Exploring the SIR model and its expansions

Koen Weverink - 14711982

October 2nd, 2023

Contents

1 Introduction

This report delves into the SIR model - a set of equations used to simulate how infectious diseases spread in a population. The model tracks the number of susceptible (S), infected (I), and recovered (R) individuals over time. The tasks involves numerical integration of the model, considering demographic factors, and exploring different SIR model variations. Much of the information used in this report has been obtained from the book Modeling Infectious Diseases in Humans and Animals By Matthew James Keeling and Pejman Rohani [2].

1.1 Numerical Integration and Epidemic Analysis

Commencing with numerical analyses, the focus shifts to understanding how the SIR model behaves in two scenarios - one with an epidemic and one without. The objective is to identify the parameter determining whether an infectious disease becomes an epidemic. Phase space diagrams will be utilized for a closer look. Subsequently, the focus turns to real-world data from a boys' school influenza outbreak. The aim is to fit the SIR model to estimate parameters. Following that, vaccination plans for the school will be brainstormed and designed. Hypotheses will be formed. Experiments will be designed to test these hypotheses.

1.2 Demography

Incorporating birth and death rates into the SIR model is the next step. The objective is to observe how the inclusion of birth and death rates alters the dynamics, particularly examining any oscillatory behavior in the fraction of the infected population that may occur. Fourier analysis will be employed to examine the frequency and amplitude of these oscillations and the conditions leading to them. Additionally, an infection-induced mortality term will be introduced to observe how the model behaves as the probability of dying while infected approaches 1.

1.3 Variants of the SIR Model

The final phase involves exploring different versions of the SIR model. Here, the SEIR model will be examined. Seasonal effects will be introduced by adding a time-varying sinusoidal rate of infection. The goal is to observe how the model responds to changing levels of seasonal influence.

2 Theory

2.1 The SIR Model

The foundation of the analysis lies in the SIR model, a compartmental model used to understand and predict the spread of infectious diseases within a population. This model partitions the population into three categories: susceptible (S), infected (I), and recovered (R). The dynamics of these categories are governed by a set of ordinary differential equations (ODEs), providing a mathematical representation of how the number of individuals in each category changes over time. The SIR model equations are as follows:

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

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

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

Here, β is the parameter for the rate of infection, and γ is the parameter for the rate of recovery. An initial assumption is made by considering the population in our model as "closed", implying that $S + I + R = 1$. [2]

2.1.1 SIR with added demography

The aforementioned SIR model does not resemble the real world accurately. Therefore, other factors need to be added to represent it better. The first one is birth and death rates. When adding these to an SIR model, the natural demographic processes of population growth and decline is taken into consideration as well. This modification introduces an additional parameter μ. In this report, it is assumed that the birth and death rates are the same and thus they are both μ. In the formula's below the ODEs with added birth and death rates can be seen. [2]

$$\frac{dS}{dt} = \mu - \beta SI - \mu S \tag{4}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I \tag{5}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{6}$$

2.1.2 SIR with added mortality rate

Next an infection induced mortality term is added to the model, which will involve a probability of dying while infected. This parameter is ρ. [2] When this term is added to the ODEs, this is what they look like:

$$\frac{dS}{dt} = \mu - \beta SI - \mu S \tag{7}$$
$$\frac{dI}{dt} = \beta SI - (\frac{\gamma + \mu}{1 - \rho})I \tag{8}$$
$$\frac{dR}{dt} = \gamma I - \mu R \tag{9}$$

2.2 The SEIR Model

Expanding the epidemiological modeling beyond the SIR framework, the SEIR (Susceptible-Exposed-Infectious-Recovered) model incorporates an additional compartment for exposed individuals. This enhancement recognizes the latent period during which individuals have been exposed to the disease but are not yet infectious. The SEIR model is described by the following set of ODEs:

$$\frac{dS}{dt} = \mu - (\beta I + \mu)\, S \tag{10}$$
$$\frac{dE}{dt} = \beta SI - (\mu + \sigma)E \tag{11}$$
$$\frac{dI}{dt} = \sigma E - (\mu + \gamma)I \tag{12}$$
$$\frac{dR}{dt} = \gamma I - \mu R \tag{13}$$

Here, S represents susceptible individuals, E denotes exposed individuals, I represents infectious individuals, R stands for recovered individuals. Parameters include: β (rate of infection), σ (rate of transition from exposed to infectious), and γ (rate of recovery) [2].

