Final Project

Analysis of compartmental models in simulating the progression of the COVID-19 disease

SI1336 Simulation and modelling

Raymond Wang

December 18, 2020

Abstract

Compartmental models in epidemiology are the mathematical frameworks wherein the progression of infectious diseases can be modeled after, in particular the COVID-19 pandemic which has been ongoing since the writing of this report. Computer models approximate reality and a qualitative one may be of great use for policymakers in implementing preventive measures. The aim of this study, therefore, is to assess two variants of basic compartmental models and their ability in predicting the progression of COVID-19 in various situations, as well as their applicabilities and constraints.

Key words: Mathematical modelling, epidemiology, computational biology, simulation, data science, COVID-19, research, public health

Contents

1 Introduction

1.1 Background

As of the writing of this report, the COVID-19 disease has been permeating through all corners of the world, causing major disruptions and a devastating pandemic. To alleviate the implementation of preventive measures, policymakers must rely heavily on scientific advice and in this case, computer modeling can be extensively and effectively utilized in predicting the spread of the disease amongst populations.

1.2 Compartmental models in epidemiology

Compartmental models in epidemiology are the mathematical frameworks wherein the progression of infectious diseases can be modeled, utilizing a system of ordinary differential equations with various epidemiological parameters [10]. The population is classified into various segments, simulating the course of the infection of an individual.

1.3 Epidemic SIR-model

The epidemic SIR-model omits vital dynamics (birth and death) and categorizes the population into the segments of **S**usceptible, **I**nfected and **R**ecovered as well as N being the sample population. The model is expressed by the following system of differential equations:

$$\frac{dS}{dt} = -\frac{\beta I S}{N} \ , \qquad \frac{dI}{dt} = \frac{\beta I S}{N} - \gamma I \qquad \frac{dR}{dt} = \gamma I \ , \qquad (1)$$

where β and γ denote the transmission and recovery rates respectively.

1.4 SEIR-model

For a number of infectious diseases and COVID-19 in particular, the pathogen undergoes a long-lasting incubation period wherein the infected individual has a low infectivity rate. In this case, an additional segment, **E**xposed is introduced.

The system of differential equations of this model summarizes to:

$$\frac{dS}{dt} = -\frac{\beta I S}{N} \ , \qquad \frac{dE}{dt} = \frac{\beta I S}{N} - \alpha E \ , \tag{2}$$

$$\frac{dI}{dt} = \alpha E - \gamma I \ , \qquad \frac{dR}{dt} = \gamma I \ , \tag{3}$$

where β, γ and α denote the transmission, recovery and infection rates respectively [4].

1.5   SIS-model

The SIS-model aims to simulate the progression for infections with long-lasting immunity and thus the R-segment is omitted. Although this does not apply for COVID-19, it will later come to great use. The system of differential equations summarizes to:

$$\frac{dS}{dt} = -\frac{\beta I S}{N} \ , \qquad \frac{dI}{dt} = \frac{\beta I S}{N} - \gamma I \ , \tag{4}$$

with the same definition for parameters as in the SIR-model.

1.6   Aims and objectives

The aim of this report is to study and compare the epidemic SIR and SEIR compartmental models in predicting the progression of the COVID-19 disease in different scenarios, with the population of Sweden being used as a reference. Two numerical integrators and their quality in modelling the compartmental models will furthermore be studied.

2 Methods

2.1 Numerical integrators

Being systems of differential equations, the SIR and SEIR models can be solved utilizing various numerical methods. In this case the fourth-order Runge-Kutta method and *odeint* from the Scipy module in Python will be utilized and studied. With the fourth-order Runge-Kutta method being of order accuracy $O(h^5)$, it should in principle yield qualitative approximations of the exact solutions, and being more exact for smaller step sizes. Also, since the analytical solution for the SIS-model is widely accepted and easily derivable [10], it has been utilized for extrapolating the quality of the numerical integrators. Nonetheless, *odeint* was utilized for the ensuing simulations, and the rationale behind it is explained in section 4.5. Python coupled with Matplotlib, Numpy, Scipy etc. are to be utilized.

