



Delft University of Technology
Faculty of Electrical Engineering, Mathematics & Computer Science
Delft Institute of Applied Mathematics

Modelling COVID-19 in The Netherlands

Master thesis submitted to the
Delft Institute of Applied Mathematics
as part to obtain

the degree of

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

DANIEL TSANG

Delft, The Netherlands
September 2021



MSc Thesis APPLIED MATHEMATICS

**“Modelling COVID-19
in The Netherlands”**

DANIEL TSANG

Delft University of Technology

Supervisor

Dr. B.J. Meulenbroek

Other members of the committee

Dr. J.L.A. Dubbeldam

Dr. D. Kurowicka

Thesis Defense Date

21 September 2021

Abstract

In this thesis, we present a study to obtain a clear and accurate overview of the progress and behaviour of COVID-19 in the Netherlands. We distinguish two parts for this study. The first part is to estimate the total number of infected people as a function of time by combining data from hospital admissions, daily reported cases and serological data. Using these data sets, we found that our estimation for the number of infected people was comparable to the estimations provided by the RIVM and Sanquin. Furthermore, we found that on average only 39.3% of the total number of cases were detected. 1.2% of the total number of infected people is admitted to the hospital and 18.6% of the hospitalized patients is admitted to the ICU.

The second part is to develop a representative model that reproduces the estimated total number of infections using a modified SEIR model. These modifications include modelling the infection rate $\beta(t)$ as a function of time using a simple linear ODE, a system of ODEs inspired by the Lotka-Volterra equations, the implementation of gamma distributed exposed and infected stages and lastly the incorporation of spatial heterogeneity. We found that our Lotka-Volterra inspired model was able to model multiple consecutive waves, which differs from the standard compartmental models. The other modifications however seemed to have only minor effects on the model and had some difficulties with matching historical data. We conclude that our Lotka-Volterra inspired model should be used to model consecutive waves for a longer period of time. The other modifications can be used to optimize the model.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 2 | The SEIR Model | 8 |
| 2.1 | Definitions and concepts of the SEIR model | 8 |
| 2.2 | Simplifications of the SEIR model | 11 |
| 2.3 | Possible solutions to the simplifications | 12 |
| 2.4 | Conclusion & Discussion | 14 |
| 3 | Estimating the number of infected persons in the Netherlands | 15 |
| 3.1 | Difference in data | 16 |
| 3.2 | Estimation methods | 17 |
| 3.3 | Sanquin data | 19 |
| 3.4 | Results | 21 |
| 3.5 | Conclusion & Discussion | 28 |
| 4 | Initial values and first results with the SEIR model | 29 |
| 4.1 | Values for γ and α and initial conditions | 29 |
| 4.2 | Values for β_{high} and $\beta(t_0)$ | 30 |
| 4.3 | Preliminary results | 31 |
| 5 | Lotka-Volterra equations for $\beta(t)$ | 34 |
| 5.1 | Lotka-Volterra and the SIR model | 34 |
| 5.2 | Second adjustment of the SIR model | 38 |
| 5.3 | Lotka-Volterra and the SEIR model | 44 |
| 5.4 | Conclusion & Discussion | 47 |
| 6 | SEIR model with gamma distributed stages | 49 |
| 6.1 | The model equations | 49 |
| 6.2 | Results with first ODE for $\beta(t)$ | 50 |
| 6.3 | Results with Lotka-Volterra | 54 |
| 6.4 | Conclusion & Discussion | 57 |
| 7 | The SEIR model and spatial heterogeneity | 58 |
| 7.1 | Model equations for the subregions | 61 |
| 7.2 | Results of the spatial SEIR model | 63 |
| 7.3 | Conclusion & Discussion | 68 |
| 8 | Discussion | 69 |
| 9 | Conclusion & Outlook | 71 |
| A | Reproduction Factor | 75 |
| A.1 | Calculating the reproduction factor | 75 |
| B | Determining Reproduction Factor | 80 |
| B.1 | Continuum model | 80 |
| B.2 | Simplifying assumptions | 81 |
| B.3 | Computations | 82 |

1 Introduction

The outbreak of the COVID-19 virus at the end of 2019 has led to a worldwide crisis which still lingers on today. With a quick search on the internet, we see that globally over 194 million people have been infected, and over 4 million people are deceased up until now. In the Netherlands, there are over 1.8 million people who have been infected and almost 18 thousand people are deceased. Note that the actual numbers of infections and deceased people are higher, as it is not possible to track all infected (and dead) cases.

One of the aspects of COVID-19 is that the virus is transmitted mainly through small droplets or aerosols when an infected person coughs, sneezes or talks. These droplets or aerosols carry the virus in the air, and standing nearby an infected person may lead to the inhalation of the virus; thus becoming infected as well. Therefore, since the outbreak of the virus, we have seen that the virus can spread rapidly among people, especially in confined spaces. We have also seen that COVID-19 affects different people in various ways. Most infected people develop mild symptoms such as sneezing, (dry) cough, tiredness and fever. Some infected people do not even notice their symptoms at all, and as a consequence, they are not aware that they are infectious to other people. However, some people develop severe symptoms such as shortness of breath and breathing problems. As a consequence, these people need to be admitted to the hospital or intensive care. And it is also known that the duration of their hospital admission is relatively long (about two weeks). And this means that comparatively low number of patients can quickly fill up a large portion of the hospital capacity. This is what happens in the Netherlands (but also in other countries). We have seen that the hospital beds (including intensive care units) were rapidly occupied by COVID-19 patients - increasing the pressure on hospitals and healthcare (personnel). Because of this increase in pressure, many other patients have their (regular) healthcare being delayed or even cancelled for an unknown period of time, such as knee operations or cancer treatments. What the consequences are of the delayed healthcare is a major question on its own.

In order to contain the virus and elevate the pressure on the healthcare, the government decided to take different measures such as (self-) quarantine and isolation, distance keeping or temporary closing of restaurants, shops, gyms and the (whole) educational system. Taking different measures however does come with consequences on health, economic and societal areas. For example, the stay-at-home policy combined with the closure of sports activities lead to worse diets and less physical activity, thus causing all kinds of health issues. Shopkeepers and restaurant owners (almost) lost their entire income due to closing. And the prohibition of social activities contributes to the lack of social contacts among people. These examples show that the government has to consider which and when certain measures must be taken. Good and effective measures can only be taken if one has an accurate and clear overview of the progress and behaviour of the virus throughout the country and the expected impact of a proposed measure on this behaviour.

In order to obtain this overview, a number of steps is needed. The first step is to acquire data about the number of hospital (and ICU) admissions and the number of positive cases per day. However, acquiring data alone is not enough for a complete overview, because the data itself may be incomplete, and simply not all people can be detected. By combining serological data and hospital admissions we can estimate the total number of infected people and compare this to the number of positive tests and thus we obtain a more reliable picture.

As a second step, we need to have a good model that accurately describes the progress and behaviour of the epidemic, i.e., it matches the historical data and also predicts the future developments reasonably well.

A third step that leads to good and effective measures is to have an overview of the benefits and costs of each measure. As it was mentioned above, each measure affects the lives of people in different aspects. So then the question arises whether or not the benefits of each measure outweigh the costs that come along. With an accurate model, we can understand how the virus is spreading, and it makes possible to incorporate measures into the model such that the effects of those measures can be monitored and predicted. Such a complete model helps policymakers to weigh costs and benefits beforehand instead of taking measures and hoping for the best¹.

¹"With 50 per cent of the knowledge, we have to make 100 per cent of the decisions" - Mark Rutte, Prime Minister of the Netherlands.

Up until now, we see for the first step that the Corona Dashboard of the Netherlands had shown that the RIVM² had estimated the total number of infected people as a function of time with regular updates. We will see in Section 3 that our estimation is close to the estimation of the RIVM. Thus, in both ways we have a reliable picture of the total number of infected people.

Although the second step is still in progress, we have seen that the RIVM has given technical briefings to the government on a regular basis about the COVID-19 situation in the Netherlands. In these briefings, they inform the government about the current situation and they show results of their models, including future predictions when different measures are taken. So we can see that there is much (media) attention on step two as well. However we note that there is still room for improvement in the second step.

Finally, we notice that in the Netherlands the third step does not receive as much attention as the previous two steps when it comes to deciding what measures should be taken. Perhaps partly due to the complexity of making an overview of the benefits and costs. In this thesis, we will focus on the first two steps as well, i.e., finding good estimates for the number of infected people as function of time, developing a model that reproduces these numbers (as reliably as possible) and we leave the third step for future research.

Often times in mathematical epidemiology, one uses compartmental models to model the infectious diseases. These models subdivide the population in different compartments, and the transmissions between these compartments are described by a set of differential equations (DEs) and its model parameters. One of the simplest compartmental models is the SEIR model, where S stands for the number of susceptibles, E the number of exposed individuals, I the number of infected individuals and R the number of recovered (or removed/deceased) individuals. Each individual is considered to be in one compartment at a given time, but it is possible to move from one compartment to another. In case of the SEIR model, individuals can move from compartment S to E , E to I and from I to R .

One of the advantages of using compartmental models is that such models are relatively easily modified and expanded with more or different compartments. However, the addition or altering of compartments also means the addition or altering of the model parameters that need to be estimated. Estimation of the model parameters is not a trivial task. Some, if not, all of the model parameters also have physical meanings that need to be taken into account. For example, in the SEIR model, we will see in this thesis that we define the parameter γ as the inverse duration of the infection. But the value of the duration of the infection differs among various studies (see Table 1). Another important parameter is the infection rate β . This parameter is based on the average number of contacts a person has each day and the probability that such a contact leads to an infection. In this thesis, we will consider β to be a function of time and we primarily focus on the modelling of β .

A large number of studies on the modelling of COVID-19 has been published. In this thesis we mention a selected number of studies that seems most relevant for our research questions. These studies have applied a compartmental model to model COVID-19 in various countries and cities. From these papers, we can see what modifications are used to certain compartmental models, how these models match with historical data, what measures can be incorporated into these models and how they can predict future developments. We note however that developing an accurate model for COVID-19 is an ongoing process that will take a significant amount of time, even when the epidemic is over.

Peng et al.(2020) [1] proposed a generalized SEIR model with the inclusion of compartments P (in-susceptible) and Q (quarantined) and D (death) to model COVID-19 in China. For the parameters, they assumed the cure rate λ and the mortality rate κ to be time-dependent, whereas the infection rate β is a fixed value. During parameter estimation, they wanted to overcome overfitting problems by choosing a fixed value for the latent time γ^{-1} , and then for each fixed γ^{-1} , they explored its influence on other parameters. By doing so, they found that the parameter β is very close to 1. They applied their model based on the data from January 20th 2020 to February 16th 2020, where their study included 24 provinces in China and 16 counties in Hubei province. Then, they carried out simulations for a longer time to forecast the course of the epidemic and compared the simulations with official data from February 17th to March 23rd. They expected that for most parts of China the epidemic will end no later than the middle of March 2020. And for Wuhan city, they expected that the severe situation back then will end up at the beginning of April 2020. Their forecasts were in well agreement with the real situation.

Godio et al.(2020) [2] adopted a generalized SEIR model from Peng et al.(2020) [1] where the com-

²Het Rijksinstituut voor Volksgezondheid en Milieu' (National Institute for Public Health and the Environment)

partments P (protected category), Q (quarantined people) and D (dead people) were added to model COVID-19 in regions of Italy, Spain and South-Korea. Model parameters were fitted in least-square sense with a deterministic approach and with a stochastic approach using a Particle Swarm Optimization (PSO) algorithm. They assumed a time-dependent β that is based on the contact rate calculated using mobility data. We have seen that the PSO algorithm provides a set of possible solutions, where each solution refers to a different set of parameters estimated by the algorithm; whereas the deterministic approach only gives one solution. The Italian and Spanish data were well fitted in the period of 1 March 2020 to 15 April 2020, and they showed the predicted evolution in 30 days according to their SEIR model. The fitting of the Korean data was not ideal, but that is also due to the oscillations in the data itself. They recommend fitting the data of the epidemic using a stochastic approach such as the PSO method, and the influence of the time-varying infection rate β on the model prediction may open interesting discussions about the effect of lockdown policies on the evolution of the epidemic in the future.

The SIR model was modified by Anand et al.(2020) [3] by adding the compartments quarantined Q and unquarantined UQ . The factor of testing ft , which stands for the fraction of infected people who are tested and quarantined is introduced. If the level of testing is increased and more number of infected people are tested, than the number of unquarantined people will reduce, thereby controlling the spread of the disease. They applied their model to the situation in the state Kerala, India. Parameter estimation has been done by formulating a cost function and using a differential evolution optimizer. After that, they compared the model predictions with the actual data, starting from March 8th 2020 to April 26th 2020. The model matched with the actual data, and they simulated several scenarios after April 26th 2020. After the simulations, they suggested 4 protective measures to control the pandemic in Kerala. First: lift lockdown, but rapidly increase testing. Second: extend lockdown but gradually increase testing. Third: intermittent or staggered lockdown, while the level of testing is the same. Fourth: lift lockdown while ensuring adequate social distancing measures.

Ansumali et al.(2020) [4] reviewed the basic SIR and SEIR models and analysed their properties such as Lyapunov stability, both with and without vital dynamics (births/deaths). After that, they introduced the SAIR model, where A stands for the asymptotic people. These people do not have symptoms, but can still infect people from the susceptibles group (which is the case with COVID-19). The difference between the SAIR and the SEIR model is that it is assumed that the exposed people are not infectious yet. Then, they examined the SAIR model both with and without vital dynamics, and analysed their stability properties. Three parameters (γ , δ and β) are estimated in analytical ways. After that, numerical solutions are shown using the estimated parameters for the countries Switzerland, Japan, France and Italy (although parameter estimation was also done for USA, Brazil, India and Iran). Their model matches well with the real data of those countries. The same methods are applied to the city Delhi, India, which was a hotspot in India back then. They found that future predictions of the progress of the disease show that for the SAIR model, herd immunity is achieved when the number of asymptotic and infected persons is around 25% of the population, whereas the classical SIR model would predict that herd immunity is achieved at 64.29%. However, there is no explicit expression for the onset of herd immunity in the SAIR model yet.

The RIVM also used a compartmental model³ for COVID-19 in the Netherlands where vaccination strategies are included. They included compartments such as IC (people in intensive care), H (hospitalized) and HIC (return to the hospital ward following treatment in IC). Different vaccination strategies were assessed. These were: old to young (vaccination begins with 50-59 year olds and then progresses through in decreasing order), young to old (vaccination starts from 18-19 year olds and then progresses through in increasing order), alternative (18-29 year olds followed by 50-59 year olds, then progresses to 40-49 year olds and then 30-39 year olds) and no vaccination. They found that regardless of the vaccination strategy, implementing a vaccination program results in fewer new infections, new cases, hospital admissions, IC admissions, new deaths and fewer life years lost compared to no vaccination program. The overall differences between these vaccination programs were very little, however the old to young strategy resulted in the smallest number disease outcomes (e.g infections, cases, hospital admissions). The studies above are examples of how compartmental models can be applied to real-life cases, and how such models can be a useful tool to help policymakers in their decision making (RIVM).

³https://www.rivm.nl/sites/default/files/2021-03/Modellingresults%20COVID19%20vaccination%20version1.0%2020210324_0.pdf

Although the studies have provided some new insights, we also mention at least three deficiencies. First, the matching of historical data and the predictions are done in a relatively short period of time (a few months). For short term policies it is a good way to start, however it would be even better if one can model future developments in the long run (1-2 years).

Second, compartmental models such as SEIR usually show only one peak of infections, followed by a decline of infections. This is what most of the results from the studies show as well. We observe however that already four waves of infections occurred in the Netherlands up until now. We want to be able to capture the different infection waves into the model.