2.2.1 SEIR Model with Seasonality

Adding seasonality to an SEIR model takes into account the influence of seasonal factors on the transmission dynamics of infectious diseases. Seasonal variations can affect the contact rates, transmission probabilities, and other parameters involved in the spread of diseases. In this report, the effects of seasonality is limited to the transmission rate. Instead of β being constant throughout the time span, it will now be an ODE along with S, I and R. It is shown in the formula's below.

$$\beta(t) = \beta_0 \cdot (1 + \epsilon \cdot sin(2\pi f t)) \tag{14}$$

Here, f is the frequency of the seasonal variation, and ϵ controls the ampli-

tude of the variation [2]. The new ODEs now look as follows:

$$\frac{dS}{dt} = \mu - (\beta(t)I + \mu)\, S \tag{15}$$

$$\frac{dE}{dt} = \beta(t)SI - (\mu + \sigma)E \tag{16}$$

$$\frac{dI}{dt} = \sigma E - (\mu + \gamma)I \tag{17}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{18}$$

2.3 Numerical Integration of Differential Equations

Numerical integration plays a crucial role in solving ODEs, providing a practical approach to understanding the behavior of complex systems like the SIR model. While analytical solutions for ODEs are often elusive, numerical methods offer a feasible means of approximating the system's evolution over time. There are multiple methods for numerically solving ODEs, such as: Euler's method, the Runge-Kutta method and Adaptive Methods. In this report the Adaptive Methods are used, specifically the `solve_ivp` module from scipy.

This method uses a variant of the Runge-Kutta method known as the Dormand-Prince method [1]. This is an adaptive algorithm that automatically adjusts the time step to ensure accuracy while efficiently handling different scales of behavior in the solution. Specifics of this method will be discussed in the the Methods section.

2.4 Phase Space Diagrams

The phase space for the SIR model is two-dimensional, encapsulating the states of susceptible (S) and infected (I) individuals. A trajectory in this space represents the evolution of the system over time. Each point along the trajectory corresponds to a specific combination of S and I at any given moment, providing a dynamic visualization of the epidemic's progression.

Fixed points within the phase space denote equilibrium states where the rates of infection and recovery achieve balance, resulting in a stable distribution of individuals in the population. Identification and analysis of these points contribute to understanding the long-term behavior of the epidemic.

In practical terms, the application of phase space diagrams in the analysis of the SIR model aids in visually unraveling complex dynamics. By seeing the trajectory paths and fixed points, a nuanced understanding of how different parameter conditions influence the transition between epidemic and non-epidemic states can be gained.

3 Methods

3.1 The `scipy.integrate.solve_ivp` module

The `scipy.integrate.solve_ivp` module, part of the `scipy` library, is employed for the numerical solution of the system of ordinary differential equations (ODEs) inherent in epidemiological models like the SIR and SEIR models. The chosen method within `solve_ivp` is the adaptive Runge-Kutta method, specifically the Dormand-Prince method (`DOP853`). This adaptive approach adjusts the step size dynamically during integration, rendering it suitable for systems with varying timescales and rapid changes in behavior [1].

3.2 Parameter estimation for an epidemiological model

In order to try to fit an SIR model to historical data and estimate the parameters (β and γ) of the model, the `scipy.optimize.minimize` function [4] was used in combination with a Mean Squared Error (MSE) calculation. Specifically, the often used 'L-BFGS-B' method was used for the optimization algorithm (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with bounds) [5], where the bounds have to be non-negative. For the specific implementation of these methods, please refer to the accompanying code.

3.3 Fourier Analysis

Fourier analysis is a technique that decomposes a function into a sum of sinusoidal functions with different frequencies. Applying Fourier analysis to an oscillating SIR model with demography can be beneficial for performing frequency analysis, where the frequency components of the oscillations in the model is analyzed [3]. This can help identify dominant oscillatory patterns and understand the periodic behavior of the infectious disease spread within a population. Also, Resonance and Stability Analysis can be performed, where by examining the Fourier transform of the model equations, the stability of the system and identify resonant frequencies can be analyzed. Understanding the stability of the model is crucial for predicting the long-term behavior of the infectious disease dynamics.

In order to do this, firstly a Fast Fourier Transformation [6] was performed on the data of the infected population. The result contains the complex values representing the amplitudes and phases of the frequency components. Next, the frequencies corresponding to the components in the Fourier-transformed data was calculated, where the absolute values of the complex amplitudes was obtained from the Fourier transform. These amplitudes represent the magnitude or strength of each frequency component.