2.2 Initial conditions

As previously mentioned, the population of Sweden will be utilized as the sample size N, and is numbered at 10 373 225 individuals as of September 2020 [6]. At the beginning of the pandemic and until now, almost none is immune to the disease which implies that everyone is susceptible. Also, assume that there is a trace of infection in the population at e.g. 8 individuals [9]. This yields initial values:

$$S(0) = N, \ E(0) = 0, \ I(0) = 8, \ R(0) = 0 \tag{5}$$

For the parameters:

- $\alpha = 1/5.2$ to reflect a 5.2 day average incubation period [2].

- $\beta = R_0\gamma$ to match the COVID-19 reproduction number of $R_0 = 2-5$ (No social distancing) [7].

- $\gamma = 1/18$ to match an average illness duration of 18 days [2].

Variations are implemented in order to simulate different scenarios.

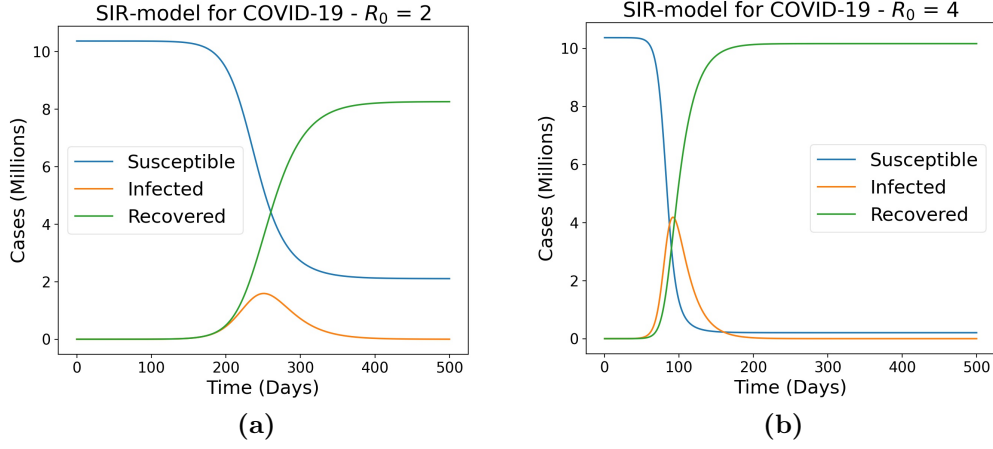# 3   Results

3.1   Epidemic SIR - Comparisons for R_0-rates

Figure 1: (a) $R_0 = 2$, $\gamma = 1/18$ (b) $R_0 = 4$, $\gamma = 1/18$

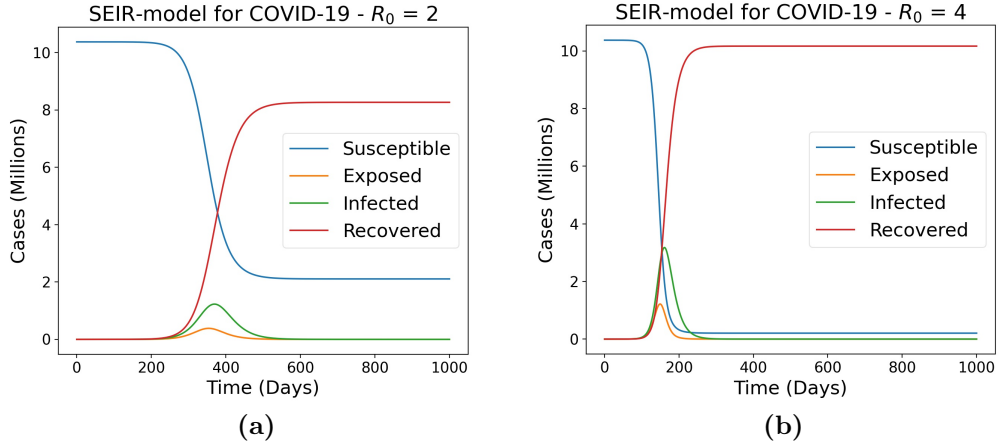## 3.2   SEIR - Comparisons for R_0-rates