Lastly, there is little to no attention to spatial heterogeneity, which will be explained below. So in this thesis, we provide three new modifications to the standard SEIR model to model COVID-19 in the Netherlands. These modifications are described in Section 2, but we will briefly mention them here.

The first modification is to develop a time-dependent function for the infection rate $\beta(t)$ because the contact rate of an individual changes when protective measures are applied. Different studies have provided a time-dependent expression for the infection rate. For example, Lin et al.(2020) [5] included a stepwise function that represents the governmental actions. The values of this stepwise function are dependent on the strength of the lockdown. And Kaushal et al.(2020) [6] defined the infection rate with use of a Heaviside function that represents the lockdown policy. They considered a lockdown as a sudden change in the infection rate. In our modification, we will present two ideas that describe the infection rate as an ordinary differential equation (ODE), where one of them is inspired by the Lotka-Volterra equations. The second modification is to implement a non-exponential distribution for the infectious time period and the incubation time period. The standard SEIR model assumes that these time periods are exponentially distributed, however that assumption is not realistic (Lloyd (2001) [7], [8]). Instead of the exponential distribution, we will use the gamma distribution that describes the incubation time and the infectious time period (Feng et al.(2007) [9]).

The third modification is to incorporate spatial heterogeneity into the model. The standard SEIR model assumes that every individual in the population interacts with one another. However when it comes to country-sized populations, there are parts of the population that are less likely to interact with one another (e.g. one from Amsterdam and one from Groningen). Moreover, the virus is more likely to spread faster in densely populated areas such as cities than in rural areas. So we will present a general model that can take account these observations.

At last, we will also provide estimation methods that estimate the total number of infected persons per day (undetected cases included). This number is useful to know for policymakers in the battle against the virus. Böhning et al.(2020) [10] provided a method to estimate the total number of infections. Their method is based statistical methods where a modified version of Chao's estimator is introduced. This links to a capture-recapture approach. They applied their method to the situation in Austria and other European countries, where data they used are the cumulative count of infections and the count of new deaths in the period of March 15th 2020 to April 7th 2020. We will see in our methods that we make use of the data from hospital admissions and the reported positive cases.

The organization of this thesis is as follows. In Section 2, we will explain the definitions and concepts of the SEIR model. We will also see that the SEIR model has its (over)simplifications, and we present multiple solutions to these simplifications. Next, we will estimate the number of infected individuals in the Netherlands using the data from the RIVM and compare it with other sources in Section 3. In Section 4, we will derive the initial values and initial conditions for the SEIR model. Finally, in Sections 5, 6 and 7, we will explain the different solutions to the simplifications of the SEIR model in more detail, and we will show numerical results of these solutions. We will end this thesis with the discussion, conclusion and outlook.

| Study | Location | Parameter (days) | Central tendency reported | Variation (days; inclusion) |
|--------------------------------|--|--|---------------------------|-----------------------------|
| Tracking studies | | | | |
| He et al. (2020) [11] | Vietnam, Malaysia, Japan, China, Taiwan, USA and Singapore | 19.3 days (Infectiousness) | Mean | 12.9 - 24 days |
| Ma et al. (2020) [12] | Various countries | 5 days (serial interval - upper limit of latent period) | Median | Range 0 - 24 days |
| Modelling studies | | | | |
| Li et al. (2020) [13] | China | 3.47 (posterior estimation from model) | Median | 3.26 - 3.67 |
| Tuite et al. (2020) [14] | Canada | 6 days (fixed parameter within a deterministic model) | Fixed parameter | |
| Lourenço et al. (2020) [15] | UK | 3 - 5 days (Posterior output, depending on scenario tested) | Mean | 3 - 6 days |
| Zhu et al. (2020) [16] | Wuhan, China | 12.53 days (posterior estimated from model with Weibull distribution) | Mean | Variance of 11.4 days |
| Davies et al. (2020) [17, 18] | UK | 5 days (used as a prior, drawn from a gamma distribution) | Mean | |
| Piccolomini et al. (2020) [19] | Italy | 20 days (prior, obtained from data estimations and clinical observations) | Fixed parameter | |

Table 1: Table of studies with their given infectious periods from Byrne et al.(2020) [20], either from modelling or tracking studies. Infectious period from tracking studies is inferred from patient histories. Infectious period from modelling studies is reported as a prior (assumed parameter value) or an posterior estimate.

2 The SEIR Model

In this section, we will explain the basic concepts and definitions of the SEIR model. Also, we will explain the different (over)simplifications of an SEIR model. It is important that we understand the concepts and ideas of such compartment models, because epidemics are often modelled using an SEIR model, or a similar compartment model. Moreover, by knowing and understanding the simplifications of such models, we can adapt the SEIR model to obtain a more realistic model.

First, we will show the equations of the SEIR model and explain its definitions. After that, we will discuss a few simplifications that are part of most SEIR models. And finally, we will show possible solutions in order to improve these models.

2.1 Definitions and concepts of the SEIR model

During modelling of an epidemic, the population in a basic SEIR model is divided into 4 compartments/stages (see Figure 1):

1. Susceptibles (S): This group represents the people that have not been infected, but are susceptible to the disease when they get in contact with infected people (I).
2. Exposed (E): The people in this group have been infected by the virus, however the symptoms are not to be seen and the people cannot infect other people yet. The time between an infection of a person and symptom onset is called the incubation time.
3. Infected (I): In this group, the people have already been infected and the symptoms have emerged. Also, these people are able to spread the disease to the people in the susceptibles group (S) as long as they are infected. The time period in which these people are infectious is called the infectious time.
4. Recovered/Resistant (R): After the infectious time, an infected individual recovers from the disease and becomes resistant to the disease as well. These people are then not able to infect other susceptibles (S) anymore. Sometimes this group also includes people who were deceased, since there is no difference between a deceased and a resistant person from a mathematical viewpoint.

The transition between the groups is modelled using the law of mass action from chemistry. The law of mass action states that *the rate of reactions is proportional to the active concentrations of the reactants*. We can illustrate and make clear the meaning of the law by considering the following example. Suppose we have an irreversible reaction described by



The number of the reactants are denoted by A , B and C . Now, the change in the number of C molecules is determined by the number of collisions between A and B molecules, the probability that such a collision will lead to the product C and the duration of the reaction. This can be formulated as

$$\underbrace{\frac{\Delta C}{\Delta t}}_{\text{Change in number of } C \text{ molecules.}} \approx \underbrace{k}_{\text{Probability that a collision between } A \text{ and } B \text{ molecules produces molecule } C \text{ per unit time.}} \cdot \underbrace{A}_{\text{Number of } A \text{ molecules.}} \cdot \underbrace{B}_{\text{Number of } B \text{ molecules.}} \cdot \underbrace{\Delta t}_{\text{Duration of the reaction.}}$$

where k is also called the rate constant and t denotes time. Rewriting the equation we will have

$$\frac{\Delta C}{\Delta t} \approx k \cdot A \cdot B.$$

Taking the limit of $\Delta t \rightarrow 0$, we obtain the differential equation

$$\frac{dC}{dt} = k \cdot A \cdot B.$$

This is called the law of mass action. At the same rate, the number of A molecules decreases when the reaction takes place:

$$\frac{dA}{dt} = -k \cdot A \cdot B.$$

From these equations, we see that with a higher number of molecules of A or B (or both), the reaction takes place at a faster rate. On the other hand with a lower number of molecules of A or B (or both), we have a slower rate of the reaction.

In case we have the reaction



then the change in the number of C molecules is determined by the number of A molecules, the probability that a molecule of A decays to a molecule of C and the duration of the reaction. This can be formulated as

$$\Delta C \approx \underbrace{k^*}_{\substack{\text{Probability that a molecule} \\ \text{of } A \text{ decays to a molecule of } C \\ \text{per unit time.}}} \cdot A \cdot \Delta t.$$

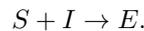
Similarly as before, we obtain the differential equation

$$\frac{dC}{dt} = k^* \cdot A,$$

and

$$\frac{dA}{dt} = -k^* \cdot A.$$

Now in the case of the SEIR model, we assume that an individual in group S can get infected by an individual in group I when they are in contact with each other. If so, the individual from S becomes exposed and is thus in group E . The corresponding 'chemical reaction' is then given by



Next, after a few days, the exposed person becomes infectious, meaning that the person from group E moves to group I . The reaction is given by



After a while, we assume that the infectious person recovers, and is unable to infect other people anymore. Similarly, the reaction is



Combining all the reactions, we have that



Using the law of mass action, we can write down a system of ordinary differential equations (ODEs) for this whole reaction:

$$\begin{aligned} \frac{dS}{dt} &= -k_1 SI, \\ \frac{dE}{dt} &= k_1 SI - k_2 E, \\ \frac{dI}{dt} &= k_2 E - k_3 I, \\ \frac{dR}{dt} &= k_3 I, \end{aligned}$$

where k_1, k_2, k_3 are the rate constants.

In terms of mathematical epidemiology, we set $k_1 = \beta \frac{1}{P_{tot}}$, $k_2 = \alpha$ and $k_3 = \gamma$. Then, we have the

well-known basic SEIR model:

$$\frac{dS}{dt} = -\beta \frac{SI}{P_{tot}}, \quad (1)$$

$$\frac{dE}{dt} = \beta \frac{SI}{P_{tot}} - \alpha E, \quad (2)$$

$$\frac{dI}{dt} = \alpha E - \gamma I, \quad (3)$$

$$\frac{dR}{dt} = \gamma I, \quad (4)$$

where P_{tot} denotes the population number. Next, we will associate physical meanings to the parameters α, γ and β .

- We will start with the parameter γ . From our example earlier, we can approximate the equation

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

by taking the limit of $\Delta t \rightarrow 0$ from the equation

$$\Delta R \approx \gamma \cdot I \cdot \Delta t.$$

Here, $\gamma \cdot \Delta t$ can be interpreted as the probability that an individual moves from I to R within Δt time, where the outcome per unit time is either that the person is recovered or not recovered yet. An analogy for this is to think about a coin flip where the first occurrence of heads (recovery) requires m independent trials (Δt). Now, we define the period of time in which a person is infectious as t_{inf} . After the infectious time, the person is then recovered. So if $\Delta t = t_{inf}$, we must have

$$\Delta R \approx I,$$

and

$$\gamma \cdot t_{inf} \approx 1,$$

because (almost) all individuals move from I to R within t_{inf} time with probability 1. Thus, it follows that

$$\gamma \approx \frac{1}{t_{inf}} = \frac{1}{\text{Infectious time}}. \quad (5)$$

- A similar procedure can be done for the parameter α , where we look at the transmission of a person from E to I . Thus, we look at the equation (omitting the last term)

$$\frac{dI}{dt} = \alpha E.$$

Defining the incubation period as t_{inc} and setting $\Delta t = t_{inc}$, it follows that

$$\alpha \approx \frac{1}{t_{inc}} = \frac{1}{\text{Incubation time}}. \quad (6)$$

- For the parameter β , we will look at the transmission of a person from S to E . This is given by the approximated equation

$$\Delta E \approx \beta \cdot S \cdot \frac{I}{P_{tot}} \cdot \Delta t.$$

We define the parameter β as

$$\beta := \text{Total number of contacts per person per day} \times \text{chance of infection per "contact"}. \quad (7)$$

where "contacts" can be either direct or indirect (e.g. by touching a contaminated surface). If t is in days with $\Delta t = 1$, then the terms to model ΔE have the meaning

$$\begin{aligned} \beta \cdot S \cdot \frac{I}{P_{tot}} \cdot \Delta t := & \text{Total number of contacts of the susceptibles} \times \text{chance of infection per contact} \\ & \times \text{chance that a contact was with an infectious person.} \end{aligned} \quad (8)$$

The number of contacts can be estimated, however the chance of infection per contact is unknown. That is why we will consider β as an unknown parameter. The parameters α and γ can be estimated from the data of the infected individuals by looking at the average incubation and infectious times from the patient data.

Note that if we add all the equations of the SEIR model, we have

$$\frac{d}{dt}(S + E + I + R) = 0 \Rightarrow S(t) + E(t) + I(t) + R(t) = P_{tot}(t) = P_{tot}(0). \quad (9)$$

This means that the total population is constant. No new susceptibles are added (e.g. no births or inflow from outside sources) and people who are deceased are merged into the R -group. The transitions between the groups can be visualised using a simple flow chart.

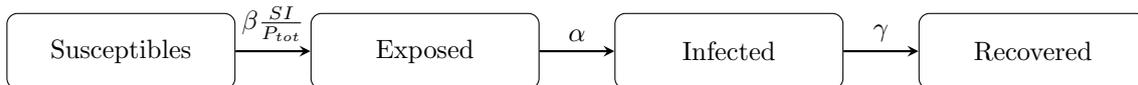


Figure 1: Transitions between the 4 groups in a basic SEIR model.

The parameters above the arrows in Figure 1 are called the transition parameters, and they determine the average rate in which the individuals move from one group to the other.

2.2 Simplifications of the SEIR model

In the case of the spread of COVID-19 through the Netherlands, such an SEIR model has the following systematic (over)simplifications:

1. In a basic SEIR model, the parameter β is usually set to a constant value. However, in real life this not the case, because people adapt their behaviour once the virus is spreading (stay at home, avoiding contact with different people). Moreover, the government also takes measures in order to control the virus (lockdown, quarantine, no gatherings etc). Therefore, it is more reasonable to assume that $\beta(t)$ is a function of time.
2. An SEIR model implicitly states that the incubation time period and the infectious time period are both exponentially distributed with mean $1/\alpha$ and $1/\gamma$ respectively (how long until the person is infectious and how long until a person recovers). In case for the infectious time period, this means that regardless of the time since infection, the probability of recovery within a given time interval is constant, which is not realistic (because of the memory-loss property of an exponential distribution). To state it in another way, the exponential distribution overestimates the number of individuals whose duration of infection is much shorter or much longer than the mean [8]. This will be explained more in Section 6.
3. The basic SEIR model is spatially homogeneous. This means that the model assumes that the spread of the virus through the country is more or less homogeneous. However, that is usually not the case. For example in the Netherlands, the first⁴ corona case was in Noord-Brabant, where they also celebrated carnaval. It followed that Noord-Brabant was the first province that had taken their first measures against the virus. After a while, we saw that the number of infected people rose in De Randstad, which is one of the most densely populated areas in the Netherlands. However, the number of infected people in other regions such as Groningen or Limburg were relatively low. We will show in Section 7 a model which includes spatial heterogeneity.
4. An SEIR model does not take into account the age of individuals. However, we saw that COVID-19 was more dangerous for older people (65+) than for younger people. Also, the role of children in the spread of COVID-19 was considered negligible compared to the role of (young) adults.

As we can see from these simplifications, it should not come as a surprise if the results of modelling are not in line with the actual data when we use the basic SEIR model. Therefore, we should try to adapt the basic SEIR model such that we obtain a more realistic model.

⁴<https://nos.nl/artikel/2324870-eerste-nederlander-met-coronavirus-opgenomen-in-tilburg-man-vierde-carnaval>

2.3 Possible solutions to the simplifications

We will briefly present four solutions to the simplifications and explain how to implement these solutions into the SEIR model.

2.3.1 $\beta(t)$ as a function of time

As it is mentioned earlier, it feels more reasonable to assume that $\beta(t)$ is a function of time. There are studies who have used a time-dependent $\beta(t)$ for their models, such as [5] and [6]. However, they implemented a stepwise function or a Heaviside function that models the governmental actions such as lockdowns. These functions are not smooth, which can cause problems when combining these $\beta(t)$ functions with the basic SEIR model. We have considered two ideas to present $\beta(t)$ as a function of time.