4 Experiments

4.1 Numerical integration of the SIR model

4.1.1 Two scenarios

Firstly, the ODEs of the standard SIR model have been numerically integrated using the `solve_ivp` module of the `scipy` library. Figure 1 illustrates the time evolution of the susceptible, infected, and recovered populations over a 100-day period. The parameter values used are $\beta = 0.3$ and $\gamma = 0.1$ ($R0 > 1$)for the first plot, and $\beta = 0.3$ and $\gamma = 0.4$ ($R0 < 1$) for the second. The initial conditions are $S_0 = 0.9$, $I_0 = 0.1$, and $R_0 = 1 - S_0 - I_0$ for both plots. The first plot clearly shows the dynamics of the epidemic, with an initial increase in infections followed by a recovery phase. The second plot shows that the amount of infected only drops over time and is, therefore, not considered an epidemic.

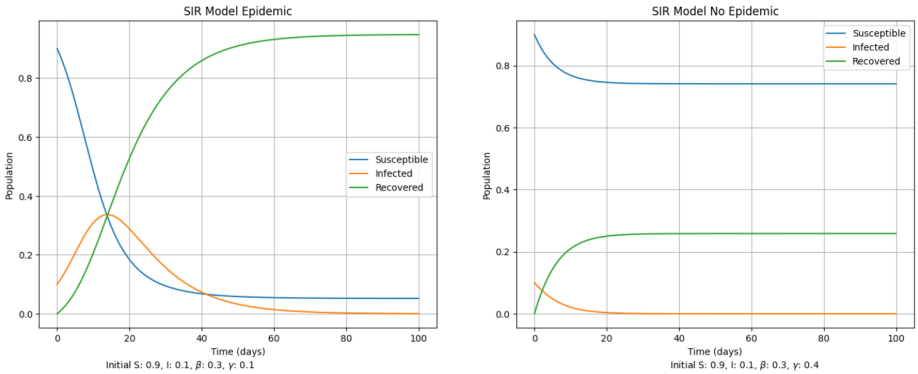

Figure 1: Two scenario's of time evolution of populations in the SIR model.

Next, the Phase Space plots of these scenario's are shown in Figure 2. The first one is what an epidemic looks like, and the second one what an epidemic does not look like. The first plot clearly shows that once the susceptible population is more than the plotted threshold value ($\frac{1}{R0}$), a significant rise in infected people can be observed. However, in the second plot, no matter how many susceptible people there are, the infected population will always drop immediately without rising first. Therefore, in the first scenario contains an epidemic and the second scenario does not.

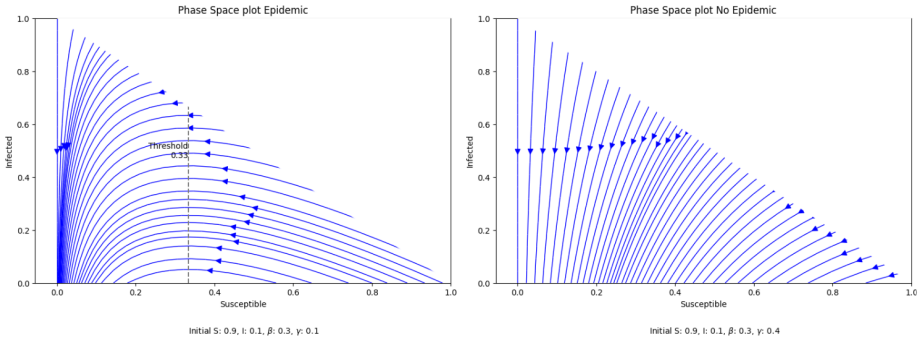

Figure 2: Phase Space plot of the two scenario's

In short, the parameter that determines if the infectious disease will become an epidemic is the reproduction number $R0$. $R0$ is a crucial epidemiological parameter used to measure the transmission potential of an infectious disease in a population. In simple terms, it represents the average number of secondary infections produced by one infected individual in a completely susceptible population. It can be calculated by doing the following: $R0 = \frac{\beta}{\gamma}$. If this number is greater than one, the disease will become an epidemic; if not, it will not become one. Therefore, if $\beta > \gamma$ it will be an epidemic, if $\beta < \gamma$ it will not be.

