point as seen in the plot.

The next step is to quantify the oscillatory behaviour around the equilibrium point. The approximate value of its frequency can be found by recalling the formula for the period T from Equation 13. The result can be empirically tested by applying Fourier analysis on the trajectory of the S values over time, as these should ultimately be a phase-offset image of the I values in the later stages, but with a higher magnitude.

The next iteration of the disease model keeps the same parameter values but embeds the newly introduced infection-induced mortality rate. An interesting study to conduct with this model is its behaviour under variation of its key component, the infected mortality rate ρ. For this, phase plots of values of $\rho = 0.25, 0.5, 0.75$ are meant to give an indication of the changes. For a more detailed observation, the coordinates of the endemic state (X^*, Y^*, Z^*) are plotted

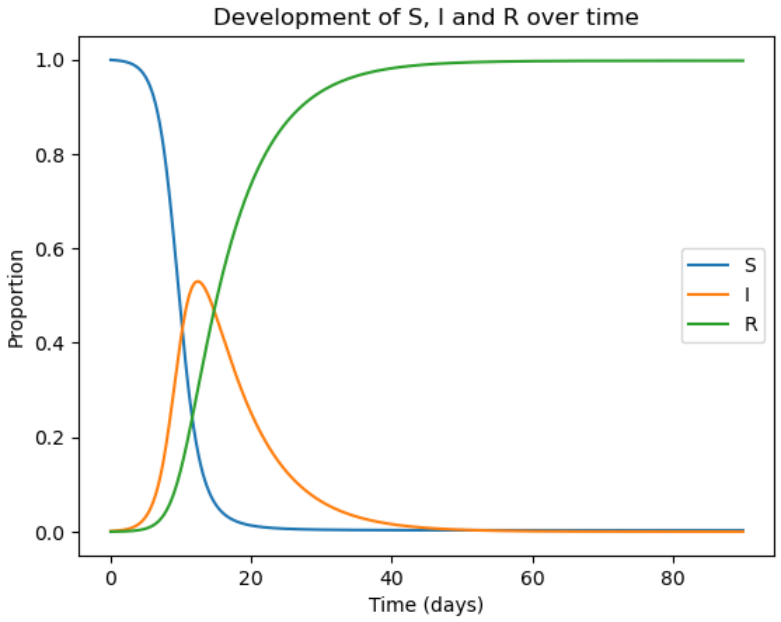

Figure 2: Changes in the proportions of the three population compartments of the naïve SIR model ($R_0 \approx 5.9$).

as a function of ρ within the range of $[0, 1]$. This is performed by substituting Equation 16 in Equations 17, 18 and 19.

The final set of experiments involves the multi-host model with vectored transmission. As with previous experiments, a plot of the population compartments for each of the species over time is presented. Two of the defining parameters are explored in their effect towards the system's behaviour: the bite rate b and the ratio between population sizes. In addition, temporal forcing is introduced in the model to represent seasonal variations in the birth rate of mosquitoes. The resulting oscillatory reactions in both populations are then analysed through plots.

6 Results

The plotted developments of the S, I and R compartments (Figure 2) show a behaviour concurrent with the examples shown in literature and throughout the course. An onset of infected individuals is noticeable shortly after the start of the simulation, with the number of susceptibles rapidly declining. Shortly after the infection onset a rise in the R numbers is observed, continuing in a sigmoid fashion as I reaches a peak at about half the population size and then begins to subside with a slightly more prolonged decay.

An internal calculation of R_0 from the input parameters shows a value of approximately 5.9. As this value is larger than 1, an epidemic onset is to be expected. The phase plot examining the trajectories of S versus I for different initial conditions shows a consistent peak of infection at $S \approx 0.169$ (Figure 3).

To capture a behaviour with no epidemic the necessary c is calculated using Equation 33, yielding $c \approx 0.017699$. This is the threshold value of the infection probability, given that all other values are fixed. This configuration is then visualised in a new set of phase plots (Figure 4).

Running the prototypical model through the curve fitting procedure results in $\beta \approx 1.66$ and $\gamma \approx 0.45$. To qualitatively evaluate the precision of these results, the development curve is plotted together with the original data points in Figure 5. The R_0 value resulting from these parameter settings is approximately 3.72, which is within the expected range for influenza documented in literature (3-4)[17].

Taking this instance of the model with its defined parameters, the first vaccination strategy examines the initial states of (S, I) for which no infectious onset would develop. A phase plot of

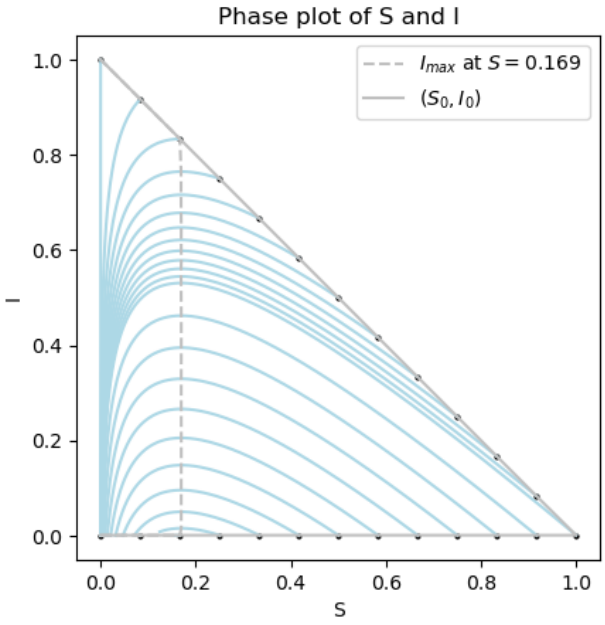

Figure 3: Phase plots of the naïve SIR model exhibiting an epidemic ($R_0 \approx 5.9$).