(a) ODE for $\beta(t)$

Intuitively, when the number of infected people rises, individuals start to be more careful and avoiding other people. Moreover, the government have to take measures as well to contain the virus spreading. Therefore, the number of contacts per person decreases, thus β decreases. On the other hand, when the number of infected people decreases, people start to meet up again, and the measures are gradually removed. This leads to an increase of β . In mathematical terms, we have the following linear ODE:

$$\frac{d\beta}{dt} = r_+ - r_-, \quad (10)$$

where the growth rates r_+ and r_- are given by

$$r_+ = k_1(\beta_{high} - \beta), \quad r_- = k_2\beta\frac{I}{P_{tot}}. \quad (11)$$

k_1, k_2 and β_{high} are unknown parameters that have to be fit according to the data, and N is the total population. β_{high} is the term for $\beta(t)$ when the situation is normal, which means just before the spreading of the virus starts. In that situation, the number of contacts per person per day is higher (because no risk of infection yet).

(b) Lotka-Volterra equations

The Lotka-Volterra equations are also known as the predator-prey equations. These equations are given by

$$\begin{aligned} \frac{dx}{dt} &= c_1x - c_2xy, \\ \frac{dy}{dt} &= c_4xy - c_3y. \end{aligned} \quad (12)$$

The idea by using the Lotka-Volterra equations to model $\beta(t)$ is still the same as mentioned in the first point. When the number of infected persons $I(t)$ rises, $\beta(t)$ decreases, and when $I(t)$ decreases, then $\beta(t)$ increases again. Therefore, we can model this interaction between $\beta(t)$ and $I(t)$ as a predator-prey model. Now, if we set $x = \beta - \beta_{low}$, $y = I$ and we ensure that $\beta(t)$ has a positive minimum, then we have

$$\begin{aligned} \frac{d\beta}{dt} &= (\beta - \beta_{low}) \cdot (-k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I), \\ \frac{dI}{dt} &= c_4(\beta - \beta_{high})I - c_3I, \end{aligned} \quad (13)$$

where $k_1 = -c_1 > 0$, $k_2 = -c_2 > 0$, β_{high} is the maximum value and β_{low} is the minimum value for $\beta(t)$.

2.3.2 SEIR model with gamma distributed stages

The use of a non-exponential distribution instead of an exponential distribution means that the probability of recovery depends on the time since infection, hence the model needs to keep track of this information [7]. One of the distributions we can choose is a gamma distribution. This corresponds to the subdivision of both the exposed compartment E into m stages and the infected compartment I into n stages. Then, we assume that the time spent in each substage of E is exponentially distributed, all

with identical mean. Hence the total time spent in the m substages of E is given by the sum of m independent exponential distributions, which is a gamma distribution. Similarly for each substage of I , we assume that the time spent in each substage is exponentially distributed with identical means, which implies that the total time spent in the n substages of I is gamma distributed.

The model equations are then given by:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{SI}{P_{tot}}, \\
\frac{dE_1}{dt} &= \beta \frac{SI}{P_{tot}} - m\alpha E_1, \\
\frac{dE_2}{dt} &= m\alpha E_1 - m\alpha E_2, \\
&\vdots \\
\frac{dE_m}{dt} &= m\alpha E_{m-1} - m\alpha E_m, \\
\frac{dI_1}{dt} &= m\alpha E_m - n\gamma I_1, \\
\frac{dI_2}{dt} &= n\gamma I_1 - n\gamma I_2, \\
&\vdots \\
\frac{dI_n}{dt} &= n\gamma I_{n-1} - n\gamma I_n, \\
\frac{dR}{dt} &= n\gamma I_n.
\end{aligned} \tag{14}$$

2.3.3 Spatial heterogeneity

When modelling spatial heterogeneity, we will divide the Netherlands in different regions. The idea is to apply the SEIR model to each region, and then connect all the SEIR models by modelling the spread of the virus between the regions.

Now for each region, the total population per region is divided into the four stages of the SEIR model:

- Susceptibles S_m ,
- Exposed E_m ,
- Infected I_m ,
- Recovered /deceased R_m ,

where $m = 1, 2, \dots, N_r$, and N_r the number of subregions. Then, the SEIR model for region m can be generalised as

$$\begin{aligned}
\frac{dS_m}{dt} &= -\sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}}, \\
\frac{dE_m}{dt} &= \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}} - \alpha E_m, \\
\frac{dI_m}{dt} &= \alpha E_m - \gamma I_m, \\
\frac{dR_m}{dt} &= \gamma I_m,
\end{aligned} \tag{15}$$

where $\beta_{m,n}(t)$ is the transmission coefficient between region m and region n . If $m = n$, we have the transmission coefficient within region m itself (intra-region spreading). In this case, the population density ρ_m of region m is used:

$$\frac{dS_m}{dt} \text{ intra} = -\beta_m(t) \frac{S_m I_m}{P_m}, \quad \beta_m = -k_m \rho_m, \tag{16}$$

and k_m is an unknown fitting parameter. If $m \neq n$, we model the transmission of the disease between different regions (inter-region spreading) as usual:

$$\frac{dS_m}{dt} \text{ inter} = - \sum_{n=1, m \neq n}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}}. \quad (17)$$

2.3.4 Age groups

We can model the spread of the virus between different age groups in a similar way when we model spatial heterogeneity. For each age group, the total population is divided into four different stages of the SEIR model:

- Susceptibles S^k ,
- Exposed E^k ,
- Infected I^k ,
- Recovered /deceased R^k ,

where $k = 1, 2, \dots, N_a$ and N_a the number of age groups. Then it follows that

$$\frac{dS^k}{dt} = - \sum_{l=1}^{N_a} \beta^{k,l} S^k I^l, \quad (18)$$

where $\beta^{k,l}$ is the transmission coefficient between age group k and l . It is possible to combine the age groups with the spatial heterogeneity. We will then have the equation

$$\frac{dS_m^k}{dt} = - \sum_{n=1}^{N_r} \sum_{l=1}^{N_a} \beta_{m,n}^{k,l} S_m^k I_n^l, \quad (19)$$

where the product $S_m^k I_n^l$ represents an encounter of a susceptible living in region m in age group k with an infected living in region n in age group l .

In this thesis however, we did not included the age groups in our modelling due to time limitations.

2.4 Conclusion & Discussion

In this section, we described the definitions, concepts and the underlying assumptions of the basic SEIR model. From this description, we have seen that the basic SEIR model has at least four different (over)simplifications. These simplifications are: a constant parameter β , the assumption of an exponential distribution, spatial homogeneity and not taking into account the age of individuals. For each of these simplifications, we described their possible solutions. The results of these solutions will be shown in later sections.

3 Estimating the number of infected persons in the Netherlands

In this section, we will estimate the number of infected people per day in the Netherlands. It is important that we know how many people are infectious at any moment for at least 2 reasons. First, implementing effective public and health policies can only be done if the number of infected people is known. And secondly, later on we want to analyse the output of the SEIR models by comparing it with the estimations, because the output of the SEIR models is usually the number of infected people at a certain moment of time.

In order to estimate the number of infected persons, we will use a data set from the RIVM that is publicly available⁵. This data set consists of the following information:

- The number of positive tests (reported cases) per day,
- The number of hospital admissions per day,
- The number of deceased per day.

The data set is shown in Figure 2. For modelling purposes, we will use the data set that starts from February 27th 2020 to February 3rd 2021. The data set ends at February 3rd 2021 because at that time the Dutch vaccination programme started. Also, we saw the rise of the British variant of the virus in the Netherlands. So, we wanted to exclude these 2 events from the modelling.

In Figure 3 the same hospital admissions per day and deceased per day are shown for clarity.

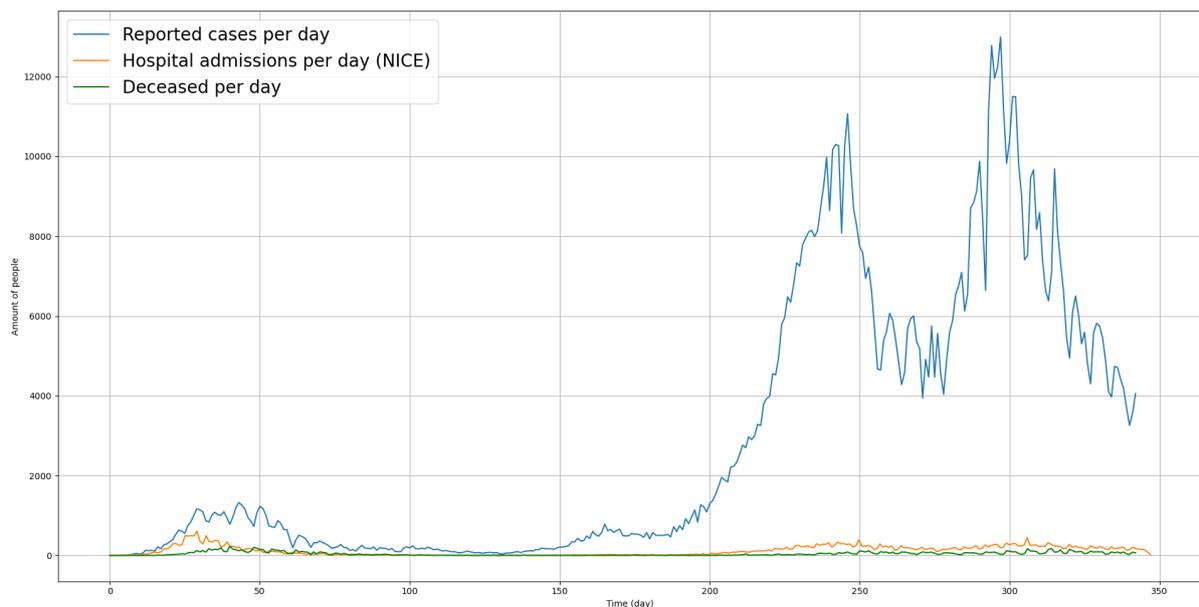


Figure 2: Data set from the RIVM with the number of reported cases, hospital admissions and deceased per day of COVID-19 in the Netherlands. The data set starts from February 27th 2020 to February 3rd 2021.

⁵<https://data.rivm.nl/covid-19/>

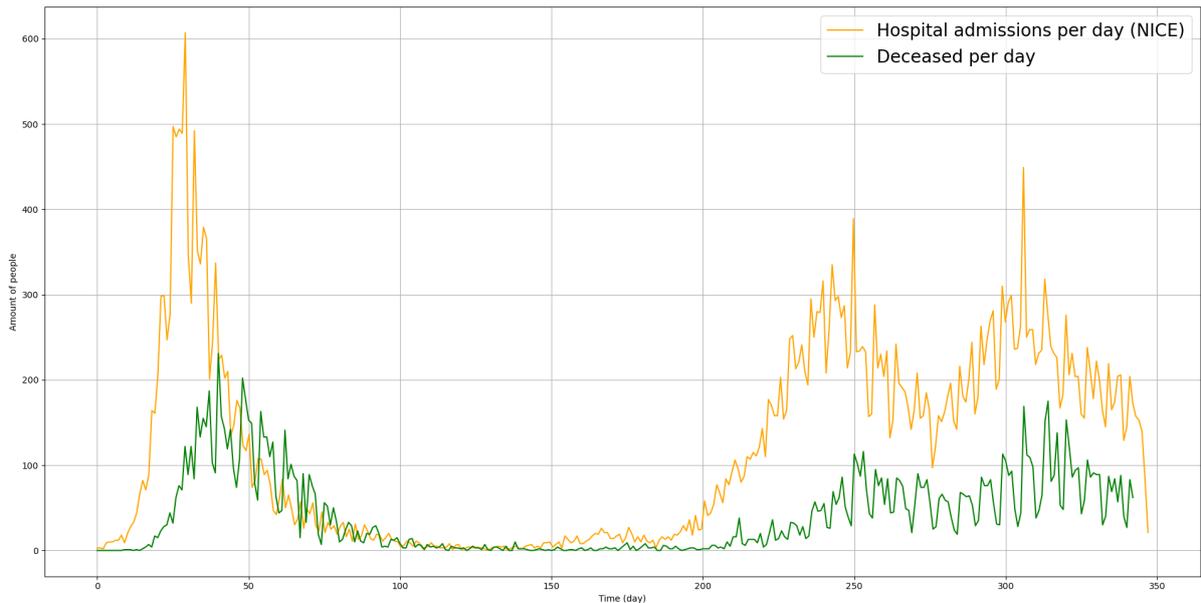


Figure 3: Hospital admissions per day and deceased per day of Figure 2 shown again for clarity.

The number of positive tests per day is not a good measure in order to estimate the number of infected people for at least 3 reasons:

1. At the start of the pandemic, there was not enough test capacity. As a consequence, not all individuals could get tested.
2. Even with sufficient test capacity, there are also people who do not have any symptoms at all (asymptomatic). Thus it is likely that a majority of the asymptomatic cases will not test themselves.
3. Some people are simply not willing to test themselves. They may be scared (children), or they do not trust the result after testing.

However, the number of hospital admissions suffer less from these shortcomings. When someone is hospitalized, we can know for sure that they are infected with visible symptoms (therefore a positive 'test'). Moreover, it is usually not up to the individuals whether or not they want to be hospitalized when their health is at risk. Thus, we will use the number of hospital admissions as a measure to estimate the number of infected people. This will be explained later in this section.

Furthermore, we assume that between the number of infected people and the number of hospital admissions, there exists a certain factor κ . We assume that this factor κ is a constant, independent of time which needs to be estimated from the given data.

Lastly, we will compare our estimation results with the results from the RIVM and Sanquin⁶, an organisation specialised in blood research. The RIVM had based their estimations on serological data from the Pienter Corona onderzoek⁷. Which means that they measured how many donated blood contained antibodies against the virus. Also, they combined the serological data with the amount of hospital admissions per age-group to estimate the number of infected people per day. Sanquin measured the amount of antibodies against COVID-19 as well.

3.1 Difference in data

Before we estimate the total number of infected people, we want to distinguish the data about the hospital admissions given by the RIVM (via the GGD) and another source called NICE (Nationale Intensive Care Evaluatie). The differences are shown in Figure 4.

⁶<https://www.sanquin.nl/>

⁷<https://www.rivm.nl/pienter-corona-studie>

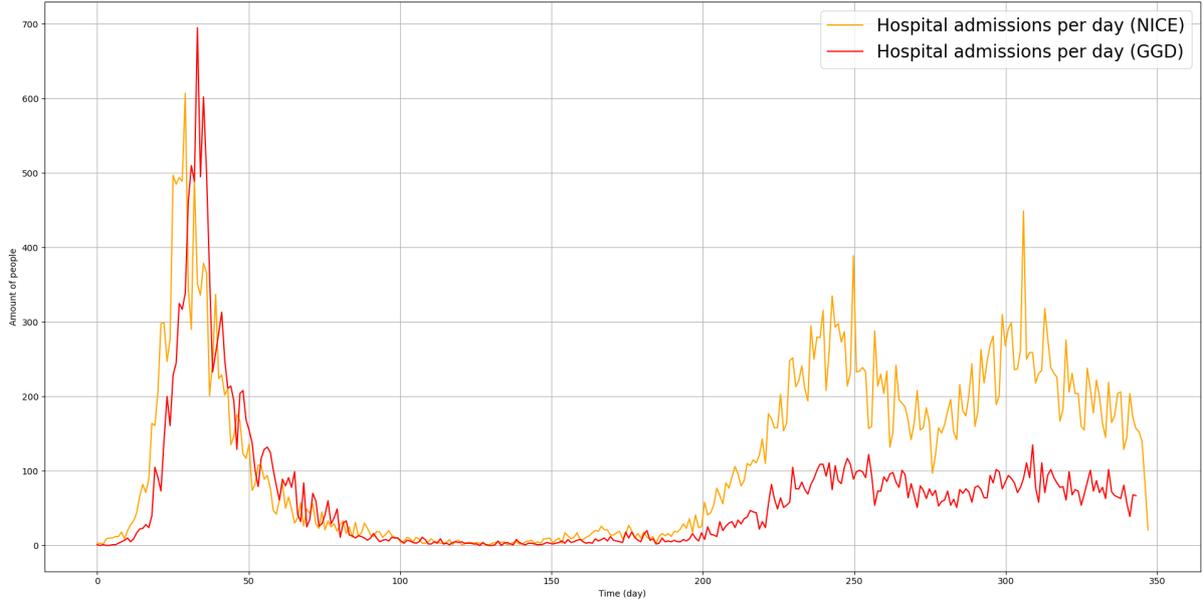


Figure 4: Hospital admissions per day from the GGD and NICE.