Figure 2: (a) $R_0 = 2$, $\alpha = 1/5.2$, $\gamma = 1/18$ (b) $R_0 = 4$, $\alpha = 1/5.2$, $\gamma = 1/18$

3.3 Analysis of the SIR and SEIR-models

As for the epidemic SIR-model, it can be concluded that larger reproduction numbers result in more pervasive pandemics. Peak infections occur earlier and with larger populations. For smaller R_0 values the progression of the pandemic follows a more gradual path, however with less serious consequences, i.e. infections and fatalities.

For the SEIR-model, it can be concluded that the dependence on the reproduction number follows a similar pattern as to its SIR counterpart. However, the progressions of the outbreaks are more gradual due to infections having an incubation time.
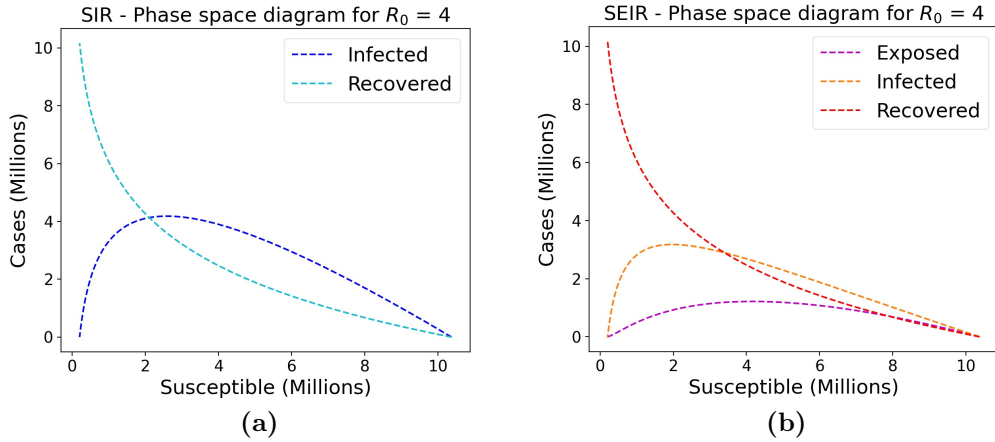
3.4 Phase space diagrams

Figure 3: (a) SIR - $R_0 = 4$ (b) SEIR - $R_0 = 4$

For both models, it can be concluded that an increase in the susceptible population increases in infected cases, while the recovered populations decrease as the susceptible decrease.

The exposed population in the SEIR-model follows the same correlation as infected, however with a more marginal behavior. This also results in the infected population correlation in the SEIR-model being more gradual, which is attributed to the fact that the sum of these should add up to the infected population of the SIR-model.

3.5 Implication of social distancing

In order to analyze the implication of social distancing, a social distancing factor $\rho \in [0, 1]$ can be introduced upon the SEIR-model [4]. As such, equations (2) summarize to:

$$\frac{dS}{dt} = -\frac{\rho \beta I S}{N} \ , \qquad \frac{dE}{dt} = \frac{\rho \beta I S}{N} - \alpha E \ . \tag{6}$$