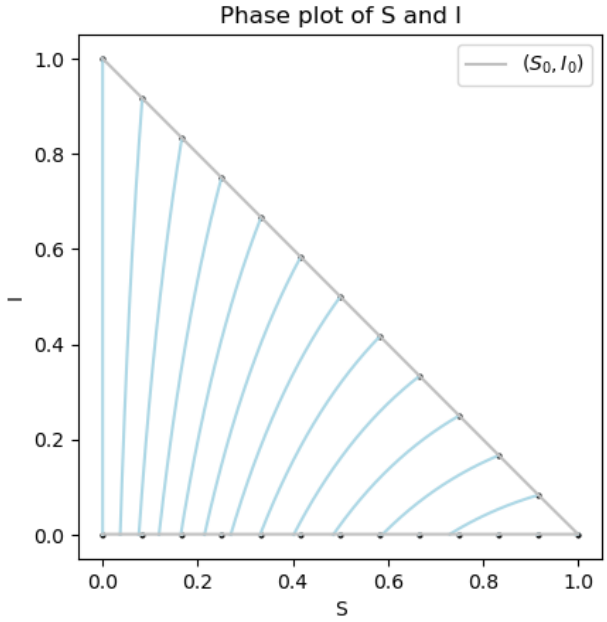

Figure 4: Phase plots of the naïve SIR model with no epidemic ($R_0 \approx 0.85$).

the system with its current settings is created to aid this process (Figure 6).

An obvious assumption is that the starting value of I should be as low as possible, implying that the vaccination should start on the first day of the hypothetical outbreak. For all trajectories containing the peak $\frac{dI}{dt} = 0$ the apex is consistently located at $S \approx 0.27$. This gives sufficient information about the minimum amount of recovered individuals on day zero ($R = 1 - S - I \approx 0.73$).

For the second, compromise-oriented vaccination solution the analysis finds a threshold value of $R \approx 0.43$ ($S \approx 0.57$) for which the infection curve reaches the threshold peak of 0.1 (Figure 7).

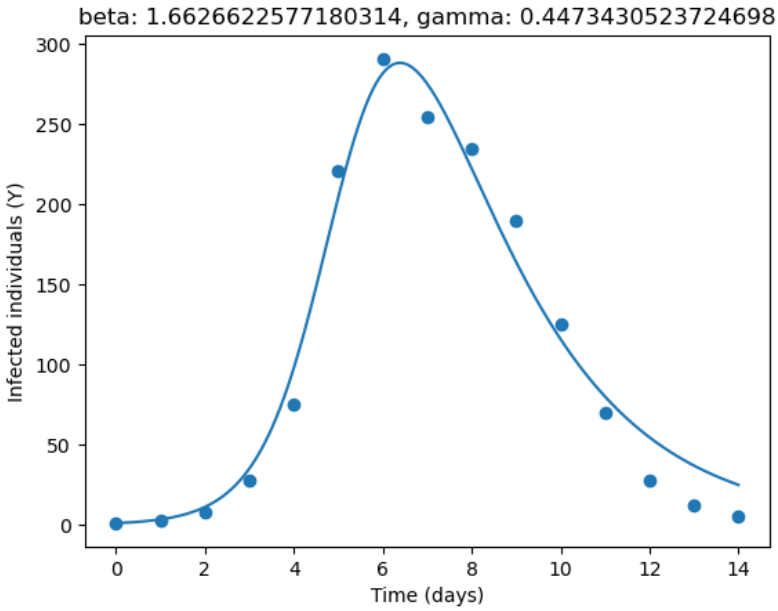

Figure 5: Fitted curve of I over the influenza case data points ($R_0 \approx 3.72$).

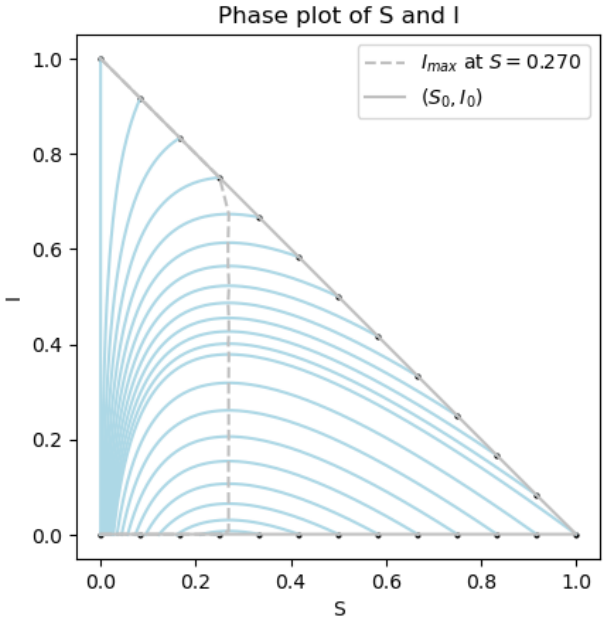

Figure 6: Phase plots of the SIR model fitted to the influenza outbreak. The peak of I values is situated at $S \approx 0.27$.

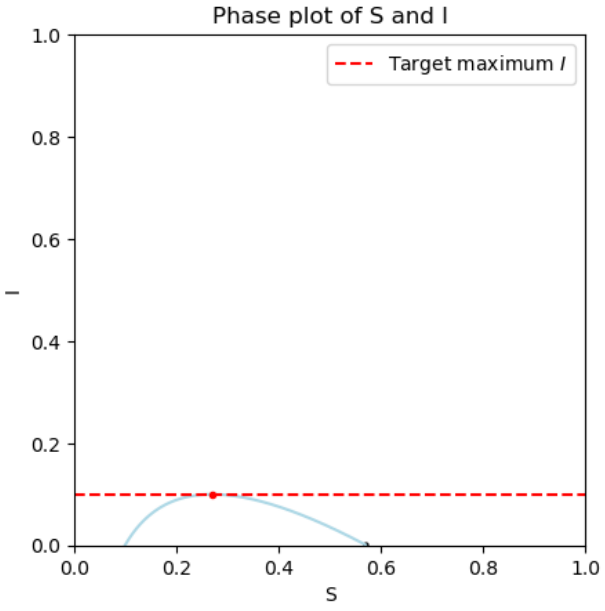

Figure 7: Phase plot of the SIR model with parameters tuned for reaching a preset threshold of $I_{max} = 0.1$.