4.1.2 Influenza outbreak fit

In this problem, the aim is to optimize the parameters of a Susceptible-Infected-Recovered (SIR) model tailored to a student population. The parameters of interest are the transmission rate (β) and the recovery rate (γ), which collectively influence the spread and control of an infectious disease. The total student population is 763. The initial conditions for the SIR model are set based on a presumed scenario where one student is initially infected, and the rest are susceptible.

$$S_0 = \frac{762}{763} \quad I_0 = \frac{1}{763} \quad R_0 = 1 - S_0 - I_0$$

The course of the disease has been given in the following table:

Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number	1	3	8	28	75	221	291	255	235	190	125	70	28	12	5

To determine the optimal values for β and γ, an optimization process is employed. The optimization method used is L-BFGS-B, which allows for bounds on the parameters β and γ. The initial guess for is set to 0.2 and 0.1 respectively. After optimization, the optimal parameter values have been calculated to be $\beta = 1.66476$ and $\gamma = 0.44803$. These values represent the best-fitting parameters for the SIR model, aligning the model's predictions with the observed target array. In Figure 3, the fit is plotted.

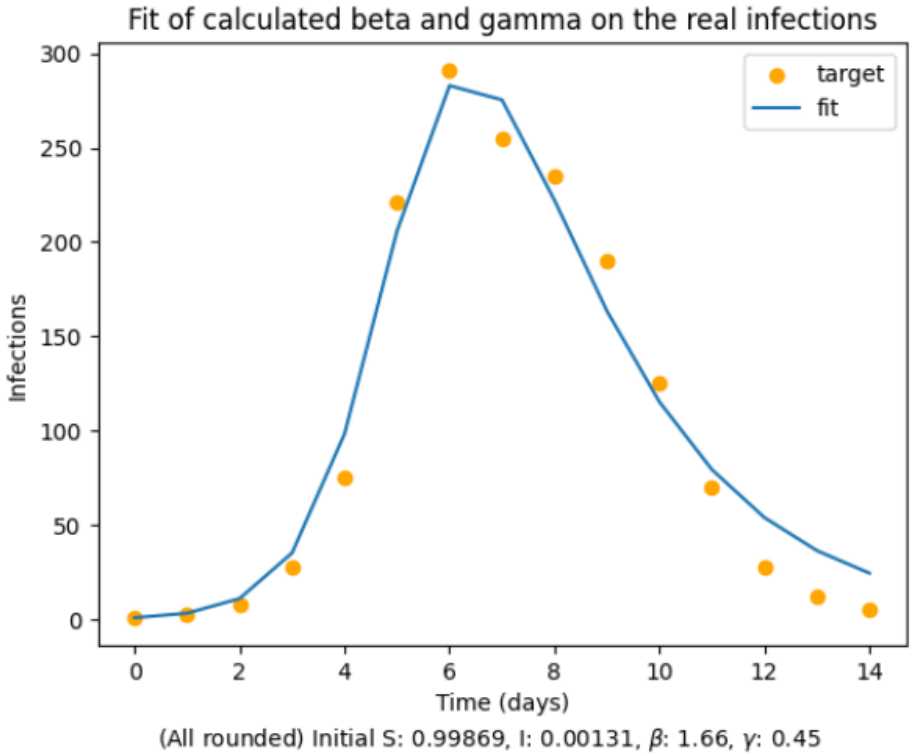

Figure 3: Fit of calculated β and γ on given infections

4.1.3 Vaccination Strategy

This section addresses the design of a vaccination strategy for the boys' school mentioned in the previous section. The two strategies that have been chosen in the experiment are:

1. Every day 5% of the class population is vaccinated;
2. A percentage of the population is vaccinated when a single infection is observed. This percentage is based on the threshold value.

A couple of assumptions have to be made. Firstly, it is assumed that vaccinations work immediately and are 100% effective. Also, for the second strategy, it has been assumed that the β and γ is known (e.g. from a previous observation of a different school with the same amount of infections in the same population).

4.1.4 Hypotheses

Strategy 1: Implementing a daily vaccination strategy, targeting 5% of the population, will significantly mitigate the spread of the infectious disease in the

boys' school. It is anticipated that a noticeable reduction in the number of infected individuals and a more controlled progression of the epidemic compared to scenarios without vaccination will be observed.

Strategy 2: Implementing a preventive vaccination strategy, where a certain percentage of the population is vaccinated when there is a single infection, based on a previously observed threshold value, will prevent an epidemic from forming.