One may note that in the case when $\rho = 1$, equations (6) are equivalent to equation (2) and this signifies no social distancing involved. By plotting the infected cases and fraction of the susceptible population and varying ρ, the following diagrams are obtained.
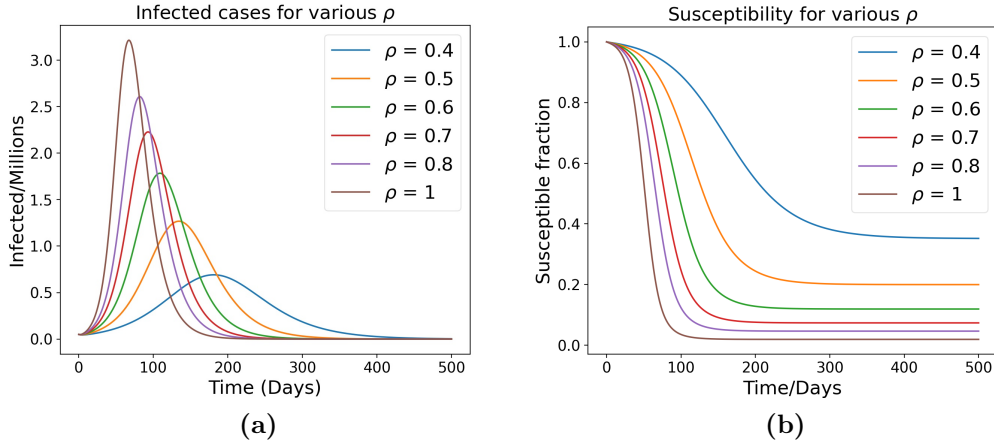

(a) (b)

Figure 4: (a) Infected cases (b) Susceptibility fraction

As can be noted, higher levels of social interactions cause higher peaks for the infected populations as well as the peaks occurring earlier. This can have severe implications such as overwhelming the healthcare capacity. Lesser levels of social distancing prolong the duration of the pandemic, however with less serious repercussions. This phenomenon is commonly referred to as *flattening the curve.*

The fraction of the susceptible population decreases as more time passes and decreases more rapidly for higher levels of social interactions.

This is attributed to the higher number of infected individuals for larger ρ. One may also note that due to this decrease in the susceptible population, a later termination of lockdown measure would result in a lower peak for a resurgent *second wave* and vice versa.

3.6 Implication of vaccination

In order to analyze the implication of vaccination in a population, the proportion of vaccinated population can be introduced upon the SIR model utilizing initial conditions:

$$S(0) = (1-p)(N-1), \ R(0) = p(N-1), \tag{7}$$

for S and R [1]. Note that as $p = 0$, equation (2) is identical to (7) which holds true when no vaccinations are implemented.
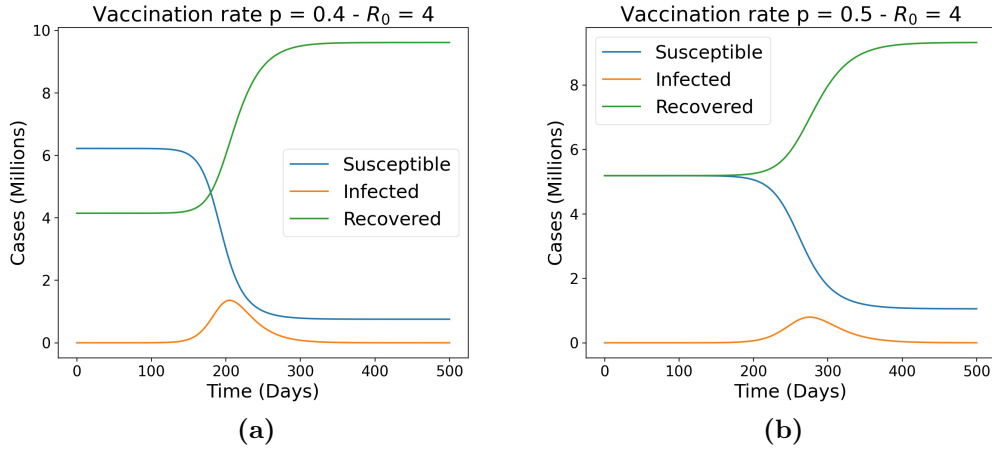
Figure 5: (a) 40 % vaccination rate (b) 50 % vaccination rate

As can be noted, the diagrams for $p = 0$ should be identical to Figure 1b and as the vaccination rate increases, the peak in infected individuals lessens and patients recover earlier. As in previous cases, the pandemic resolves more gradually but with less serious repercussions.