Running the extended SIR model for the third vaccination strategy, a steady decrease in the infection peak is observable from approximately $\alpha > 0.15$ (Figure 8). Varying the α value does not seem to shift the S component of the infection peak.

The S, I and R plots of the model with demography exhibit oscillations as expected (Figure 9). It is also noticeable that the undulations of I happen at a relatively small value for these settings, so the plot in Figure 10 is used to indicate that for very small values of β (low infection rate) and γ (long recovery and prolonged infectious state) the value of I^* becomes more significant. Therefore, the parameters are tuned so that $c = 0.00025$ and the infectious period is five

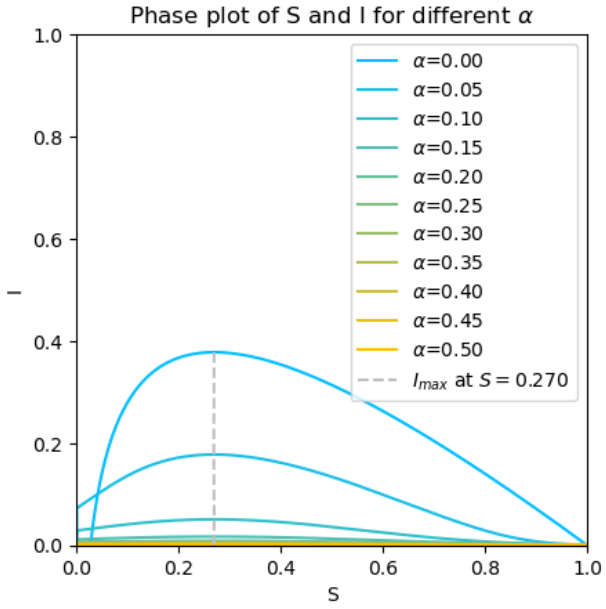

Figure 8: Phase plots of the modified SIR model with vaccination for different values of α. All starting points are at maximum S and $I = 1/N$.

years. These parameter settings result in the phase plot in Figure 11, where a spiraling attraction towards the equilibrium point is clearly noticeable.

Estimating the oscillation period using Equation 13 yields a period T of approximately 21384.23 days, roughly equivalent to 58.55 years. Taking the inverse to represent the frequency, we arrive at 0.017. As the oscillations emerge at a relatively late point in the development and their amplitude gradually decreases, it is hard to pinpoint a precise baseline amplitude. However, we can infer the decay rate of the amplitude using the real term in Equation 10, which

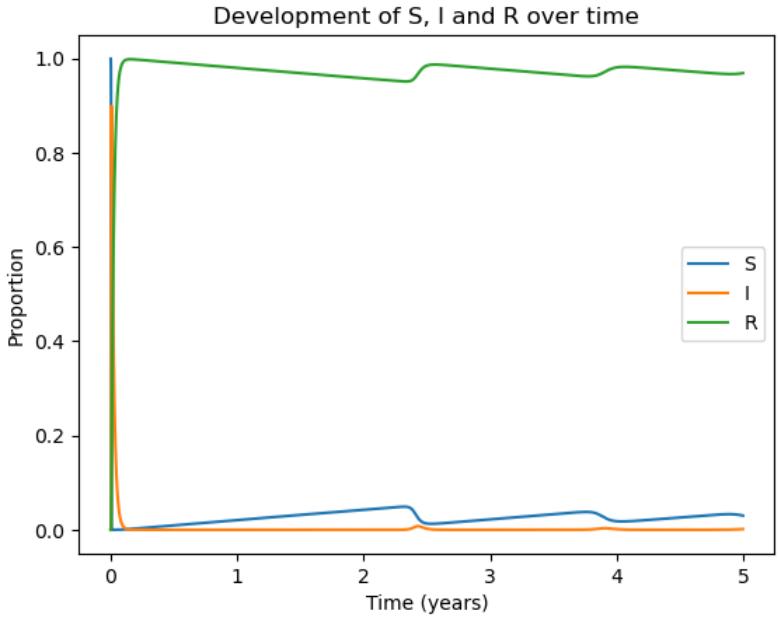

Figure 9: Changes in the proportions of the three population compartments of the SIR model with demography ($R_0 \approx 38.8$).

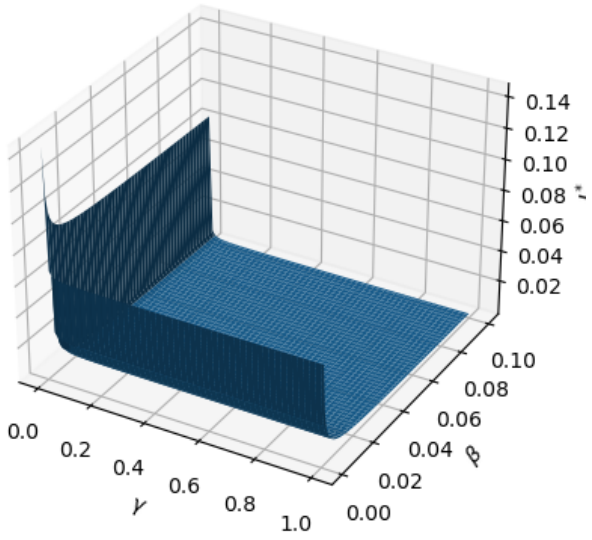

Figure 10: Articulation of the $I*$ component of the equilibrium state for different combinations of β and γ

leads to a value of 10e-5. To verify these analytical derivations of the system a Fourier analysis is performed and the resulting frequency spectrum plotted in a Figure 13. Indeed, beside the zero-frequency accumulation due to the flattening of the oscillations, a peak can be observed around 0.02, which is a close approximation of the predicted value.