4.1.5 The Experiment

Firstly, the differential equation for the susceptible population need to be altered to include the vaccination rate, ρ. This will result in the following:

$$\frac{dS}{dt} = -\beta SI - \rho S \tag{19}$$

Secondly, the threshold value needs to be calculated for strategy 2:

$threshold = \frac{1}{R0}$
$R0 = \frac{\beta}{\gamma} \rightarrow R0 = \frac{1.66476}{0.44803} \approx 3.715758$
$threshold = \frac{1}{3.715758} \approx 0.26912$

This means that approximately 27% of the population does not have to be vaccinated in order for the infection to die out before becoming an epidemic. So, for the new initial susceptible $S0$ is chosen $\frac{762}{763} \cdot 0.26912 \approx 0.242211$. Which means that about 205 boys can be unvaccinated for this strategy to work and so at least 558 boys need to be vaccinated.

4.1.6 The Results

In Figure 4, three distinct scenarios are shown: a baseline without any vaccination, a strategy where 5% of the population is vaccinated each day, and a prevention vaccination strategy. All vaccinated or insusceptible people have been added to the Recovered population.

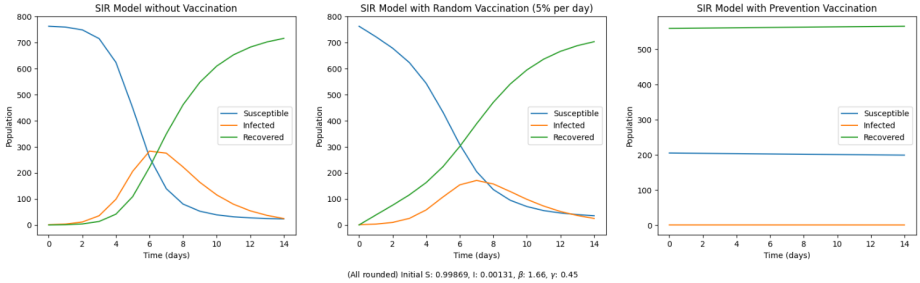

Figure 4: Baseline + two vaccination strategies

4.1.7 Evaluation

Figure 4 illustrates that the second plot shows a significantly lower peak in the number of infected people compared to the first plot. This means that implementing a daily vaccination strategy, targeting 5% of the population, will significantly mitigate the spread of the infectious disease in the boys' school. Though it is important to note that there is still an epidemic happening.

In the third plot of Figure 4, it can be seen that no epidemic will take place, since the threshold value has not been reached. Therefore, implementing a preventive vaccination strategy, where a certain percentage of the population is vaccinated when there is a single infection, based on a previously observed threshold value, will prevent an epidemic from forming. An important note, though, is that the β and γ need to be known beforehand, which in reality is very hard to achieve.

4.2 Demography

The SIR model with demography was extended to incorporate birth and death rates (μ). The aim was to observe changes in the dynamics, specifically running the model to capture an endemic state and investigating potential oscillatory behavior in the fraction of the infected population.

4.2.1 Change in Dynamics

First, the model was expanded with the new parameter. Assumed was that the birth and death rate is equal. These can be seen in the Theory chapter. Next, numerical integration was performed on this model, similar to previous times. The initial conditions were: $S0 = 0.99, I0 = 0.01$ and $R0 = 1 - S0 - I0$. The parameter values were: $\mu = 1/60, \beta = 0.3, \gamma = 0.1$. The results are illustrated in Figure 5.

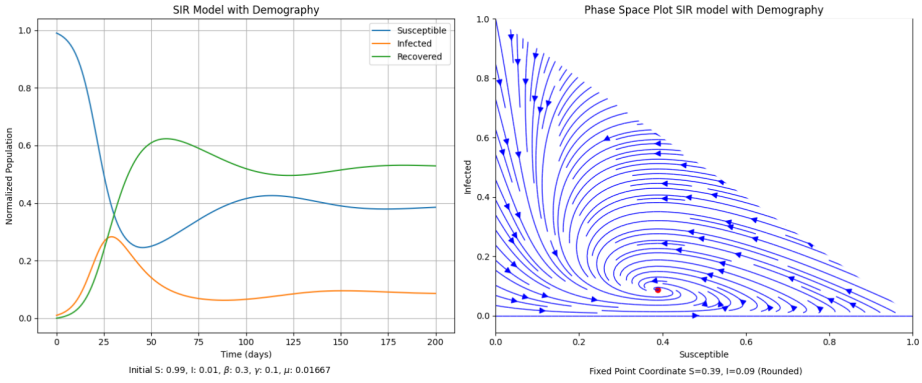

Figure 5: SIR with Demography

Due to the introduction of birth and death rates, the susceptible pool was

constantly filled with new susceptible people. Because of that, the infection does not die out, but keeps being reintroduced into the population. This is the cause of the oscillation in the fraction of infected people, which is called an endemic state. After a certain amount of time, an equilibrium is reached. This equilibrium point is dependent on all the factors with which the model was run. To find this equilibrium, Fixed Point Analysis was performed using the root-finding method. The result of this can be seen in the Phase Space plot as a red dot.