4   Discussion

4.1   SIR and SEIR models

As the reproduction number increases, the more infectious a disease is and causes a larger degree of devastation. It can also be concluded that the SEIR-model is more gradual compared to epidemic SIR and thus more realistically modeling an outbreak. By having only two parameters, the SIR model allows for simple predictions of disease behavior but this simplicity is also its main constraint.

The SIR model does not consider incubation time, i.e. the period in which an individual becomes infected but not contagious. SEIR does however take account of this parameter but additional extensions are required to model in order more desirably match reality. Such extensions are for instance vital dynamics (endemic) and modeling of time-dependent mitigation measures but for simplicity's sake, these have been omitted. However, since pandemics at most times involve quick measures and recovery, vital dynamics have a more minor impact.

Both models make various assumptions, assuming a homogeneous mixing of the population, a closed population as well as equal contact and transmission rates for all individuals in society. One would note that the social structures of humans are far more complex and as previously stated, many more factors must be taken into account. Closed populations, however, can be presumed as of the current day considering the imposed travel restrictions in most countries.

Assumptions are also made in quantifying the parameters, with them only being rough estimations and may vary considerably for populations. This constraint and lack of precision could be alleviated by utilizing distributions for the parameters and thus simulating possible futures.

Finally, one must note that the SIR and SEIR models should only be interpreted as suggestive predictions of the outbreak and when different mathematical models yield qualitatively considerably different results, one may question some assumptions and deduce their inefficiencies. Nonetheless, even a moderately qualitative approximation can be highly useful for informing policymakers.

4.2 Social distancing

It can be noted that social distancing measures are highly effective at curbing the severity of the outbreak and once such mandates are terminated, one would expect a quick resurgence. Longer lockdowns result in less severe second waves due to a decrease in susceptible population over time. This phenomenon can be interpreted as *herd immunity* and even though more social interaction (greater ρ-values) results in a smaller susceptible population, more fatalities and devastation will occur as compared to the more gradual progression for less social interactions (lesser ρ-values).

4.3 Vaccination

Just as social distancing, vaccination is also an effective measure in restraining a pandemic and could even be considered to be more desirable due to individuals becoming immune and falling into the *Recovered* segment immediately. However vaccines take time to develop and do not appear at an instant. One should also note that there is no need for all individuals in a population to be vaccinated and this phenomenon can once again be attributed to *herd immunity*. *Herd immunity* in this case is attained when a sufficient proportion of the population is immune and thus not prone to furthering the transmission.

The proportion of vaccinated population required for basic *herd immunity* can be represented by the formula:

$$p_c = 1 - \frac{1}{R_0},\tag{8}$$

with R_0 being the reproduction number [1]. For COVID-19 with R_0 falling in the range of $2 - 6$, this proportion should fall between $50\% - 83.3\%$ of the population. This corresponds well with Figure 5, where the outbreak essentially dwindles at an approximate 8 million recovered population.

4.4 Statistical accuracy

As previously stated, the analytical solution for the SIS-model will be compared with the solutions generated by the aforementioned numerical integrators. The analytical solution for the SIS-model is given as:

$$\frac{(1 - \gamma/\beta)N}{1 + \big(\frac{(1-\gamma/\beta)N}{I_0} - 1\big)e^{-(\beta-\gamma)t}}, \tag{9}$$