After the enhancement of the demography model with an infection-induced mortality term, the phase plots in Figure 14 illustrate the different behaviour for the selected values of ρ. There is a noticeable trend toward disappearance of the endemic stable point for increased ρ, with $\rho = 0.75$ producing a de facto non-epidemic development. This tendency is reviewed in more detail

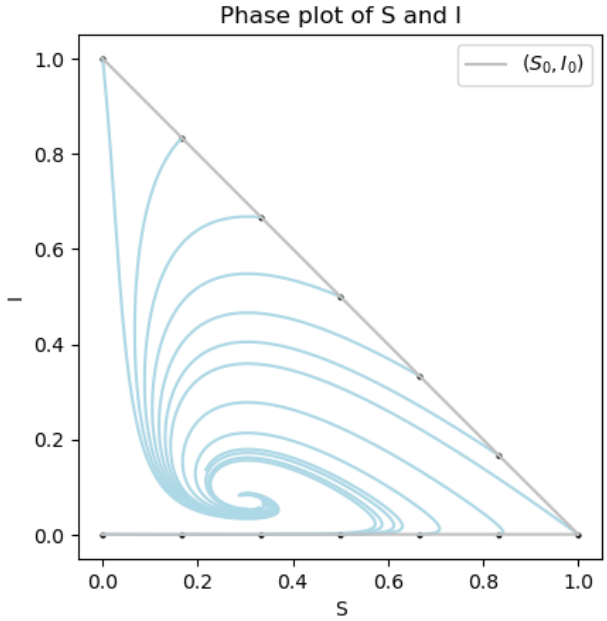

Figure 11: Phase plots of the SIR model with demography exhibiting oscillatory behaviour around an endemic equilibrium ($R_0 \approx 3.28$).

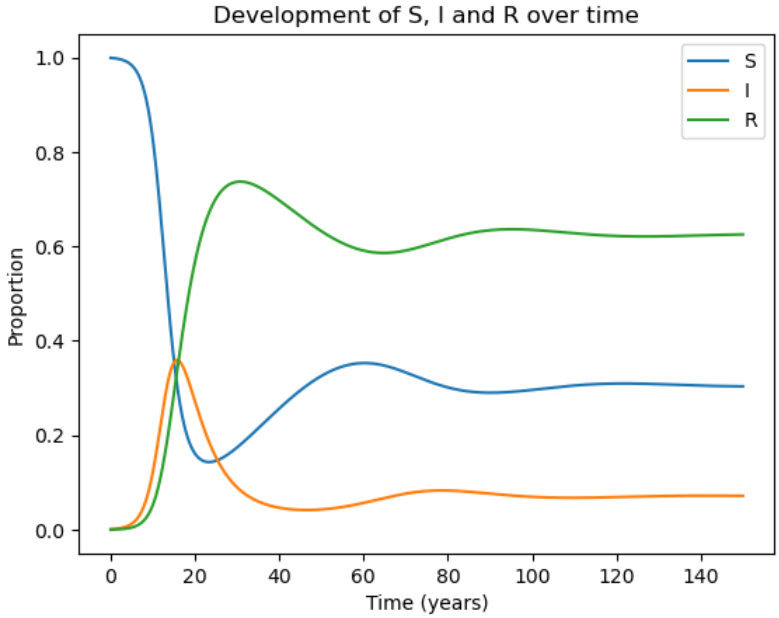

Figure 12: Changes in the proportions of the three population compartments of the SIR model with demography ($R_0 \approx 3.28$).

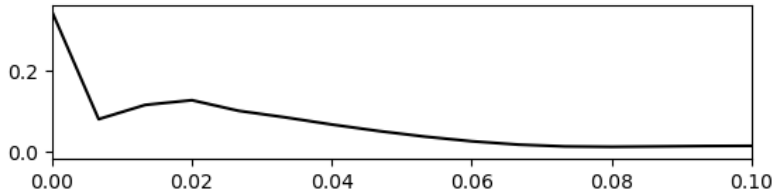

Figure 13: Fourier analysis of the values of S from Fig. 12.

in the continuous graphs of the three coordinates of the equilibrium state for $0 \leq \rho \leq 1$ (Figure 15). As can be deduced from the relationships between (X^*, Y^*, Z^*), ρ via R_0, Y^* and Z^* have a reverse linear relationship to ρ, while X^* is inversely proportional to a decrease in ρ.

A study of the multi-host system also reveals oscillations, albeit around relatively small values of the infected compartment compared to the population size (Figure 16 and 17). A qualitative first glance at the phase plots for different bite rates confirms the recognition that the infected population always shrinks to a relatively low size, but with a noticeably wider range of susceptibles for $b = 0.05$. Furthermore, the curvature of the trajectories seems to be influenced by the variation in bite rate (Figure 18). The study on the effect of different population ratios in Figure 19 also shows qualitatively different trajectories at the beginning of the infection, but eventually these also converge to a small-valued fixed point. In all plots a clear resemblance is found in the curve that the trajectories approach prior to the equilibrium state. Finally, the introduced seasonal forcing of birth rates appears to produce complex series of peaks and valleys which are indicative of the presence of bifurcations in the system's stability.