In order to truly prove that this is an endemic state, Fourier analysis was performed on the infected population; the details of which are described in the Methods chapter. In summary, the Fourier transform of the infected population was computed, revealing the frequency components present in the dynamics of the system. Also the corresponding frequencies and the magnitudes (amplitude) of these frequency components were calculated. In Figure 6, the result of this analysis is shown.

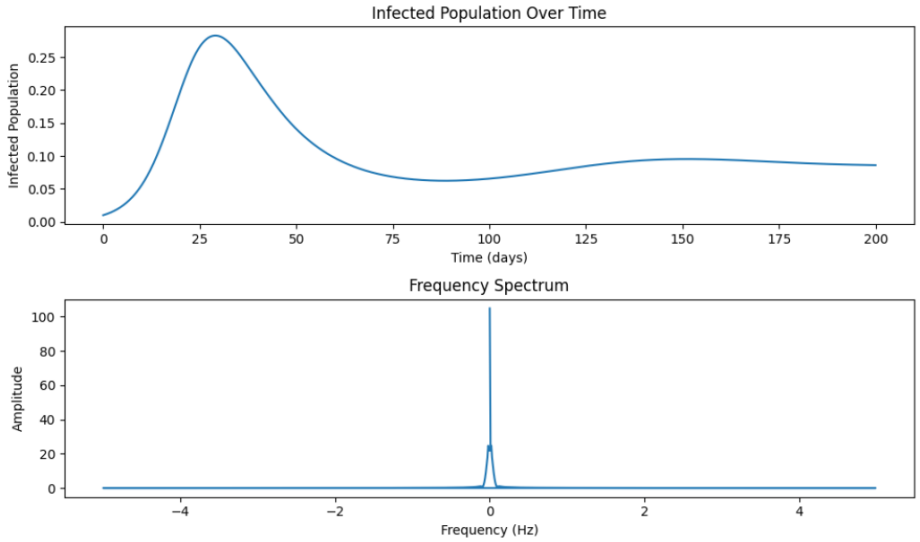

Figure 6: Fourier Analysis

The frequency component at 0 Hz is dominant in the Fourier analysis. It suggests that there is a significant constant or average component in the infected population. In the context of an SIR model with demography, this might indicate that the system has reached a steady state or equilibrium where the number of infected individuals is not changing over time. This is because the infection has become endemic, and the system has settled into a stable pattern. The proof that this is really the case can be seen in Figure 5.

4.2.2 Infection induced mortality term

Infectious diseases often pose a significant mortality risk. To explore the impact of infection-induced mortality in the SIR model, a mortality probability (ρ) for individuals in the infected class is incorporated. Think of ρ as the probability of an individual in the infected class succumbing to the infection before recovering or dying from natural causes. In scenarios of density-dependent transmission, where the total population size (N) decreases due to disease-induced mortality, there is a reduced interaction between hosts.

The adapted ODEs can be seen in the Theory chapter. These ODEs were numerically solved and its result can be seen in Figure 7. All the initial values and parameters are the same as before, only $\rho = 0.3$ has been added. This means that at every time step, an infected person has a 30% chance of dying.

It is clear that the infection spread is much less fierce. Some portion of the infected individuals does not recover but succumbs to the disease. This can be seen in the figure as the decrease of the population size. Also, the infections peak is much lower, since there are less infected people that can give the disease to the susceptible people. Therefore, the amount of susceptible people that are left at the end of the simulation is higher than in the models without infection induced mortality rate. This is also apparent in the Phase Space Diagram, where the fixed point tends more towards the higher susceptible part and the slightly lower infected part, as opposed to the Phase Space Diagram in Figure 6.