which can be derived by making a transformation of variables [5]. By plotting the absolute differences between the numerical and analytical solutions for variables S and I, the following diagrams were obtained.
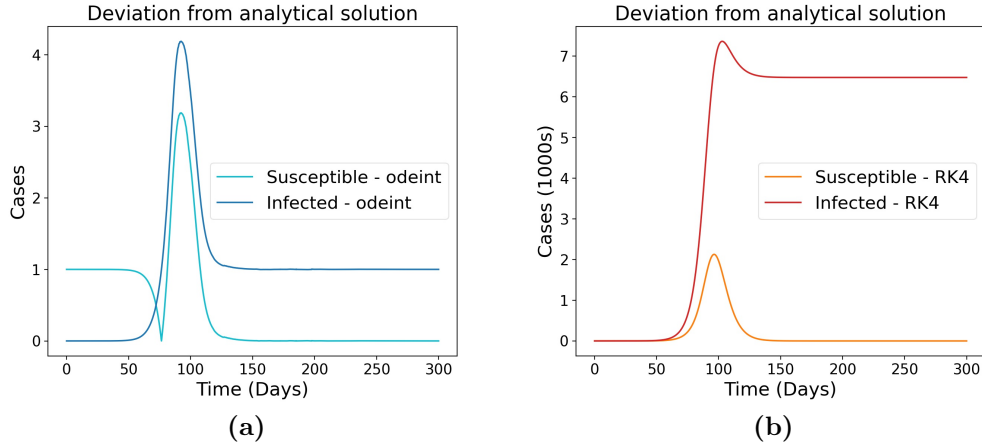

Figure 6: (a) Error margin for *odeint* (b) Error margin for RK4 - $h = 0.01$

As can be noted, *odeint* yields far more exact results than the fourth order RK4 with step size $h = 0.01$ and that is the overwhelming reason for using it as the preferred numerical integrator. For large populations such as in this case, the magnitude difference is negligible and the two integrators both yield qualitative results. However, for smaller populations sizes, *odeint* will undoubtedly outshine its counterpart.

Despite having peak deviations, both methods approximate the susceptible population better than the infected population, attributing to the fact that they converge towards the analytical solution (deviation of 0) for long times.

As for the susceptible populations, both methods show an immature increase in value and converging for longer times, however not towards the analytical solution. Despite the fact that RK4 for the given step size does not attain the same precision as *odeint*, the step size can be further incremented. For $h = 0.001$ and $h = 0.0001$, the following error margins were to be observed.
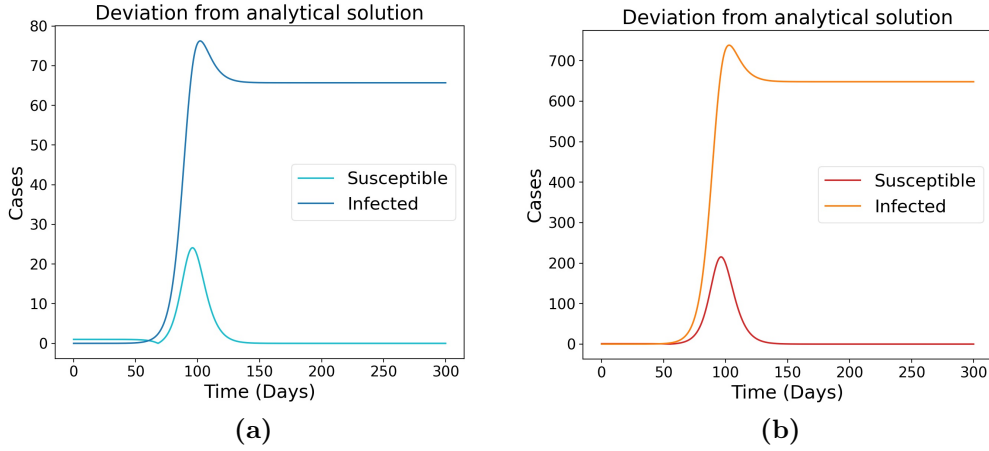

Figure 7: (a) RK4 - $h = 0.0001$ (b) RK4 - $h = 0.001$

As can be noted from the diagrams, the error margin for RK4 decreases for smaller step sizes and should attain the same magnitude as in *odeint* and be comparably precise. However, this comes at the expense of the computational time, with *odeint* being considerably faster for margins of error of the same magnitudes.