The difference between the data is most noticeable during the second wave, where NICE reported more hospital admissions than the GGD. Apparently, the RIVM uses the data of NICE on their website to show the hospital admissions per day. So we will use the data of NICE as well.

3.2 Estimation methods

We will describe 2 methods to estimate the number of infected people with use of the hospital admissions data and the reported positive cases data.

3.2.1 Hospital admissions

For the first method, we define the following notations

$$\begin{aligned} I_{new}(t) &= \text{total number of new infections at the time interval } [t, t + \Delta t], \\ R_{new}(t) &= \text{total number of new recoveries at the time interval } [t, t + \Delta t], \\ H_{new}(t) &= \text{total number of new hospital admissions at the time interval } [t, t + \Delta t]. \end{aligned}$$

With these definitions, it follows that

$$I_{tot}(t) = \int_0^t I_{new}(\tau) d\tau, \quad R_{tot}(t) = \int_0^t R_{new}(\tau) d\tau \quad (20)$$

are the total number of people that have ever been infected and the total number people that have been recovered until time t respectively.

We can relate $I_{new}(t)$ and $H_{new}(t)$ by setting

$$H_{new}(t) = \kappa \left(\int_0^\infty I_{new}(t-a) \rho_H(a) da \right), \quad (21)$$

where $\rho_H(a)$ is a probability mass/density function describing the probability that an individual is hospitalized a days after being infected, and κ is a constant, independent of time that contains the percentage of infections that leads to hospital admission (which we mentioned before).

Similarly, the relation between $I_{new}(t)$ and $R_{new}(t)$ is as follows:

$$R_{new}(t) = \int_0^\infty I_{new}(t-a) \rho_R(a) da, \quad (22)$$

where $\rho_R(a)$ is a probability mass/density function describing the probability that an individual is recovered from the infection a days after being infected.

Now, we make a few assumptions that will simplify the estimation using the model:

1. The RIVM updated their data once per day. This means we set t in days and $\Delta t = 1$.
2. From the serological data of Sanquin, we see that from the first half year of 2020, about 10^6 people had been infected up to that moment (see Table 2). At the same time, the RIVM reported that there were in total 10^4 hospital admissions up to that moment. This means that about 1% of the infections leads to hospital admission. Though, we started with a slightly higher percentage, about 1.25%. For now, this implies that $\kappa = \frac{1}{80}$.
3. The probability density function $\rho_H(a)$ is in reality unknown. Therefore we have to choose a distribution in order to do the calculations. Hospital admission takes place (on average) 14 days after start of the first symptoms. This number is based on the presentation given by Jaap van Dissel on 20th May 2020⁸. So for now, we set $a_H = 14$ and

$$\rho_H(a) = \begin{cases} 1 & \text{if } a = a_H, \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

4. Similarly to assumptions 3, the probability density function $\rho_R(a)$ is unknown. We assume that the infectious time (the time in which a person can transmit the disease) is on average 8 days per person. After the infectious time, we say then that the person is recovered. The infectious time varies among different studies [20]. But for now, we will fix the infectious time to 8 days. So similarly, we set $a_R = 8$ and

$$\rho_R(a) = \begin{cases} 1 & \text{if } a = a_R, \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

From assumptions 2 and 3, equation (21) reduces to

$$\begin{aligned} H_{new}(t) &= \kappa \left(\int_0^\infty I_{new}(t-a) \rho_H(a) da \right), \\ &= \frac{1}{80} I_{new}(t - a_H). \end{aligned} \quad (25)$$

We can rewrite the above equation to obtain the number of new infections per day as

$$I_{new}(t) = 80 \cdot H_{new}(t + a_H). \quad (26)$$

Similarly, by assumption 4, equation (22) reduces to

$$\begin{aligned} R_{new}(t) &= \int_0^\infty I_{new}(t-a) \rho_R(a) da, \\ &= I_{new}(t - a_R). \end{aligned} \quad (27)$$

Consequently, we can now estimate the current number of infected persons at time t . This will be done by subtracting the total number of people that have been recovered until time t from the total number of people that have ever been infected until time t .

$$I(t) = I_{tot}(t) - R_{tot}(t). \quad (28)$$

We can rewrite this in a recursive formula:

$$\begin{aligned} I(t) &= I_{tot}(t) - R_{tot}(t), \\ &= I_{tot}(t-1) + I_{new}(t) - (R_{tot}(t-1) + R_{new}(t)), \\ &= I(t-1) + I_{new}(t) - R_{new}(t), \\ &= I(t-1) + I_{new}(t) - I_{new}(t - a_R). \end{aligned} \quad (29)$$

So using equation (29), we can estimate the total number of infected people at any time t , with use of the data from hospital admissions.

⁸<https://www.tweedekamer.nl/kamerstukken/detail?id=2020D19084&did=2020D19084>

3.2.2 Reported positive cases

In a similar way as in the first method, we can estimate the total number of infected people using the data from the reported cases per day. We define

$$\begin{aligned} I_{new,rep}(t) &= \text{total number of new reported cases at the time interval } [t, t + \Delta t], \\ R_{new,rep}(t) &= \text{total number of new recoveries from the reported cases at the time interval } [t, t + \Delta t]. \end{aligned}$$

Then it follows that

$$I_{tot,rep}(t) = \int_0^t I_{new,rep}(\tau) d\tau, \quad R_{tot,rep}(t) = \int_0^t R_{new,rep}(\tau) d\tau \quad (30)$$

are the total number of reported cases and the total number of recovered people from the reported cases until time t respectively. We will use assumption 4 again that people are recovered 8 days after infection. This means that expression for $R_{new,rep}(t)$ will be similar to equation (27):

$$R_{new,rep}(t) = I_{new,rep}(t - a_R). \quad (31)$$

Consequently, we can estimate the current number of infected people from the reported cases at time t as

$$\begin{aligned} I_{rep}(t) &= I_{tot,rep}(t) - R_{tot,rep}(t), \\ &= \int_0^t I_{new,rep}(\tau) d\tau - \int_0^t R_{new,rep}(\tau) d\tau, \\ &= \int_0^t I_{new,rep}(\tau) d\tau - \int_0^t I_{new,rep}(\tau - a_R) d\tau, \\ &= \int_0^t I_{new,rep}(\tau) - I_{new,rep}(\tau - a_R) d\tau, \\ &= \int_{t-a_R}^t I_{new,rep}(\tau) d\tau, \\ &= \sum_{k=0}^{a_R} I_{rep}(t - k), \quad t \geq a_R. \end{aligned} \quad (32)$$

Where in the last line, the integral changes to a sum because we work with discrete times (per day). Later on, we will compare both estimations from equations (29) and (32) with the results from the RIVM and Sanquin.

3.3 Sanquin data

As we have mentioned earlier, Sanquin measured the amount of antibodies against COVID-19 in the blood of people in the Netherlands. In this way, they can also estimate how many people have been infected with COVID-19 in total. We will call this $I_{tot,sq}$.

The results of their research for different time periods are shown in Figure 5⁹, Table 2 and Figure 6. From their research, we see for example that at May 18th, 5.4% of the people that donated their blood have antibodies, which means that they have been infected to COVID-19. This comes down to around $9.2 \cdot 10^5$ total infected persons in the Netherlands. Remarkably, we see a decline in the percentages (and the corresponding number of people) around the summer months July, August and September. We can think of at least 2 reasons for this:

- The antibodies in individuals disappear gradually over a period of a few months. This is supported by articles such as [21] and [22], where they found that antibodies disappeared within 3 or 4-5 months respectively. This does not mean that these people lost all of their immunity, because the human body has other components as well that protects us against viruses, such as B and T cells.
- Some factors such as (low) sample sizes or inaccurate measurement instruments may have impacted the results negatively. In the table we see that most measurements had around $8 \cdot 10^3$ blood donors. We do not know if such sample size is large enough to estimate the total infections accurately.

⁹<https://www.sanquin.nl/over-sanquin/nieuws/2021/01/antistoffen-tegen-coronavirus-bij-13-procent-van-de-donors-januari->

As a consequence, we note that the numbers shown in the column of $I_{tot,sq}$ are lower estimates of the number of infected people.

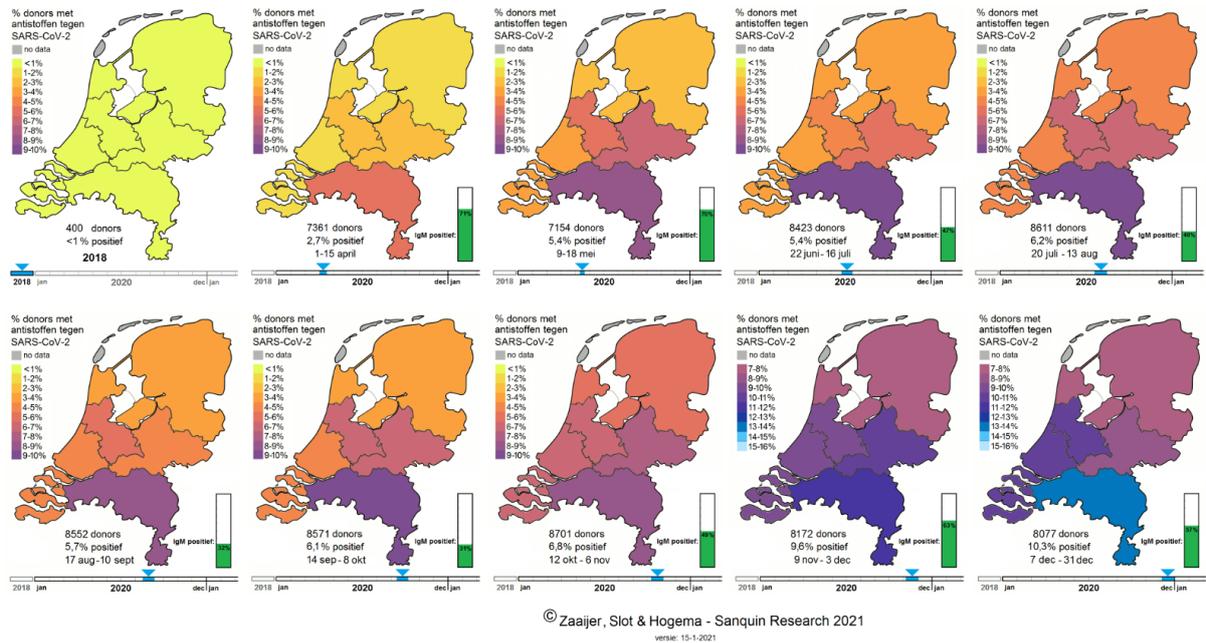


Figure 5: Monthly distribution of the antibodies found in donated blood in the Netherlands. The data start from April 2020 to January 2021.

| Date | Donors | Percentage donors with antibodies | $I_{tot,sq}$ in millions |
|---------------------|--------|-----------------------------------|--------------------------|
| 2018 | 400 | < 1% | < 0.17 |
| 1-15 April 2020 | 7361 | 2.7% | 0.459 |
| 9-18 May | 7154 | 5.4% | 0.918 |
| 22 June - 16 July | 8423 | 5.4% | 0.918 |
| 20 July - 13 August | 8611 | 6.2% | 1.054 |
| 17 August - 10 Sept | 8552 | 5.7% | 0.969 |
| 14 Sep - 8 Oct | 8571 | 6.1% | 1.037 |
| 12 Oct - 6 Nov | 8701 | 6.8% | 1.156 |
| 9 Nov - 3 Dec | 8172 | 9.6% | 1.632 |
| 7 Dec - 31 Dec | 8077 | 10.3% | 1.751 |
| 4 - 12 Jan 2021 | 3057 | 13.3% | 2.261 |

Table 2: Results of Sanquin from Figure 5 in a table.

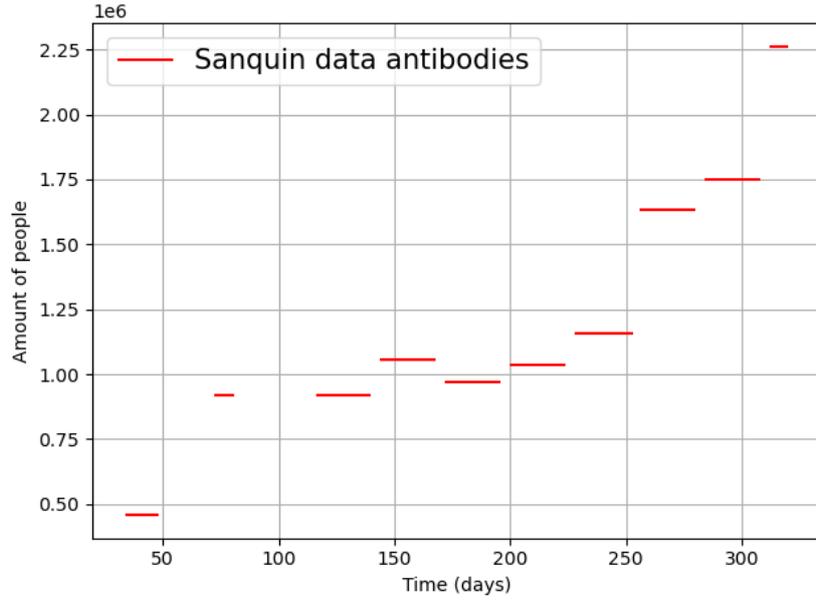


Figure 6: Graphical representation of the Sanquin data from Figure 5.

3.4 Results

We apply the methods described in the earlier section to estimate the number of infected people. Also, we show the number of people that are recovered using the same methods. Furthermore, we compare our estimation results with the data from the RIVM and Sanquin to see how accurate the results are. Lastly, we will also have a look at the intensive care admissions compared to the hospital admissions.