7 Discussion

As identified by other authors[18] the selection of an epidemiological model needs to be appropriate to the use case and the assumptions made in representing the disease can have a dramatic effect on its prediction power. Due to its relatively high level of abstraction, the naïve SIR model has only limited application but provides a foundation which is easy to comprehend and to build upon. It captures some of the essential aspects in the epidemiological understand-

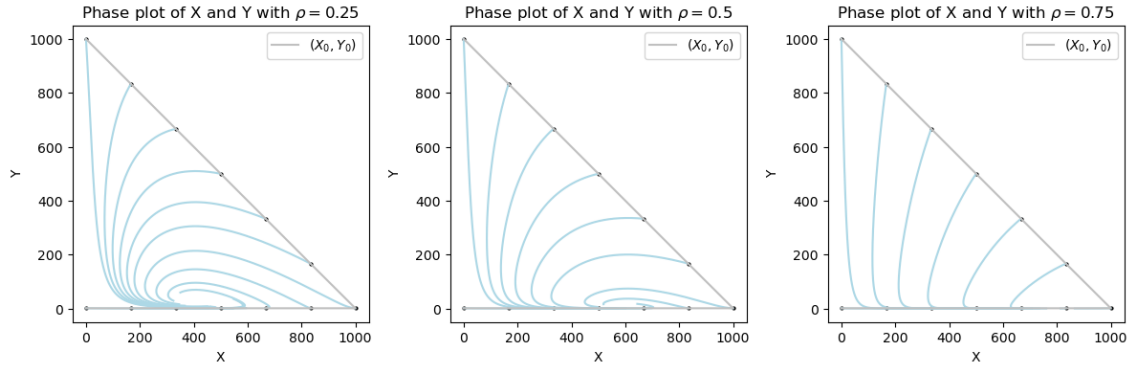

Figure 14: Comparison of the phase plots for the demographic model with infection-induced mortality for different values of $/rho$.

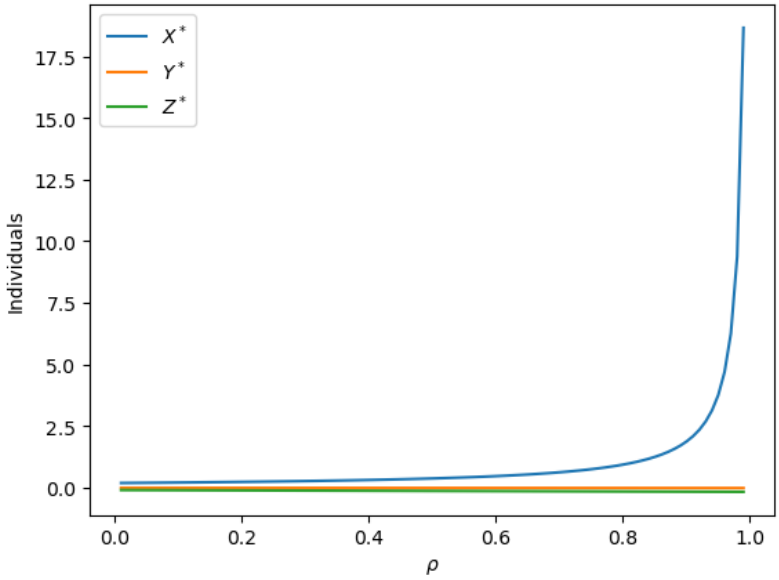

Figure 15: Plot of the three coordinates of the fixed point (X^*, Y^*, Z^*) for varying ρ.

ing of a pathogenic development - an onset of infection dependent on the characteristic mechanisms of the disease, and a decline caused by acquired immunity and respectively by a lack of susceptibles. A state of epidemic becomes present in the model whenever the rate of change in infected individuals experiences any positive increase at all ($R_0 > 1$), otherwise the disease solely undergoes decline. The predictive power of the basic SIR model appears to perform well against the examined influenza case, presumably because demography does not play a significant role in short-term epidemics. Still, some deviations between the epidemic curve and the data points are to be observed at the decay stage, which could be traced back to several possible causes - an inaccurate estimation of the γ parameter due to inherent flaws of the curve fitting algorithm, a false assumption about the mechanisms of recovery, or a stochastic factor which cannot be accounted for in the deterministic model.

The calculations made in support of the three vaccination strategies lead to the following conclusions:

- First vaccination strategy: 73% of the susceptible people need to be vaccinated on the first day of the onset.
- Second vaccination strategy: 43% of the susceptible people need to be vaccinated on the first day of the onset to prevent a peak of infected individuals equivalent to 10% of the

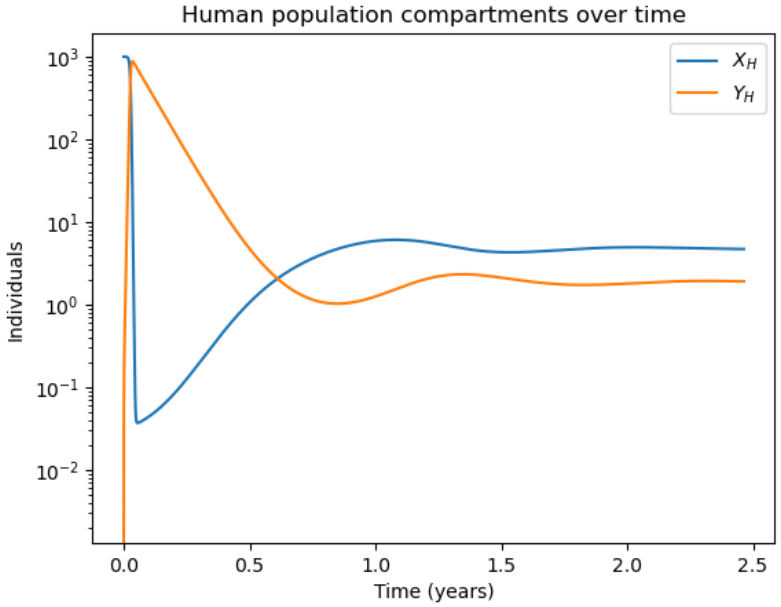

Figure 16: Changes in the proportions of the X_H and Y_H human population compartments of the multi-host model ($R_0 \approx 211.72$).