According to the Scipy library, there is no simple description of the algorithmic complexity of *odeint*, which is based on an *lsoda* integrator from FORTRAN. The *lsoda* integrator alternates between a stiff BDF and non-stiff Andrews method, which changes step size in accordance with internal error estimates [8]. This should be the primary reason for the precise and qualitative behavior shown by *odeint* in solving the compartmental models.

5   Conclusions

It can be concluded that the SIR and SEIR are simplified representations of reality and due to this simplification, a huge amount of constraints are bestowed upon them. Although being so uncomplicated, the compartmental model does an excellent job in giving a rough approximation of possible outcomes of an outbreak, with SEIR being the more realistic of them. The compartmental models are highly flexible so by introducing more parameters and variables, they should better correspond to reality.

One could also note that terms such as *herd immunity* and *flattening* the curve may initially seem vague but with mathematical visualizations, these phenomena can be simulated by varying the parameters. As such computer simulations and mathematical modeling are powerful tools in assessing the outcomes of various phenomenons in reality.

The SARS-CoV-2-virus has spread extremely hastily and the COVID-19 pandemic is often compared to other recent pandemics, such as the SARS-outbreak in 2002-2004. Despite being less lethal, COVID-19 is much more infectious than SARS and that can be explained by comparing their reproduction numbers, with SARS falling within the range of 0.19-1.08 [3]. This demonstrates that by varying R_0, one could sufficiently assess the infectivity for various pathogens.

Some degree of chaotic behavior can be noted for both the SIR and SEIR models with and without modifications, considering that variation of the model parameters, α, β, γ, ρ could result in substantial alterations for the S and I trajectories if large enough. However, the effect is not substantial for small variations and thus it can be argued that the models are not chaotic in nature.

As for the numerical integrators, it can be included that both *odeint* and the fourth-order Runge-Kutta methods yield qualitative approximations for the analytical solutions, with *odeint* having an advantage over its counterpart. RK4 has a constant but high order accuracy of $O(h^5)$ but from the simulations, it seemingly appears that constantly altering the order accuracy for internal error estimates is a more effective course of action. Although no order accuracies for the *odeint* are listed, it can be concluded that it should be higher than RK4 for some instances.

6 References

[1] Najma Ashraf. *Extending the Basic SIR Model*. July 2020. URL: https://towardsdatascience.com/extending-the-basic-sir-model-b6b32b833d76.

[2] Andrew Atkeson, Karen Kopecky, and Tao Zha. "Estimating and Forecasting Disease Scenarios for COVID-19 with an SIR Model". In: (2020). DOI: 10.3386/w27335.

[3] Gerardo Chowell et al. "Model Parameters and Outbreak Control for SARS". In: *Emerging Infectious Diseases* 10.7 (2004), pp. 1258–1263. DOI: 10.3201/eid1007.030647.

[4] Stephen Gill. *How computer modelling has been used during the coronavirus (COVID-19) outbreak to inform political decisions*. URL: https://www.mewburn.com/news-insights/how-computer-modelling-has-been-used-during-the-covid-19-outbreak-to-inform-political-decisions.

[5] Simon A. Levin, T. G. Hallam, and Louis J. Gross. "119–144". In: *Applied mathematical ecology*. Springer-Verlag, 1989.

[6] *Population statistics*. URL: http://www.scb.se/be0101-en.

[7] Steven Sanche et al. "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2". In: *Emerging Infectious Diseases* 26.7 (2020), pp. 1470–1477. DOI: 10.3201/eid2607.200282.

[8] *scipy.integrate.LSODA*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.LSODA.html.

[9] David Smith and Lang Moore. *The SIR Model for Spread of Disease - Introduction*. URL: https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-%20disease-introduction.

[10] Henrik Söderheim. *Disease Spreading Simulations*. Feb. 2019. URL: https://github.com/henriksoder24%20/SI1336%20-Simulations-Modelling.