3.4.1 Estimation of the number of infected people

We use equations (29) and (32) to with the parameters

$$a_H = 14, \quad a_R = 8, \quad \kappa = \frac{1}{80}.$$

Combining these estimations with the estimation published by the RIVM, we get the following plots in Figure 7:

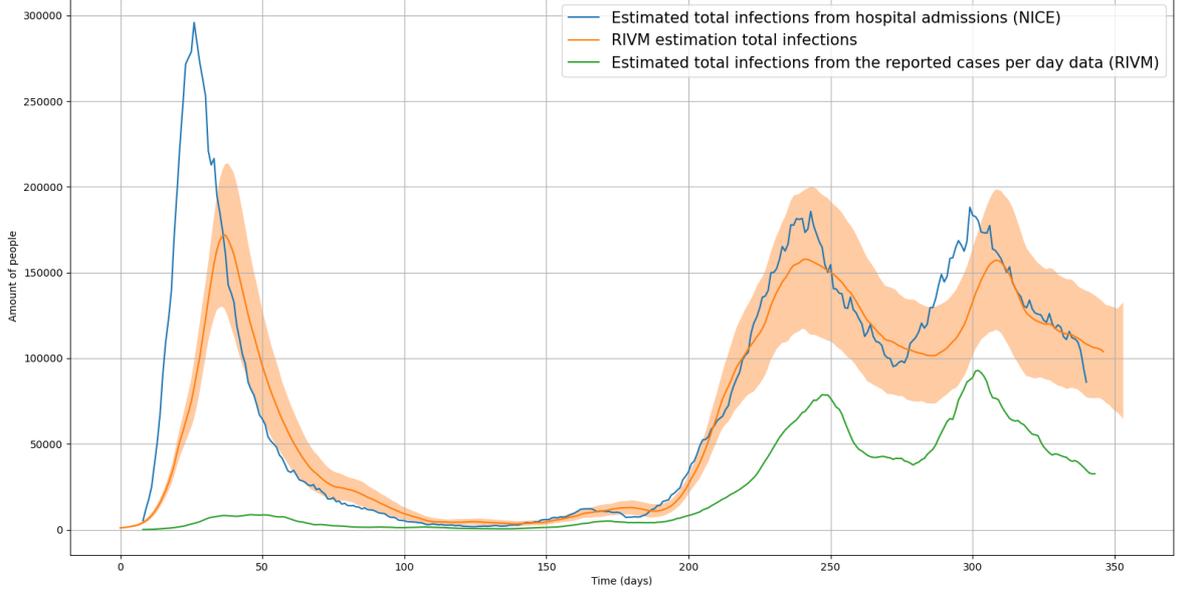


Figure 7: Estimated amount of infected people per day using equation (29) (blue), the estimation from the RIVM with its uncertainty bandwidths (orange), and the estimation using equation (32) (green). The timeline starts from February 17th 2020 to February 3rd 2021.

We see that the order of magnitude of the estimation from the hospital admissions (blue) is similar to the RIVM estimation (orange). Whereas the estimation from reported cases (green) is much lower. However, we see that the blue graph is slightly shifted to the left. A reason for this is because at first we assumed that hospital admissions take place 14 days after start of the first symptoms from the presentation of Jaap van Dissel. However in [23], we can find that 75% of the hospitalizations occur within 8 days after symptom onset. This means that we will have to change our parameter a_H . Furthermore, we want our estimation to be matched with the results from Sanquin, so we will change the parameter κ as well. The results with Sanquin can be seen later on.

Now, changing the parameters to

$$a_H = 8, \quad a_R = 8, \quad \kappa = \frac{1}{82},$$

we have that equation (26) changes to:

$$I_{new}(t) = 82 \cdot H_{new}(t + a_H). \quad (33)$$

After that, we see that the blue graph in Figure 8 is shifted to the right.

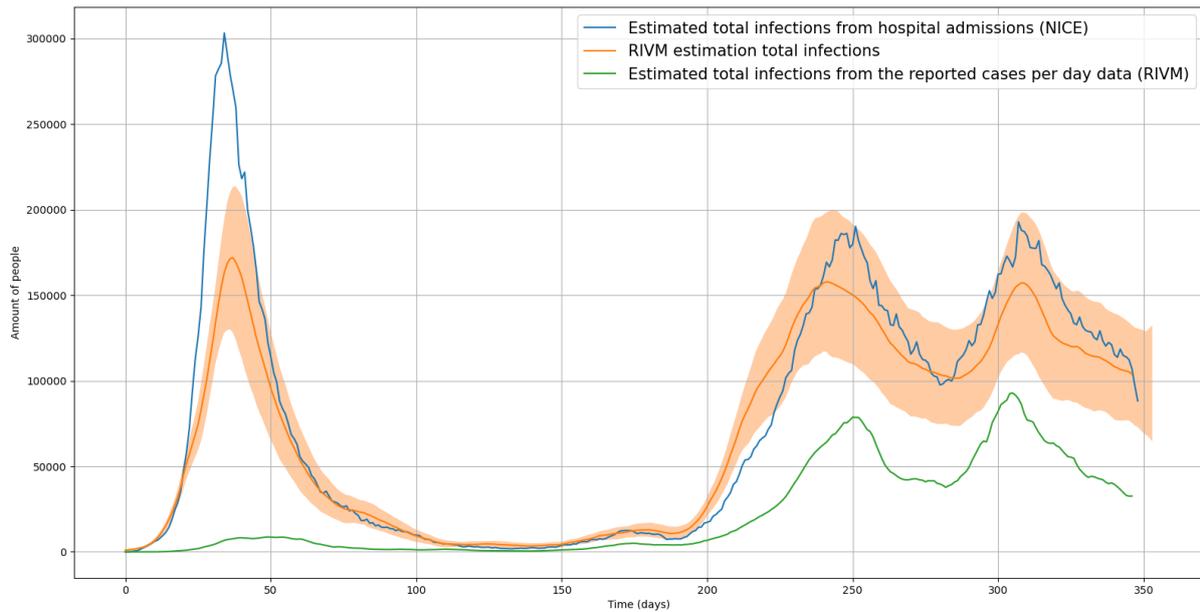


Figure 8: Similar plots as in Figure 7, but now with 8 days between symptoms onset and hospitalization instead of 14 days. Moreover, percentage of infected people that are hospitalized is set to 1.22%.

Now in Figure 8, we can remark several points. At the first wave, we see that the difference between our blue estimation and the RIVM estimation is quite large. We have to keep in mind that when the first wave came, there was not enough capacity to test the infected people, and the whole situation was a bit of a mess. So it is hard to tell if the estimations from ours and the RIVM at the first wave are accurate.

During the second wave, the Netherlands had enough test capacity and there was a better overview of the epidemic. The moment the Netherlands had enough capacity was around the 1st of October¹⁰. We see that after the assumptions changes, the peaks at the second wave are now shifted in the shaded area of the RIVM estimation. Consequently the peaks at the first wave are now aligned as well. So apart from the first wave, we have that the number of infected people is similar to what the RIVM estimated.

We will also look at the ratio between the estimated total infections from the reported cases (green) and the estimated total infections from hospital admissions (blue) during the second wave. By doing so, we can know on average the percentage of the number of total infections per day that have been reported. The ratio plot is in Figure 9.

¹⁰<https://www.volkskrant.nl/nieuws-achtergrond/door-inbreng-buitenlandse-labs-schiet-de-testcapaciteit-omhoog-bac5d620/>

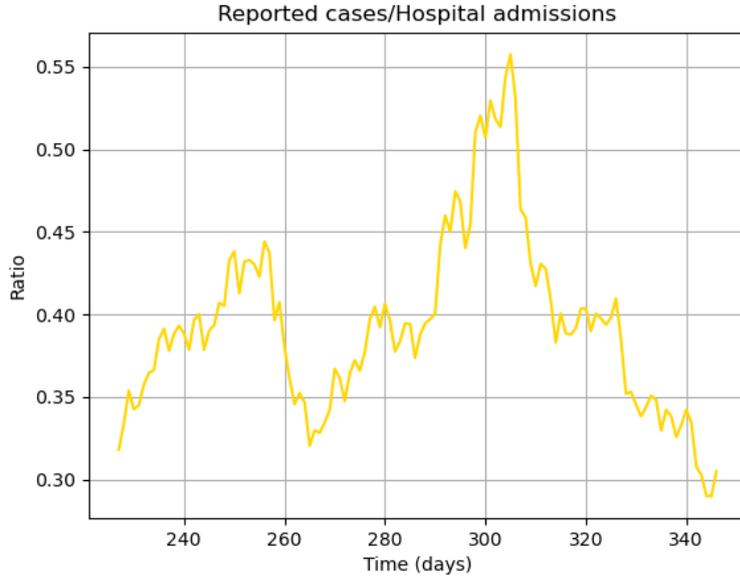


Figure 9: Ratio between the estimated total infections from the reported cases and the estimated total infections from hospital admissions, starting from the 1st of October.

From Figure 9, we can calculate the average ratio over the given time, which is 39.3%. This means that on average, the total infections from the reported cases is merely 39.3% of the true number of total infections. Most of the infections are therefore not reported, for reasons such as not having symptoms (and thus not reported), not wanting to test or other factors.

3.4.2 Number of recovered persons

From equations (27) and (33), we can plot the number of recovered people per day by setting

$$R_{new}(t + a_R) = I_{new}(t) = 82 \cdot H_{new}(t + a_H). \quad (34)$$

With the parameters

$$a_H = 8, \quad a_R = 8, \quad \kappa = \frac{1}{82},$$

we see the number of recovered people per day in Figure 10.

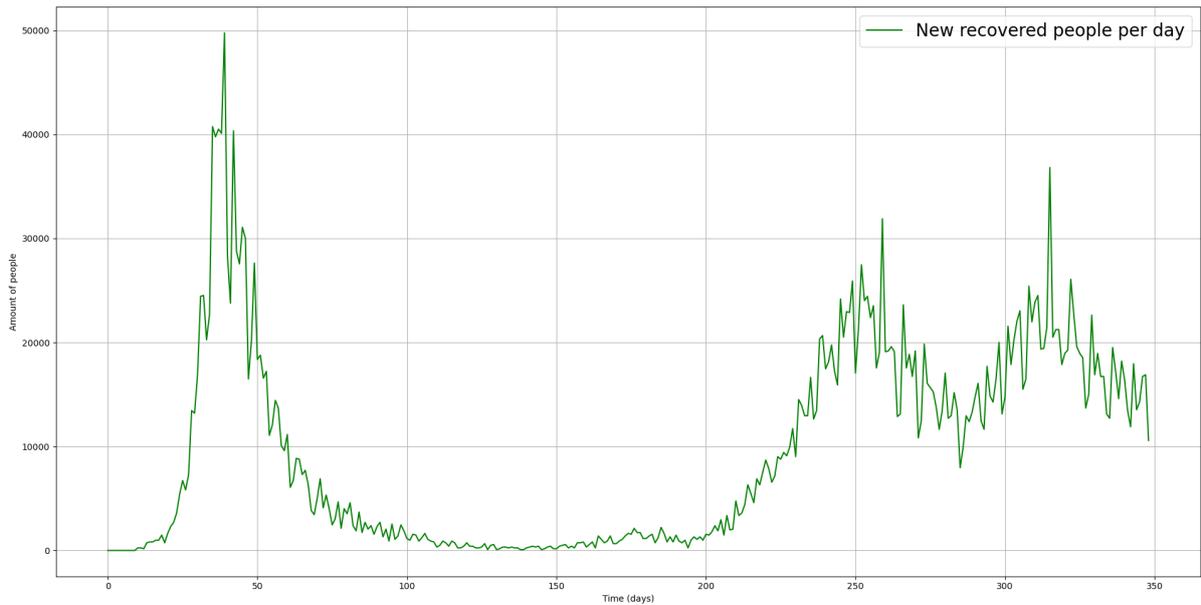


Figure 10: Number of new recoveries per day. The time period starts from February 17th 2020 to February 3rd 2021.

Now, integrating the recoveries per day over the entire time period, we get the total amount of recovered people. This comes down to about 3.3 million people. This is in line with the initial conditions that the RIVM used to run their model¹¹, where they set the total number of recovered persons at 3.13 million, and the model starts at the 1st of February 2021.

3.4.3 Results including Sanquin data

Now, we include the results of Sanquin with the results of the RIVM and our estimation. The results of Sanquin were the total number of infected people up until a certain time, whereas the estimations from the RIVM and ours were the number of infected people per day. So in order to compare all of the results, we will transform the estimations from the RIVM and ours by taking the cumulative sum of the plots in 8, and then we divide by 8 days to take into account the infection duration. Then, the results are shown in Figure 11.

¹¹https://www.rivm.nl/sites/default/files/2021-03/Modellingresults%20COVID19%20vaccination%20version1.0%2020210324_0.pdf

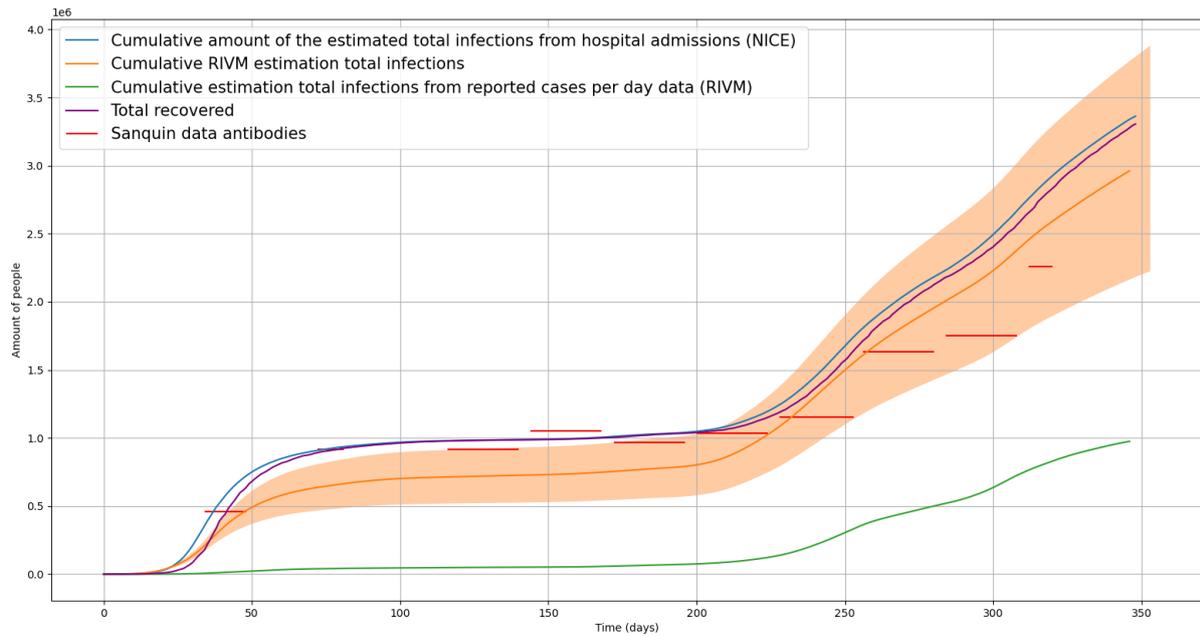


Figure 11: Comparison of the Sanquin data (red) with the estimation using hospital admissions from NICE (blue), the estimation of the RIVM (orange) and the reported cases from the RIVM (green). Additionally, the total recovered persons per day (purple) is shown as well.

In Figure 11, we see that past the 200 days (the second wave), the difference between the Sanquin data and our estimation is fairly large, around a million people. Where our estimation and the RIVM estimation reaches to around 3 million infected persons, the Sanquin data is 'merely' 2.2 million. One possible explanation for this difference is that the antibodies in people are disappearing over time, as we had mentioned that earlier. Before day 200, the amount of people that were infected was around 1 million by our estimation, which matches nicely with the Sanquin data. But at the end of the time period, the difference is around 1 million. So we think that the missing 1 million people from the Sanquin data are those who were infected during the first wave, but their antibodies are not measurable anymore.

3.4.4 Intensive Care admissions

It is also worthwhile to look at the intensive care admissions per day, since the whole policy of the government against COVID-19 is based on alleviating the pressure on hospitals. Especially on the amount of intensive care units (ICU's). In Figure 12 we will plot the both the hospital admissions per day and the intensive care admissions per day. In this way, we can estimate the percentage of hospital admissions that were also admitted to the intensive care.