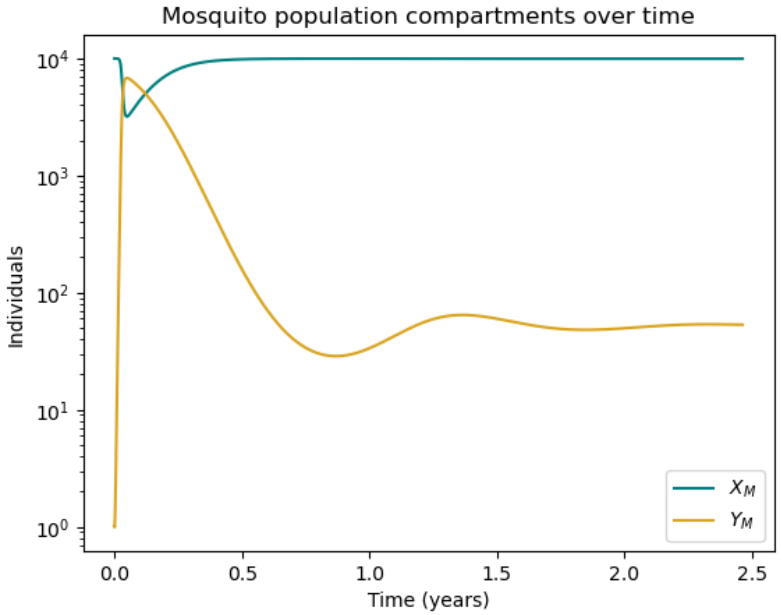

Figure 17: Changes in the proportions of the X_M and Y_M mosquito population compartments of the multi-host model ($R_0 \approx 211.72$).

population (for instance, because of health care system capacity).

- Third vaccination strategy: a vaccination rate of 15% per day can shrink the peak of the epidemic curve to a negligible low.

As noted previously, the introduced vaccination rate needs more rigorous formulation so that it can be related to actual policy measures. A rough assumption may interpret it as a percentage of the population that needs to be vaccinated per day.

The necessity for lowering the disease parameters of the demographic SIR model in order to reproduce a high infectious equilibrium may illustrate some parallels to real-life phenomena. For

instance, this effect would be analogical to the statement that if a pathogen is very mildly infectious, but lingers for a long time in the infected host, it may persist at a relatively high stable abundance. While intuition makes this sound plausible, further research is necessary to identify concrete examples of such diseases.

Regarding the analysis of the oscillations around the equilibrium state, it is noted that their calculated period is an approximation due to the omission of higher order terms in its derivation[4]. However, it effectively proves to be a fairly precise measure for the periodicity phenomenon. The error in value captured by the Fast Fourier Transform method can be explained with the low resolution of the signal, which is an aspect that can be improved for future studies.

The experiments with the infection-induced mortality term ρ clearly indicate that a higher lethality of the disease leads to a faster burnout. It can therefore be speculated that if a pathogen is to prevail for an extended period of time, it must avoid killing its hosts with great efficiency. This points to the notion of an ecosystemic balance, in which parasitic species can only benefit from the survival of their host.

Such is also the case for parasitic vectors, as can be concluded from the results of the multi-host model. The indication that a vector-transmitted disease has an inherent potential to prevail for longer terms among a small number of individuals may explain some of the difficulties in fighting infections like malaria. Since a lot of articulation plays out on a scale much smaller than the overall population size, further analysis and representation methods need to be explored for observing these effects at various levels of magnification. Regarding the modeling of the disease, some intriguing properties of the system are identified, such as the effect of the bite rate on the shape of the stable manifold leading to the convergence point. While it is hard to provide an immediate mathematical explanation of this phenomenon and its relevance for the real-life representativity of the model, it nonetheless reveals characteristic features of this model. As its full network of mechanisms is hard to capture in its entirety due to the high dimension of the variable vector, such observations can prove useful for further investigation.

Finally, the effects caused by the interaction of different oscillation periods in the multi-host model demonstrate that temporally forced birth rates can exhibit complex behaviour just like the seasonal forcing of the β parameter in other models. This invites a more detailed study of the bifurcations under different degrees of forcing as well as an inquiry in the presence of such effects in natural scenarios with diseases like malaria.

8 Conclusion

The experimental series presented in this report follows a transition from a basic, prototypical system to increasingly sophisticated frameworks for the representation of disease transmission. The consecutive steps of this process outline the potentials and limitations of each model type, enable a visual expression of their inner workings through plots and invite interpretations of the simulated phenomena. The models dissected in the report are the product of long-term research conducted in relevant fields, therefore the comprehension of their properties and mechanisms is essential for the goal of applying them in an appropriate way and eventually improving them. The method of implementing and structuring knowledge acquired throughout the course in the construction of the simulation models demonstrates a typical approach for the field of Computational Science. Such an approach aims to unite multifaceted interdisciplinary expertise in the common domain of digital computing in order to provide a better understanding of the complex nature of our world.

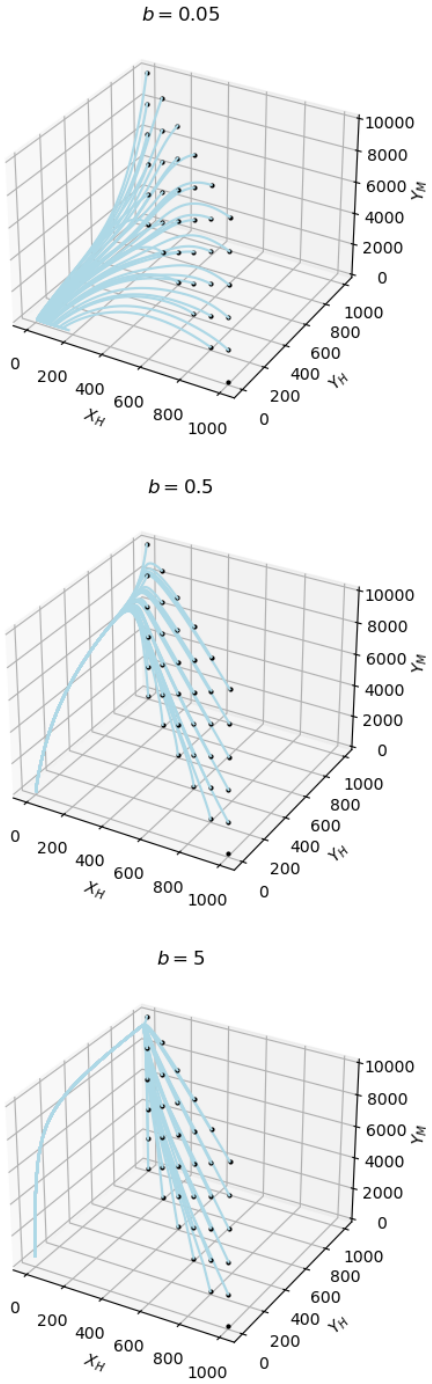

Figure 18: Phase plots of X_H, Y_H and Y_M for different initial numbers of infected and bite rates $b = 0.5, 5, 50$.