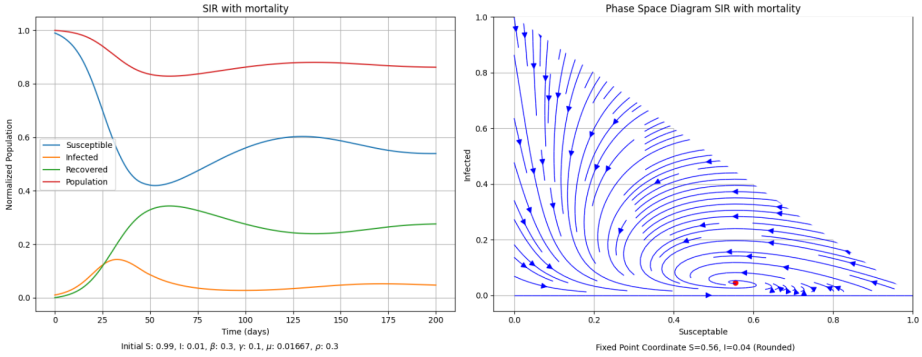

Figure 7: SIR with Demography

If the mortality rate were to approach 1, it significantly alters the dynamics of the model. It means that nearly every individual who gets infected will die. This significantly increases the severity of the disease and makes it highly lethal. In a standard SIR model, individuals move from the infectious compartment to the recovered compartment after recovering from the disease. However, if a high mortality rate is introduced, the infectious period effectively becomes shorter as individuals transition more rapidly to mortality than recovery. The introduction of a high mortality rate also tends to decrease the effective R_0 because a signifi-

cant proportion of infected individuals do not have the opportunity to transmit the infection before succumbing to mortality. The peak of the epidemic is likely to occur earlier than in a scenario without a high mortality rate. This is because a higher proportion of the infected individuals exit the infectious compartment due to mortality, leading to a more rapid decline in the number of infectious individuals. Finally, depending on the parameters of the model, there might be a potential for exponential growth in the number of deaths, especially if the mortality rate is close to 100%.

4.3 Variants of the SIR model

4.3.1 The SEIR model

The specifics of an what an SEIR model is, can be found in the Theory chapter. In this part, the impact of the σ parameter has been examined. The results of this examination are presented in Figure 8. The expectation was that a higher σ means that exposed individuals spend less time in the exposed state before becoming infectious. This effectively shortens the incubation period of the disease. This implies a shorter average duration of the exposed state, meaning individuals move more quickly from being exposed to becoming infectious. Also, with a shorter incubation period, individuals become infectious more quickly after exposure. This can lead to a faster spread of the disease through the population. Moreover, the epidemic curve, representing the number of infectious individuals over time, may show an earlier and steeper rise with a higher σ.

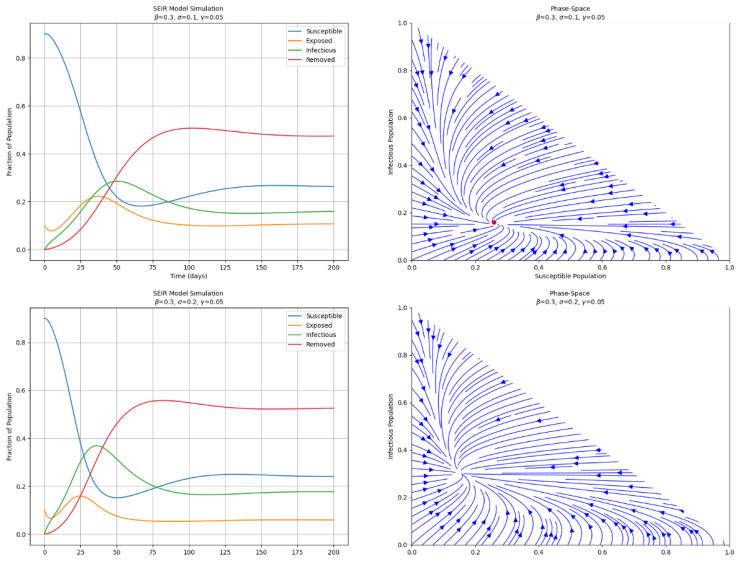

Figure 8: SEIR with differing σ

As can be seen, the expectations were true. Since individuals move more

quickly to being infectious, the line plot of the exposed people reaches its equilibrium more quickly with a higher σ. Also, the infected plot rises more quickly and has a higher peak, implying a faster spread of the disease.

4.3.2 Adding seasonal effects

When introducing a seasonal effect to the SEIR model, it will be accounting for the impact of seasonal variations on disease transmission. Seasonal effects can include factors like temperature, humidity, school schedules, and human behavior that change with the seasons. The impact of these factors can influence the transmission rates of the disease and the dynamics of the population compartments. In this report it has been chosen to focus on Periodic Oscillations, where adding seasonality may introduce periodic oscillations in the model's output, reflecting the cyclic nature of the disease's dynamics. This has been done by changing the β value every time step, based on how far along the sinusoidal track the time currently is. The specifics can be seen in the Theory Chapter.