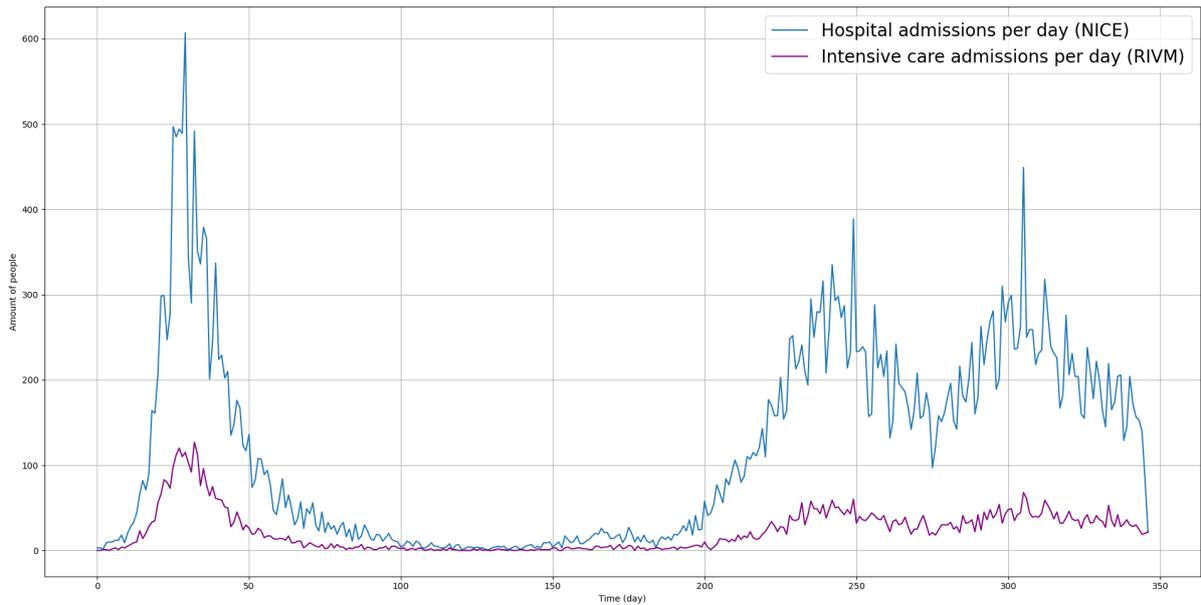


Figure 12: Hospital and intensive care admissions per day in the Netherlands. Data made available by the RIVM, starting from February 27th 2020 to February 3rd 2021.

We are interested in the ratio's between the hospital admissions and the ICU admissions as well, to see the average percentage of people that go to the ICU after hospitalization. Note that we need to take into account the time between hospital admission and ICU admission so that the ratio's are from the same dates. From [24], we see that the time between hospital admission and the ICU is on average 2 days. So we will use that as well. After that, the ratio's are shown in Figure 13:

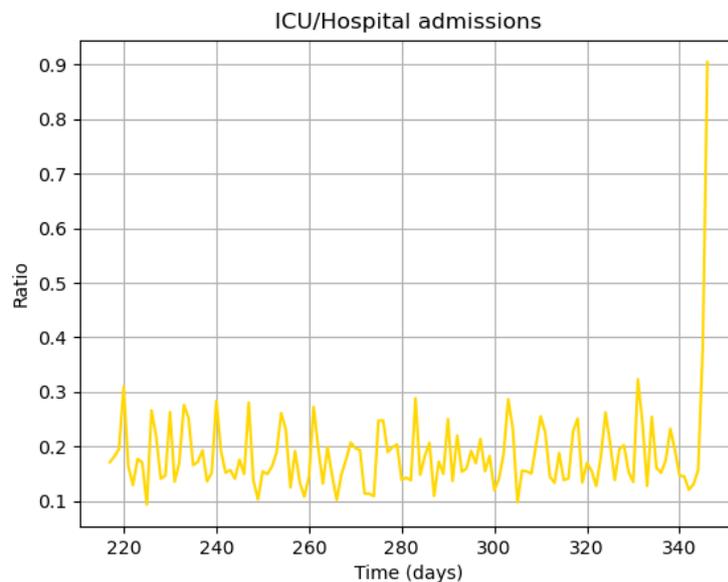


Figure 13: Ratio between the hospital admissions and the ICU admissions, starting from the 1st of October.

Calculating the average ratio over the given time, we have an average ratio of 0.186. This means that 18.6% of the hospitalized patients are admitted to the ICU on average.

3.5 Conclusion & Discussion

From the data acquired from the RIVM, we have seen that by setting the percentage of new infected people that are hospitalized to 1.22%, our estimation for the total infected people is in line with the estimation given by the RIVM and the serological data from Sanquin. Also, we estimated the total infected persons from the reported cases per day. Comparing these 2 estimations, we have seen that only 39.3% of the true number of total infected people are reported. Furthermore, we have seen that 18.6% of the hospital admissions lead to intensive care admissions.

During the estimations and analysis of the data, we set a fixed value of 8 days for the average time between symptom onset and hospitalization, where we retrieved this value from [23]. In equation (33), we see how this value works out in our estimations. During the writing of this thesis, there are still many uncertainties and unknowns about COVID-19, especially about parameter estimations. Over time, values such as the time between symptom onset and hospitalization may change.

This is also the case for the infectious time period (or recovery time), which means that on average, people are recovered 8 days after infection. For example the infectious time period is used in equation (29). However in this study [20], it mentioned several other studies that have found or used different values for the duration of infectiousness. For example in [12] and [18], 5 days was mentioned, whereas 12.5 days was mentioned in [16]. With such variety in time for both the infectious time period and the time between symptom onset and hospitalization, it suggests that a different (continuous) probability distribution for $\rho_H(a)$ and $\rho_R(a)$ might be a better choice. As a consequence, the estimation curves will be smoother when a continuous probability density function is used.

We expected that the ratio plot in Figure 9 would be similar to the plot in Figure 13 qualitatively, because the green and blue graphs in Figure 8 both have similar patterns (2 spikes after day 200) on top of each other. However, the plot in Figure 13 oscillates around a certain value (mean), whereas in Figure 9 there is an outlier with ratio 0.55. Perhaps the alignment of the blue and green graphs is shifted a few days, which is affected by the chosen infectious and recovery time.

Nevertheless, comparing our current estimations with the data from Sanquin and the RIVM (for example in Figure 11), the estimations have come close to that of RIVM and Sanquin. This suggest that our current parameter values are also close to the actual values.

4 Initial values and first results with the SEIR model

In order to solve the system of DE's (35), we need to choose a starting date, initial conditions and values for the parameters: $\alpha, \beta_{high}, \gamma, k_1, k_2$. As starting date we choose July 16th 2020, which is slightly before the second wave started in the Netherlands. As it was mentioned before in Section 3, the Netherlands had enough test capacity and a better overview of the epidemic during the second wave.

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{SI}{P_{tot}}, \\ \frac{dE}{dt} &= \beta \frac{SI}{P_{tot}} - \alpha E, \\ \frac{dI}{dt} &= \alpha E - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}\tag{35}$$

$$\frac{d\beta}{dt} = k_1(\beta_{high} - \beta) - k_2 \frac{\beta I}{P_{tot}}.$$

Next, we will discuss the choice of the parameters α, γ and the initial conditions in Section 4.1. After that, the initial value for $\beta(t_0)$ and the parameter β_{high} will be explained in Section 4.2. And finally, we note that k_1 and k_2 are fitting parameters. We conclude this section with a few numerical solutions in Section 4.3 and we compare those solutions with the estimated data that was shown in Section 3.

4.1 Values for γ and α and initial conditions

We recall from Section 2 that we derived estimations for the parameters γ and α . These were given by

$$\gamma \approx \frac{1}{t_{inf}} = \frac{1}{\text{Infectious time}}$$

and

$$\alpha \approx \frac{1}{t_{inc}} = \frac{1}{\text{Incubation time}}.$$

In Section 3, we assumed that the infectious time $t_{inf} = 8$ days. For consistency, we therefore hold on to this assumption, meaning that we set

$$\gamma = \frac{1}{8}.$$

For the incubation time t_{inc} , we follow the information¹² given by the RIVM, where it is mentioned that after being infected, symptom onset mostly take place after 5-6 days. So we will set $t_{inc} = 5$, and thus we have

$$\alpha = \frac{1}{5}.$$

On July 16th 2020, we have the following numbers (rounded) based on the RIVM data and our own estimation from the previous section:

$$\begin{aligned}P_{tot} &= 17 \cdot 10^6, \\ S(t_0) &= 16 \cdot 10^6, \\ E(t_0) &= 4510, \\ I(t_0) &= 3854, \\ R(t_0) &= 0.988 \cdot 10^6,\end{aligned}$$

where t_0 is the date July 16th 2020. Furthermore, we had set $t_{inc} = 5$. This means that we have that $E(t_0) = I(t_0 + t_{inc}) = I(t_0 + 5)$.

¹²<https://www.rivm.nl/coronavirus-covid-19/ziekte>

4.2 Values for β_{high} and $\beta(t_0)$

We also have to estimate the value β_{high} and the initial value $\beta(t_0)$. Recall that β_{high} is the term for $\beta(t)$ when the situation is normal, which was the case before the pandemic started. This means that no measures were taken yet. At the start of the pandemic, we assume that $S \approx P_{tot}$ and $\beta = \beta_{high}$, which implies that the system of ODEs (35) becomes

$$\begin{aligned}\frac{dS}{dt} &= \beta_{high}I, \\ \frac{dE}{dt} &= \beta_{high}I - \alpha E, \\ \frac{dI}{dt} &= \alpha E - \gamma I, \\ \frac{dR}{dt} &= \gamma I.\end{aligned}\tag{36}$$

From the equations of $\frac{dE}{dt}$ and $\frac{dI}{dt}$, we can write those equations as

$$\frac{d}{dt} \begin{pmatrix} E \\ I \end{pmatrix} = \begin{pmatrix} -\alpha & \beta_{high} \\ \alpha & -\gamma \end{pmatrix} \begin{pmatrix} E \\ I \end{pmatrix}.\tag{37}$$

Now, (37) is a system of the form $\mathbf{x}' = A\mathbf{x}$, where

$$A = \begin{pmatrix} -\alpha & \beta_{high} \\ \alpha & -\gamma \end{pmatrix}.\tag{38}$$

This is a homogeneous system of DEs. Thus, the general solution of this system is given by

$$\mathbf{x}(t) = c_1 \mathbf{v}_1 e^{\lambda_1 t} + c_2 \mathbf{v}_2 e^{\lambda_2 t}.\tag{39}$$

From linear algebra, there exists a theorem stating that the trace of a matrix equals the sum of its eigenvalues. This means that we have

$$-\alpha - \gamma = \text{Tr}(A) = \lambda_1 + \lambda_2.\tag{40}$$

Since $-\alpha - \gamma < 0$, it follows that $\lambda_1 + \lambda_2 < 0$ as well. This means that we either have both $\lambda_1, \lambda_2 < 0$ or one of the eigenvalues is positive. However when both eigenvalues are negative, then the epidemic would not grow, which is not an interesting case. So we assume that one of the eigenvalues λ_+ is positive and the other λ_- is negative. Thus we have that

$$\mathbf{x}(t) = c_1 \mathbf{v}_1 e^{\lambda_+ t} + c_2 \mathbf{v}_2 e^{\lambda_- t}.\tag{41}$$

At the beginning of the epidemic, the term with the negative eigenvalue decays quickly, meaning that we can roughly estimate the solution as

$$\mathbf{x}(t) \approx c_1 \mathbf{v}_1 e^{\lambda_+ t}.\tag{42}$$

Now we define the term t_{reg} as the regeneration time, which means the average time it takes for a person to infect another person (secondary infection). We also mention the basic reproduction number \mathcal{R}_0 , which is defined as the number of people who are infected, on average, by one person (with COVID-19). So for example if $\mathcal{R}_0 = 2$, it means that the number of people being infected doubles with each ‘generation’ of transmission. This is called exponential growth. As long as $\mathcal{R}_0 > 1$, the number of people infected increases faster (exponentially). More details on the reproduction number are written in Appendix A and B.

So at the start of the epidemic, we can estimate the growth as

$$e^{\lambda_+ t_{reg}} \approx \mathcal{R}_0.\tag{43}$$

It follows that

$$\lambda_+ \approx \frac{\ln(\mathcal{R}_0)}{t_{reg}}.\tag{44}$$

The characteristic polynomial of matrix A is given by

$$\begin{aligned}(-\alpha - \lambda)(-\gamma - \lambda) - \alpha\beta_{high} &= 0, \\(\alpha + \lambda)(\gamma + \lambda) - \alpha\beta_{high} &= 0.\end{aligned}\tag{45}$$

Using (44) and rewriting, it follows that

$$\beta_{high} \approx \frac{1}{\alpha} (\alpha + \lambda_+) (\gamma + \lambda_+).\tag{46}$$

Furthermore, we assumed that the incubation time is 5 days, and the infectious period is 8 days. So the time it takes for an individual to infect another person is on average $t_{reg} = 5 + 8/2 = 9$ days. Also, at the start of the epidemic in the Netherlands, the RIVM measured¹³ $\mathcal{R}_0 = 2.2$. Thus, with $\alpha = 1/5$ and $\gamma = 1/8$, we have that

$$\beta_{high} \approx 0.31.\tag{47}$$

To estimate the initial value $\beta(t_0)$, with t_0 starting at July 16th 2020, we can follow the same estimation used for β_{high} . The reproduction number on July 16th 2020 was $\mathcal{R}_0 = 1.21$. So we have that

$$\begin{aligned}\beta(t_0) &\approx \frac{1}{\alpha} \left(\alpha + \frac{\ln(1.21)}{t_{reg}} \right) \left(\gamma + \frac{\ln(1.21)}{t_{reg}} \right), \\ &\approx 0.16.\end{aligned}\tag{48}$$

So we will use these initial values in our preliminary results.

A less rigorous approach to estimate the values for β_{high} and $\beta(t_0)$ is that in our SEIR model, there exists a relation between \mathcal{R}_0, β and γ in the form

$$\mathcal{R}_0 = \frac{\beta}{\gamma}.\tag{49}$$

Then with $\mathcal{R}_0 = 2.2$, we have that

$$\begin{aligned}\beta_{high} &= 2.2 \cdot \gamma, \\ &\approx 0.28.\end{aligned}\tag{50}$$

With $\mathcal{R}_0 = 1.21$, we have that

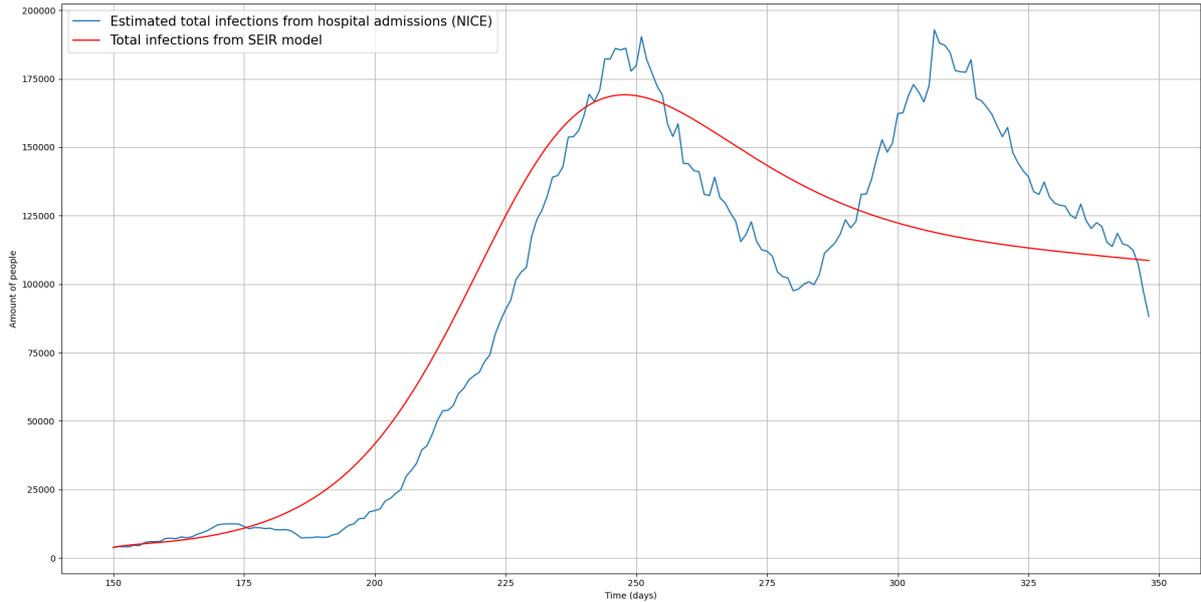
$$\begin{aligned}\beta(t_0) &= 1.21 \cdot \gamma, \\ &\approx 0.15.\end{aligned}\tag{51}$$

4.3 Preliminary results

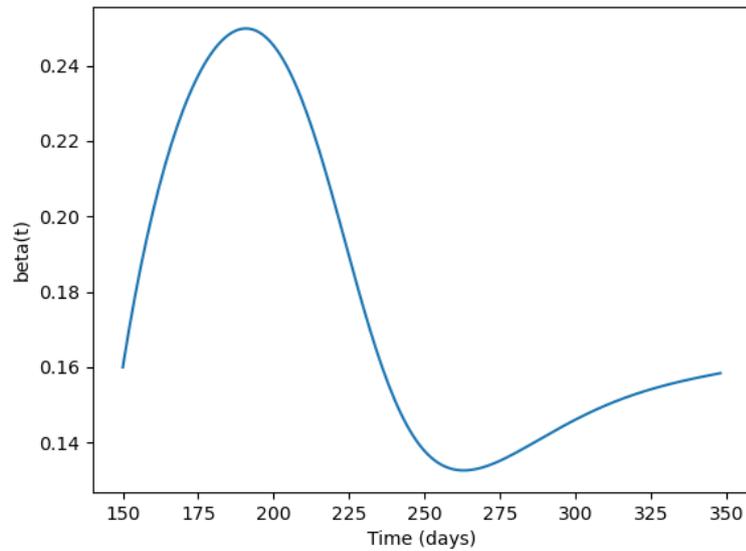
With the initial conditions and values from the previous sections, we can generate a few results with the model given by (35).