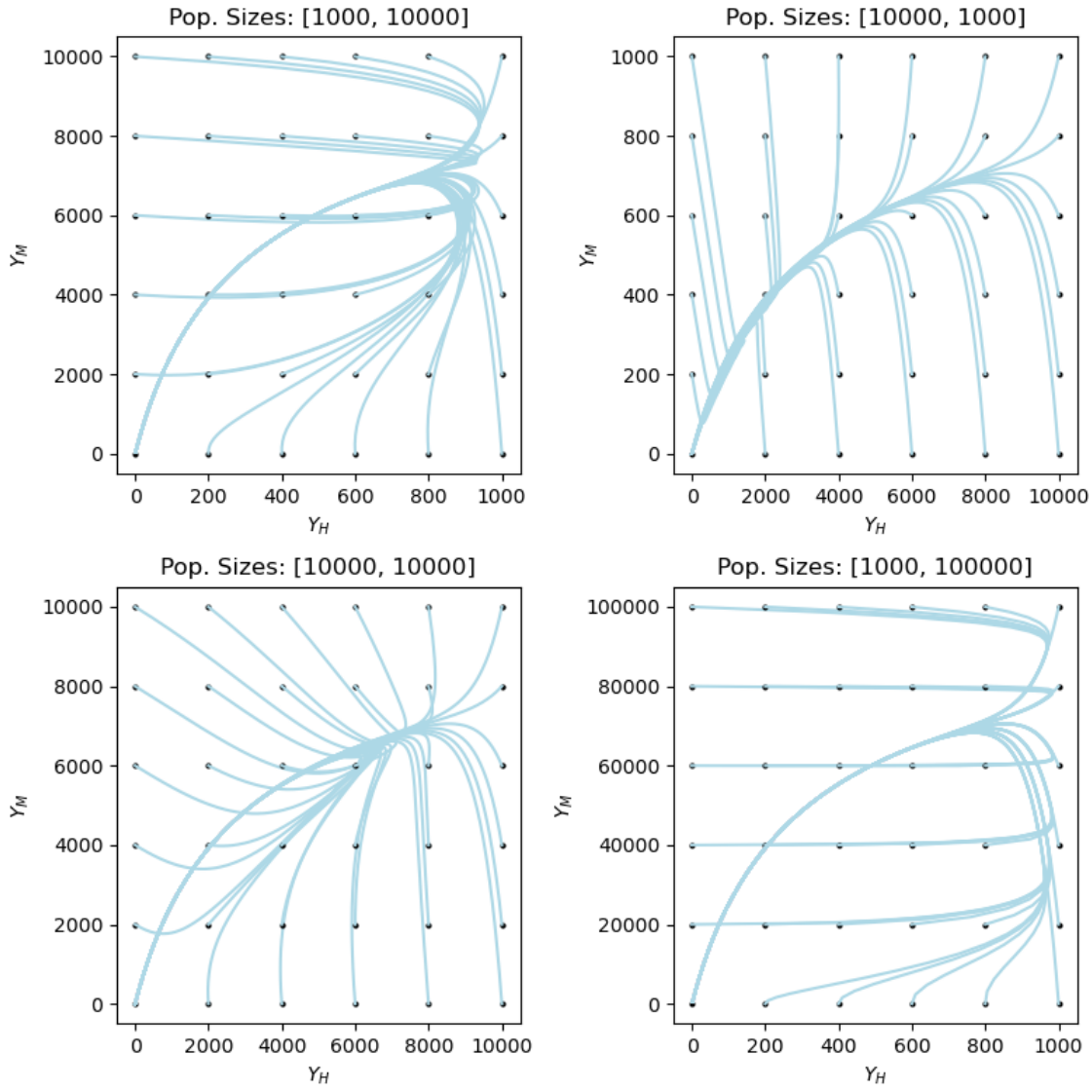

Figure 19: Phase plots of Y_H and Y_M for different initial values Y_H^*, Y_M^* and different ratios of the population sizes.

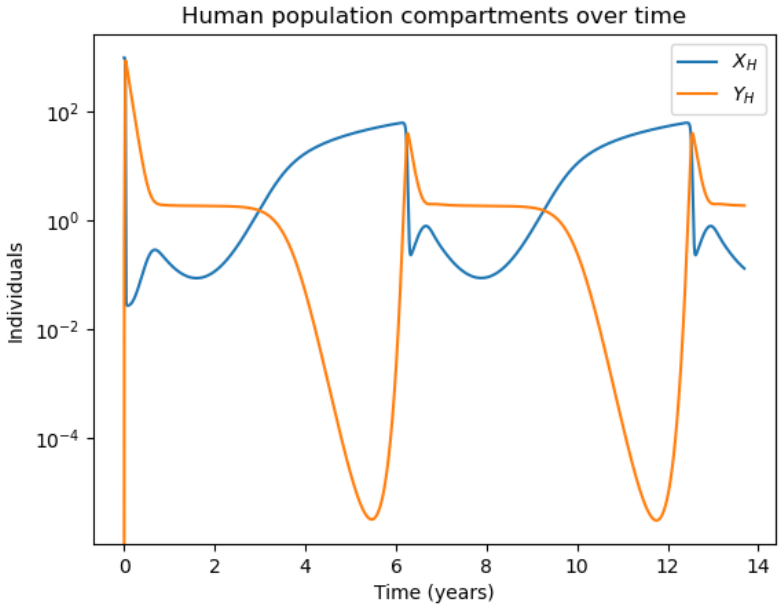

Figure 20: Changes in the proportions of the X_H and Y_H human population compartments of the multi-host model under seasonal forcing with a period of 1 year.