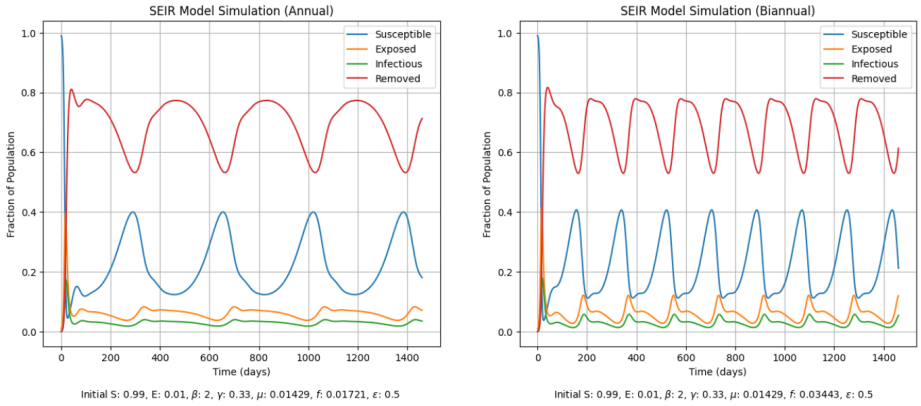

Figure 9: SEIR with seasonality

Here, it can be seen that the seasonal variations occur more frequently within a given time period (e.g., more cycles of within a year). This change can have several effects on the dynamics of the disease model. A higher frequency leads to more rapid and pronounced changes in transmission rates, causing increased variability in disease dynamics. Also, the seasonal patterns in disease transmission, recovery rates, or other parameters exhibits shorter periodic oscillations, reflecting the more frequent occurrence of seasonal variations. Finally, an increased frequency leads to a higher percentage of exposed and infected people.

5 Discussion

5.1 Simplified Models

When discussing model simplicity, it's essential to consider the trade-offs with model complexity. The models that have been used are simplifications of real infectious disease models. The discussion on model simplicity emphasizes that it is not a static attribute but a characteristic that should be continuously evaluated. As new data becomes available and our understanding of the disease and disease modelling evolves, revisiting the balance between simplicity and complexity ensures that the model remains a relevant and effective tool for decision-makers.

5.2 Parameter effects

The S(E)IR model relies on various parameters to describe the dynamics of disease spread. Key parameters examined include the transmission rate, recovery rate, and in case of an SEIR model the progression rate. Having small changes in these parameters could have a pronounced effect on the overall behavior of the model. For instance, variations in the transmission rate significantly influence the rate of infection spread.

5.3 Initial Conditions

The initial conditions of the SIR model represent the starting state of the population and play a pivotal role in shaping the trajectory of the disease outbreak. Through an analysis of the sensitivity of the model to different initial conditions, scenarios where minor changes in the starting state lead to divergent outcomes were observed. This underscores the importance of accurately specifying the initial conditions for reliable predictions and reproductions.

5.4 Assumptions

Throughout this report multiple assumptions have been made in order to generate easily interpretable results. In the reality, factors like immunity, sinusoidal births, noise, stochasticity, meta-population and more complex vaccination factors all contribute to more complex models. Here, most of these factors have been assumed not to be applicable or, at least, to be more simplified. This has to be taken into account when interpreting these results.

6 References

References

[1] J. R. Dormand and P. J. Prince. "A family of embedded Runge-Kutta formulae". In: *Journal of Computational and Applied Mathematics* 6.1 (1980), pp. 19–26.

[2] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[3] Qingkai Kong, Timmy Siauw, and Alexandre M. Bayen. "Chapter 24 - Fourier Transform". In: *Python Programming and Numerical Methods*. Ed. by Qingkai Kong, Timmy Siauw, and Alexandre M. Bayen. Academic Press, 2021, pp. 415–443. ISBN: 978-0-12-819549-9. DOI: https://doi.org/10.1016/B978-0-12-819549-9.00034-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128195499000348.

[4] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7 (1965), pp. 308–13.

[5] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer New York, 2006.

[6] Robert Oshana. "4 - Overview of Digital Signal Processing Algorithms". In: *DSP Software Development Techniques for Embedded and Real-Time Systems*. Ed. by Robert Oshana. Embedded Technology. Burlington: Newnes, 2006, pp. 59–121. ISBN: 978-0-7506-7759-2. DOI: https://doi.org/10.1016/B978-075067759-2/50006-5. URL: https://www.sciencedirect.com/science/article/pii/B9780750677592500065.