In Python, we can use the function `curve_fit` from `scipy.optimize` to fit the output of the SEIR model to the estimated data. This function uses least-squares method to fit the model to the data. The parameters to be fit are k_1 and k_2 from (35), however we also allow the parameters $\beta_{high}, \alpha, \gamma$ and $\beta(t_0)$ to be fit, though within a small range of the original value of order 10^{-2} to 10^{-3} . Starting with $\beta_{high} = 0.31$ and $\beta(t_0) = 0.16$, we have the following result:

¹³<https://coronadashboard.rijksoverheid.nl/landelijk/reproductiegetal>



(a) Estimated data of the second wave of infections from Figure 8 and the result of the SEIR model. The data starts from July 16th 2020 until January 30th 2021.

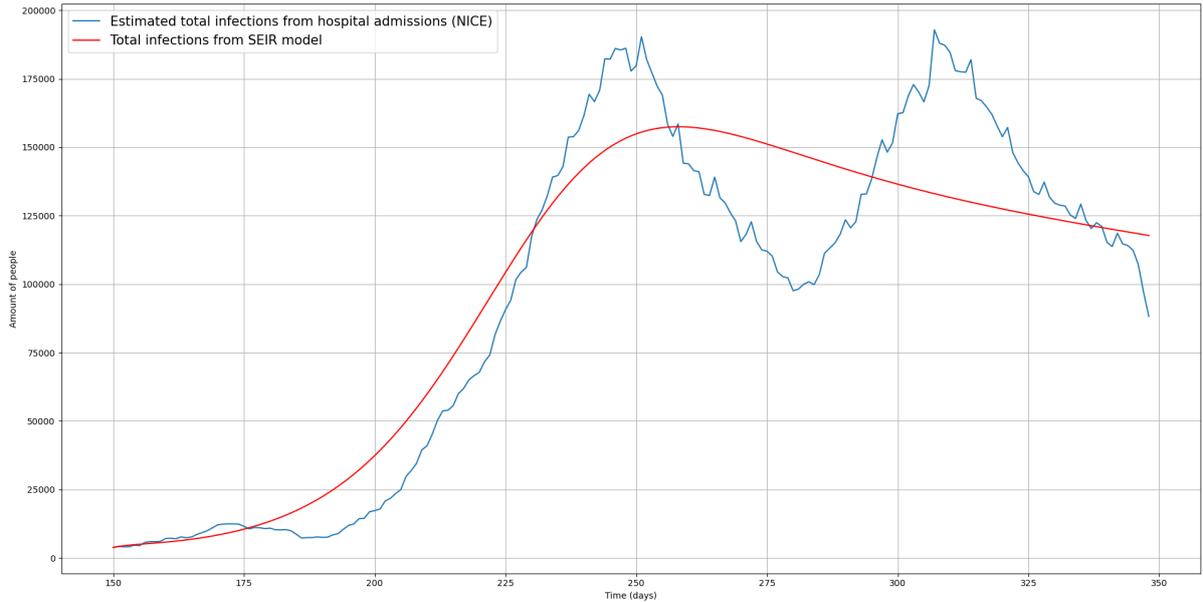


(b) Plot of $\beta(t)$ with parameters $k_1 = 0.037$, $k_2 = 5.016$, $\beta_{high} = 0.30$ and $\beta(t_0) = 0.16$.

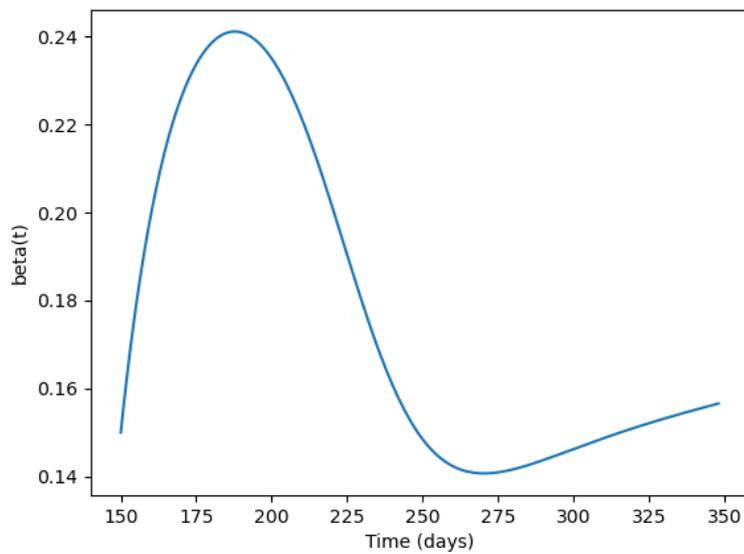
Figure 14: First results of the SEIR model with use of the initial conditions discussed earlier.

As a first fitting, we see in Figure 14a that the model captures the first peak of the estimated data quite well, where the maxima of both the model and the data are around the same days. However the model does not capture the second peak at all. Note also that after the fitting, the value of β_{high} is slightly changed to $\beta_{high} = 0.30$.

Now if we use the β -values obtained from the less rigorous approach, i.e. , $\beta_{high} = 0.28$ and $\beta(t_0) = 0.15$, we have the following plots:



(a) Estimated data of the second wave of infections from Figure 8 and the result of the SEIR model, now with the initial values $\beta_{high} = 0.28$ and $\beta(t_0) = 0.15$.



(b) Plot of $\beta(t)$ with parameters $k_1 = 0.057$, $k_2 = 5.69$, $\beta_{high} = 0.27$ and $\beta(t_0) = 0.15$.

Figure 15: Results of the SEIR model with use of the initial values from the less rigorous approach.

We see that the plots from Figure 15 are similar to the plots from Figure 14. In both figures, the model reaches a maximum at or around the same days, followed by a monotone decrease of the infections. Also, the number of infected people in both the model and the estimated data are of the same order. A small difference is that the curve of the model output in Figure 15a is slightly more flattened than the curve in Figure 14a after the first peak. So we see that the values obtained from the less rigorous approach and the estimation in Section 4 do not affect the numerical solutions that much. Although the model could not capture the second peak, we were able to implement an ODE for $\beta(t)$ into the standard SEIR model, resulting in decent numerical solutions.

5 Lotka-Volterra equations for $\beta(t)$

In Section 4, we showed a few results where the time-dependent function $\beta(t)$ was modelled using a linear ODE (mentioned in Section 2 as well). In this section, we will discuss a second method that models $\beta(t)$ as a function of time. This method is based on the well-known Lotka-Volterra equations, also known as the predator-prey equations. The Lotka-Volterra equations are given by

$$\begin{aligned}\frac{dx}{dt} &= c_1x - c_2xy, \\ \frac{dy}{dt} &= c_4xy - c_3y.\end{aligned}\tag{52}$$

Here x denotes the prey (growth inhibited by y) and y denotes the predator (growth increased by x). Furthermore, c_1, c_2, c_3 and c_4 are positive constants. The Lotka-Volterra model has periodic solutions, which implies that oscillations in the number of preys and predators will occur. Our goal is to use this periodicity to model the waves of infections that have occurred in the Netherlands.

First, we will discuss how to adapt the SIR model inspired by the Lotka-Volterra model in Section 5.1. After that, stability analysis of the adapted SIR model is done. Then, we will show a few numerical solutions. In these solutions however, we observe that $\beta(t)$ can become negative. We adjust the ODE for $\beta(t)$ slightly in order to prevent this in Section 5.2. With this adjustment, stability analysis is done again, and numerical solutions are shown. Finally, we will show results where the adaptations are used in the SEIR model in Section 5.3.

5.1 Lotka-Volterra and the SIR model

Instead of the SEIR model we have used so far, we take a small step back to apply the Lotka-Volterra equations on the basic SIR model. The SIR model is given by

$$\begin{aligned}\frac{dS}{dt} &= -\beta\frac{SI}{P_{tot}}, \\ \frac{dI}{dt} &= \beta\frac{SI}{P_{tot}} - \gamma I, \\ \frac{dR}{dt} &= \gamma I.\end{aligned}\tag{53}$$

In the SIR model, the exposed compartment is not included, meaning that the susceptibles directly move to the infectious compartment after being infected. Although this is not entirely in line with what we have observed from COVID-19, we want to know if stability is achieved with use of Lotka-Volterra in the model. Analysis of the stability in the SIR model simplifies the calculations.

Now, translating the Lotka-Volterra equations (52) to the SIR model (53), we have that β is the prey (because β decreases for larger I) and I is the predator (growth in β means an increase in the amount of infected people). Shifting the value of β upwards by the factor β_{high} (which is the value of β before the epidemic), and by setting $x = \beta - \beta_{high}$, $y = I$, we have

$$\frac{d\beta}{dt} = c_1(\beta - \beta_{high}) - c_2(\beta - \beta_{high})I,\tag{54}$$

$$\frac{dI}{dt} = c_4(\beta - \beta_{high})I - c_3I.\tag{55}$$

If we choose $c_4 = S/P_{tot}$ and $c_3 = \gamma - c_4\beta_{high}$, we have the same ODE as in the SIR model for I again:

$$\frac{dI}{dt} = \beta\frac{SI}{P_{tot}} - \gamma I.$$

Furthermore, we set $k_1 = -c_1 > 0$ and $k_2 = -c_2 > 0$, which yields

$$\frac{d\beta}{dt} = -k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I,\tag{56}$$

where the signs are chosen to obtain the correct phase space behaviour.

5.1.1 Stability analysis of the first adapted SIR model

We will analyse the stability of the system

$$\begin{aligned}\frac{d\beta}{dt} &= -k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I, \\ \frac{dI}{dt} &= c_4(\beta - \beta_{high})I - c_3I.\end{aligned}\tag{57}$$

The analysis is done using standard techniques by finding critical points, calculating eigenvalues and classifying the stability of the critical points.

We find the critical points of the system by setting $\frac{d\beta}{dt} = 0$ and $\frac{dI}{dt} = 0$. For $\frac{d\beta}{dt} = 0$, it follows that

$$\frac{d\beta}{dt} = 0 \Leftrightarrow \beta = \beta_{high} \vee I = \frac{k_1}{k_2}.\tag{58}$$

If $\beta = \beta_{high}$, we have

$$\begin{aligned}\frac{dI}{dt} &= 0 \Leftrightarrow -c_3I = 0, \\ &\Leftrightarrow I = 0.\end{aligned}\tag{59}$$

So the first critical point is

$$(\beta_{high}, 0).$$

For the case $I = \frac{k_1}{k_2}$, we have that

$$\begin{aligned}\frac{dI}{dt} &= c_4(\beta - \beta_{high})\frac{k_1}{k_2} - c_3\frac{k_1}{k_2} = 0, \\ &\Leftrightarrow c_4(\beta - \beta_{high}) = c_3, \\ &\Leftrightarrow \beta = \frac{c_3}{c_4} + \beta_{high}.\end{aligned}\tag{60}$$

So the second critical point is

$$\left(\frac{c_3}{c_4} + \beta_{high}, \frac{k_1}{k_2}\right).$$

Using that $c_4 = S/P_{tot}$ and $c_3 = \gamma - c_4\beta_{high}$, we have

$$\left(\frac{c_3}{c_4} + \beta_{high}, \frac{k_1}{k_2}\right) = \left(\frac{\gamma - \frac{S}{P_{tot}}\beta_{high}}{S/P_{tot}} + \beta_{high}, \frac{k_1}{k_2}\right) = \left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right).$$

So again, the critical points are

$$(\beta_{high}, 0) \text{ and } \left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right).$$

Now, let $f(\beta, I)$ be

$$f(\beta, I) = \begin{bmatrix} -k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I \\ \beta I \frac{S}{P_{tot}} - \gamma I \end{bmatrix}.\tag{61}$$

Then calculating the partial derivatives to β and I for both equations yields the Jacobi matrix

$$\frac{\partial f(\beta, I)}{\partial(\beta, I)} = \begin{bmatrix} -k_1 + k_2I & k_2(\beta - \beta_{high}) \\ I \frac{S}{P_{tot}} & \beta \frac{S}{P_{tot}} - \gamma \end{bmatrix}.\tag{62}$$

Substitution of the equilibrium point $(\beta_{high}, 0)$ yields the matrix

$$\begin{bmatrix} -k_1 & 0 \\ 0 & \beta_{high} \frac{S}{P_{tot}} - \gamma \end{bmatrix}.\tag{63}$$

From this we see that the eigenvalues of the matrix are

$$\lambda_1 = -k_1, \quad \lambda_2 = \beta_{high} \frac{S}{P_{tot}} - \gamma. \quad (64)$$

From these eigenvalues we see that $\lambda_1 < 0$, and it follows that while $\beta_{high} \frac{S}{P_{tot}} < \gamma$, we have that $\lambda_2 < 0$. So that the critical point $(\beta_{high}, 0)$ is an attractor. Else if $\gamma < \beta_{high} \frac{S}{P_{tot}}$ we have a saddle node.

Substitution of the equilibrium point $\left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right)$ yields the matrix

$$\begin{bmatrix} 0 & k_2 \left(\gamma \frac{P_{tot}}{S} - \beta_{high}\right) \\ \frac{k_1}{k_2} \frac{S}{P_{tot}} & 0 \end{bmatrix}. \quad (65)$$

Setting the characteristic polynomial to zero gives us

$$\begin{aligned} \lambda^2 - \frac{k_1}{k_2} \frac{S}{P_{tot}} \cdot k_2 \left(\gamma \frac{P_{tot}}{S} - \beta_{high}\right) &= 0, \\ \Rightarrow \lambda^2 - k_1 \frac{S}{P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{high}\right) &= 0. \end{aligned} \quad (66)$$

Now if $\gamma \frac{P_{tot}}{S} - \beta_{high} > 0$ ($\Leftrightarrow \gamma > \beta_{high} \frac{S}{P_{tot}}$), we have that

$$\lambda_1, \lambda_2 = \pm \sqrt{k_1 \frac{S}{P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{high}\right)}, \quad (67)$$

and we have a saddle node.