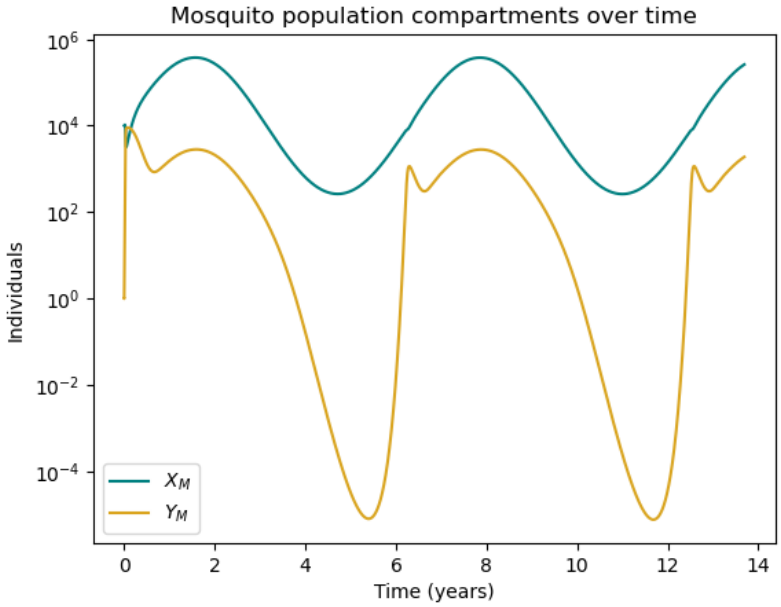

Figure 21: Changes in the proportions of the X_M and Y_M mosquito population compartments of the multi-host model under seasonal forcing with a period of 1 year.

References

[1] Andrew P. Dobson and E. Robin Carper. "Infectious Diseases and Human Population History". In: *BioScience* 46.2 (1996). Publisher: [American Institute of Biological Sciences, Oxford University Press], pp. 115–126. ISSN: 00063568, 15253244. DOI: 10.2307/1312814. URL: http://www.jstor.org.proxy.uba.uva.nl/stable/1312814 (visited on 09/26/2023).

[2] Benjamin Roche and Jean-François Guégan. "Ecosystem dynamics, biological diversity and emerging infectious diseases". In: *Biodiversity in face of human activities / La biodiversite face aux activites humaines* 334.5 (May 2011), pp. 385–392. ISSN: 1631-0691. DOI: 10.1016/j.crvi.2011.02.008. URL: https://www.sciencedirect.com/science/article/pii/S1631069111000709.

[3] Steven H Strogatz. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

[4] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. English. Princeton University Press, 2008. ISBN: 978-0-691-11617-4.

[5] W. O. Kermack and A. G. McKendrick. "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927). Publisher: The Royal Society, pp. 700–721. ISSN: 09501207. URL: http://www.jstor.org/stable/94815 (visited on 09/25/2023).

[6] Ottar N. Bjørnstad, Bärbel F. Finkenstädt, and Bryan T. Grenfell. "Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model". In: *Ecological Monographs* 72.2 (2002), pp. 169–184. DOI: https://doi.org/10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9615%282002%29072%5B0169%3ADOMEES%5D2.0.CO%3B2. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9615%282002%29072%5B0169%3ADOMEES%5D2.0.CO%3B2.

[7] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Vol. 5. John Wiley & Sons, 2000.

[8] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.

[9] H. E. Soper. "The Interpretation of Periodicity in Disease Prevalence". In: *Journal of the Royal Statistical Society* 92.1 (1929), pp. 34–73. ISSN: 09528385. URL: http://www.jstor.org/stable/2341437 (visited on 09/29/2023).

[10] David JD Earn et al. "A simple model for complex dynamical transitions in epidemics". In: *science* 287.5453 (2000), pp. 667–670.

[11] M. Di Pierro. *Annotated Algorithms in Python: With Applications in Physics, Biology, and Finance*. Experts4Solutions, 2013. ISBN: 9780991160402. URL: https://books.google.nl/books?id=cZyPngEACAAJ.

[12] Travis E Oliphant et al. *Guide to numpy*. Vol. 1. Trelgol Publishing USA, 2006.

[13] W. Kutta. *Beitrag zur näherungsweisen Integration totaler Differentialgleichungen*. Teubner, 1901. URL: https://books.google.nl/books?id=Zc4TAQAAIAAJ.

[14] Peter Wegner. "Concepts and Paradigms of Object-Oriented Programming". In: *SIGPLAN OOPS Mess.* 1.1 (Aug. 1990), pp. 7–87. ISSN: 1055-6400. DOI: 10.1145/382192.383004. URL: https://doi.org/10.1145/382192.383004.

[15] Steven W Smith et al. *The scientist and engineer's guide to digital signal processing*. 1997.

[16] Nakul Chitnis, James M. Hyman, and Jim M. Cushing. "Determining Important Parameters in the Spread of Malaria Through the Sensitivity Analysis of a Mathematical Model". In: *Bulletin of Mathematical Biology* 70.5 (July 2008), pp. 1272–1296. ISSN: 1522-9602. DOI: 10.1007/s11538-008-9299-0. URL: https://doi.org/10.1007/s11538-008-9299-0.

[17] JD Murray. "1989Mathematical biology". In: *Berlin: SpringerVerlag. MurrayMathematical biology1989* (1989).

[18] Helen J Wearing, Pejman Rohani, and Matt J Keeling. "Appropriate Models for the Management of Infectious Diseases". In: *PLOS Medicine* 2.7 (July 2005), null. DOI: 10.1371/journal.pmed.0020174. URL: https://doi.org/10.1371/journal.pmed.0020174.

A References LaTeX code

Bijgevoegd zijn de code en bibliografie.