Otherwise, if $\gamma \frac{P_{tot}}{S} - \beta_{high} < 0$ ($\Leftrightarrow \gamma < \beta_{high} \frac{S}{P_{tot}}$), we have that

$$\lambda_1, \lambda_2 = \pm i \sqrt{k_1 \frac{S}{P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{high}\right)}, \quad (68)$$

This means that the critical point $\left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right)$ is a center in the linearized system (for the non-linear case, there is actually more analysis needed, but it is known from the Lotka-Volterra equations that a center exists in the non-linear case).

To summarize the stability analysis, we have 2 cases:

1. If $\gamma < \beta_{high} \frac{S}{P_{tot}}$, then $(\beta_{high}, 0)$ is a saddle node, and $\left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right)$ is a center (in the linearized system).
2. If $\gamma > \beta_{high} \frac{S}{P_{tot}}$, then $(\beta_{high}, 0)$ is an attractor and $\left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right)$ is a saddle node.

Note that the condition

$$\frac{S}{P_{tot}} = \frac{\gamma}{\beta_{high}} \quad (69)$$

means that we have reached herd immunity. So initially, we have $S > \gamma \cdot P_{tot} / \beta_{high}$ and we are in case **1**. After reaching herd immunity, the phase space changes. Then $(\beta_{high}, 0)$ becomes stable and the epidemic dies out.

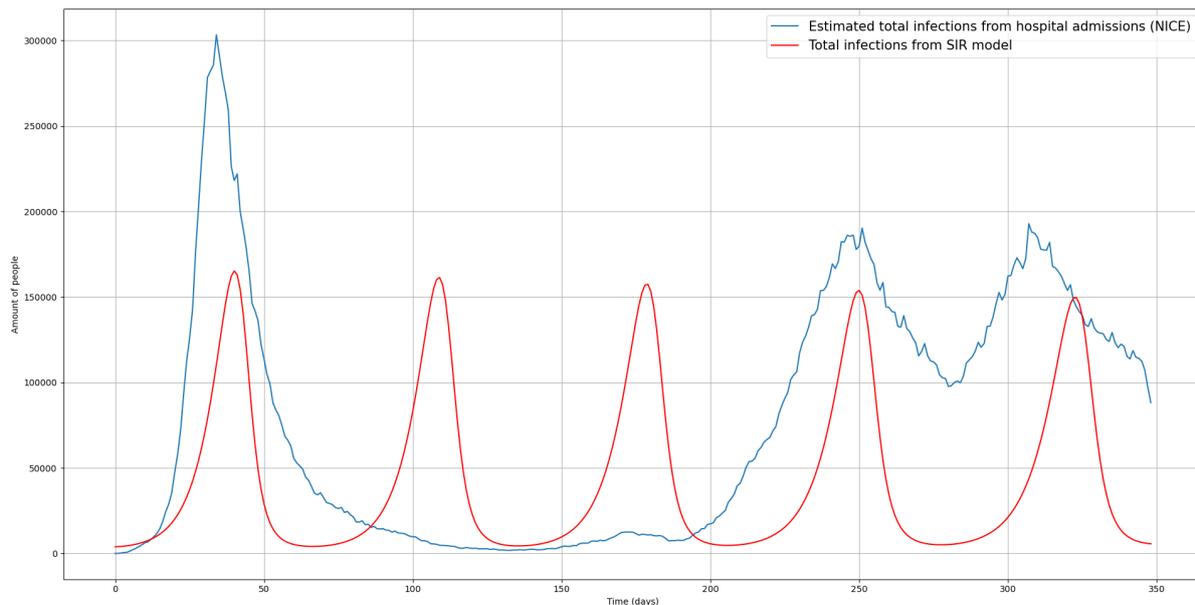
5.1.2 Results of the first adaptation of the SIR model

From the stability analysis, we see that there exists a center in the phase plane, meaning that we can expect periodicity in the numerical solutions. The periodicity of these solutions can be viewed as the COVID-19 'waves' of infected persons. Therefore, we want to try to fit the model to the entire estimated data set, such that the model matches with the waves of infections that have taken place in the Netherlands.

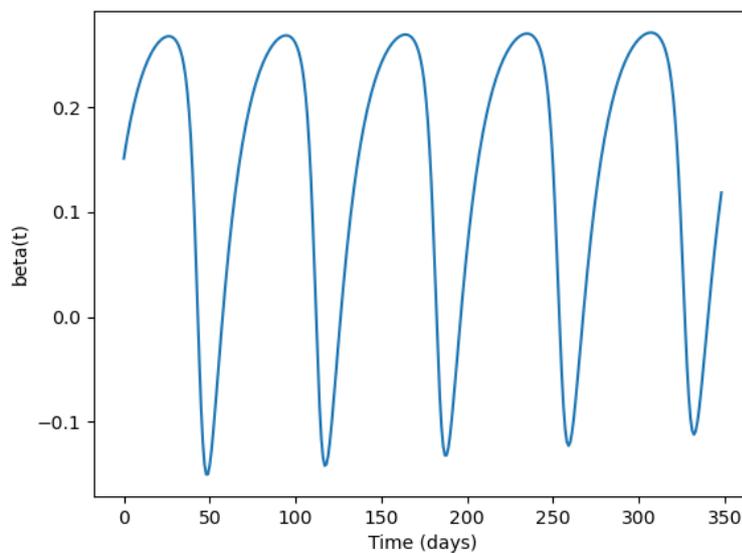
With the initial conditions

$$\begin{aligned}
 P_{tot} &= 17 \cdot 10^6, \\
 S(t_0) &= 16 \cdot 10^6, \\
 I(t_0) &= 3854, \\
 R(t_0) &= 0.988 \cdot 10^6, \\
 \beta(t_0) &= 1.21 \cdot \gamma, \\
 \beta_{high} &= 0.31,
 \end{aligned}$$

we show an example of the output of the first adaptation of the SIR model in Figure 16:



(a) Periodicity of the first adaptation of the SIR model (red) with the estimated data for reference (blue).



(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.31, k_1 = 0.074$ and $k_2 = 28.55$. Here, we also see periodicity of the $\beta(t)$ function.

Figure 16: Example of an output of the first adaptation of the SIR model and its $\beta(t)$ -plot.

We observe from Figure 16 that the model solutions are periodic, and that apparently the epidemic has

not died out yet. Another interesting (and more important) feature that we see is that the value of $\beta(t)$ becomes negative on several days. This is of course not possible in reality, having negative contacts per person per day. So we need to make sure that $\beta(t) > 0$ for all t . This will be shown in the next section.

5.2 Second adjustment of the SIR model

In order to prevent that $\beta(t)$ becomes negative, we can adjust the ODE for $\beta(t)$ from equation (57) by multiplying the RHS of the ODE by β . However, we will generalise this adjustment by introducing a new parameter β_{low} , which can be zero. It follows that this adjustment yields the equations:

$$\begin{aligned}\frac{d\beta}{dt} &= (\beta - \beta_{low}) \cdot (-k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I), \\ \frac{dI}{dt} &= c_4(\beta - \beta_{high})I - c_3I,\end{aligned}\tag{70}$$

where β_{low} is the lowest value of β during the pandemic. Like before, we choose $c_4 = S/P_{tot}$ and $c_3 = \gamma - c_4\beta_{high}$ to obtain the standard ODE equation for I :

$$\frac{dI}{dt} = \beta \frac{SI}{P_{tot}} - \gamma I.$$

5.2.1 Stability Analysis of the second adjustment of the SIR model

With this adjustment, we have the following three critical points:

$$(\beta_{low}, 0), (\beta_{high}, 0) \text{ and } \left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2} \right).$$

Let $g(\beta, I)$ be

$$g(\beta, I) = \begin{bmatrix} (\beta - \beta_{low}) \cdot (-k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I) \\ \beta I \frac{S}{P_{tot}} - \gamma I \end{bmatrix}.\tag{71}$$

Then the Jacobi matrix is given by:

$$\frac{\partial g(\beta, I)}{\partial(\beta, I)} = \begin{bmatrix} -k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I + (\beta - \beta_{low})(-k_1 + k_2I) & (\beta - \beta_{low})(k_2(\beta - \beta_{high})) \\ I \frac{S}{P_{tot}} & \beta \frac{S}{P_{tot}} - \gamma \end{bmatrix}.\tag{72}$$

Substitution of the critical point $(\beta_{low}, 0)$ yields the matrix

$$\begin{bmatrix} -k_1(\beta_{low} - \beta_{high}) & 0 \\ 0 & \beta_{low} \frac{S}{P_{tot}} - \gamma \end{bmatrix}.\tag{73}$$

We see that the eigenvalues are

$$\lambda_1 = -k_1(\beta_{low} - \beta_{high}) > 0, \quad \lambda_2 = \beta_{low} \frac{S}{P_{tot}} - \gamma.\tag{74}$$

If $\beta_{low} \frac{S}{P_{tot}} < \gamma$ ($\Leftrightarrow \beta_{low} < \gamma \frac{P_{tot}}{S}$), then $\lambda_2 < 0$ and thus we have a saddle point.

If $\beta_{low} \frac{S}{P_{tot}} > \gamma$ ($\Leftrightarrow \beta_{low} > \gamma \frac{P_{tot}}{S}$), then $\lambda_2 > 0$ and thus we have an unstable node.

Next, substitution of the point $(\beta_{high}, 0)$ yields

$$\begin{bmatrix} -k_1(\beta_{high} - \beta_{low}) & 0 \\ 0 & \beta_{high} \frac{S}{P_{tot}} - \gamma \end{bmatrix},\tag{75}$$

with eigenvalues

$$\lambda_1 = -k_1(\beta_{high} - \beta_{low}) < 0, \quad \lambda_2 = \beta_{high} \frac{S}{P_{tot}} - \gamma.\tag{76}$$

If $\beta_{high} \frac{S}{P_{tot}} < \gamma$ ($\Leftrightarrow \beta_{high} < \gamma \frac{P_{tot}}{S}$), then $\lambda_2 < 0$ and thus we have an attractor.

If $\beta_{high} \frac{S}{P_{tot}} > \gamma$ ($\Leftrightarrow \beta_{high} > \gamma \frac{P_{tot}}{S}$), then $\lambda_2 > 0$ and thus we have a saddle point.

Lastly, substitution of the point $\left(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2}\right)$ yields

$$\begin{bmatrix} 0 & (\gamma \frac{P_{tot}}{S} - \beta_{low}) (k_2 (\gamma \frac{P_{tot}}{S} - \beta_{high})) \\ \frac{k_1}{k_2} \frac{S}{P_{tot}} & 0 \end{bmatrix}. \quad (77)$$

Now setting the characteristic polynomial to zero:

$$\lambda^2 - a = 0, \quad (78)$$

with

$$a = \frac{k_1}{k_2} \frac{S}{P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\gamma \frac{P_{tot}}{S} - \beta_{high} \right) \right), \quad (79)$$

it follows that we have three cases:

If $\gamma \frac{P_{tot}}{S} < \beta_{low} < \beta_{high}$, then $a > 0$. Thus we have $\lambda_{1,2} = \pm \sqrt{a}$, a saddle node.

If $\gamma \frac{P_{tot}}{S} > \beta_{high} > \beta_{low}$, then $a > 0$. Thus we have $\lambda_{1,2} = \pm \sqrt{a}$, a saddle node.

If $\beta_{low} < \gamma \frac{P_{tot}}{S} < \beta_{high}$, then $a < 0$. Thus we have $\lambda_{1,2} = \pm i\sqrt{a}$, a center (in the linearised system).

To summarize the stability analysis of all three critical points, we observe that we can distinguish three cases:

1. If $\gamma \frac{P_{tot}}{S} < \beta_{low} < \beta_{high}$, then $(\beta_{low}, 0)$ is an unstable node, $(\beta_{high}, 0)$ is a saddle and $(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2})$ is a saddle node.
2. If $\beta_{low} < \gamma \frac{P_{tot}}{S} < \beta_{high}$, then $(\beta_{low}, 0)$ is a saddle node, $(\beta_{high}, 0)$ is a saddle and $(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2})$ is a center.
3. If $\beta_{low} < \beta_{high} < \gamma \frac{P_{tot}}{S}$, then $(\beta_{low}, 0)$ is a saddle node, $(\beta_{high}, 0)$ is an attractor and $(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2})$ is a saddle.

Case **1** happens when the epidemic is extremely severe, because the recovery rate γ is very low. This leads to an unstable situation. Case **2** is when the situation is more controllable, however (local) outbreaks and different waves of infections happen. Lastly, case **3** happens when the epidemic starts to die out. Note that these three cases happen in the order of **1** \rightarrow **2** \rightarrow **3**, and it is dependent on the specific values in which case a disease starts. So with COVID-19 in the Netherlands, we observe that we start more or less in case **2**.

5.2.2 Analytical solution - First method

For case **2**, we can find an analytical solution around the critical point $(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2})$. Substitution of this critical point in the Jacobi matrix yields

$$\begin{bmatrix} 0 & (\gamma \frac{P_{tot}}{S} - \beta_{low}) (k_2 (\gamma \frac{P_{tot}}{S} - \beta_{high})) \\ \frac{k_1}{k_2} \frac{S}{P_{tot}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -b \\ c & 0 \end{bmatrix}, \quad (80)$$

where

$$b = \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\beta_{high} - \gamma \frac{P_{tot}}{S} \right) \right) > 0, \quad c = \frac{k_1}{k_2} \frac{S}{P_{tot}} > 0. \quad (81)$$

Then the eigenvalues of the matrix are

$$\lambda_{1,2} = \pm i \sqrt{\frac{k_1}{k_2} \frac{S}{P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\beta_{high} - \gamma \frac{P_{tot}}{S} \right) \right)} \quad (82)$$

$$= \pm i \sqrt{cb}, \quad (83)$$

with corresponding eigenvectors

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \pm i \begin{bmatrix} \sqrt{b} \\ \sqrt{c} \\ 0 \end{bmatrix}. \quad (84)$$

Then the solution corresponding to λ_1 and its eigenvector is given by

$$\begin{aligned} \begin{bmatrix} \beta(t) \\ I(t) \end{bmatrix} &= \exp(i\sqrt{bct}) \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} + i \begin{bmatrix} \frac{\sqrt{b}}{\sqrt{c}} \\ 0 \end{bmatrix} \right), \\ &= \left(\cos(\sqrt{bct}) + i \sin(\sqrt{bct}) \right) \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} + i \begin{bmatrix} \frac{\sqrt{b}}{\sqrt{c}} \\ 0 \end{bmatrix} \right), \\ &= \begin{bmatrix} -\frac{\sqrt{b}}{\sqrt{c}} \sin(\sqrt{bct}) \\ \cos(\sqrt{bct}) \end{bmatrix} + i \begin{bmatrix} \frac{\sqrt{b}}{\sqrt{c}} \cos(\sqrt{bct}) \\ \sin(\sqrt{bct}) \end{bmatrix}. \end{aligned} \quad (85)$$

Then by Lemma 1 in chapter 3.9 from the book [25], the solution is given by

$$\begin{bmatrix} \beta(t) \\ I(t) \end{bmatrix} = \alpha_1 \begin{bmatrix} -\frac{\sqrt{b}}{\sqrt{c}} \sin(\sqrt{bct}) \\ \cos(\sqrt{bct}) \end{bmatrix} + \alpha_2 \begin{bmatrix} \frac{\sqrt{b}}{\sqrt{c}} \cos(\sqrt{bct}) \\ \sin(\sqrt{bct}) \end{bmatrix}, \quad (86)$$

where α_1, α_2 are constants.

5.2.3 Analytical solution - Second method

Substitution of the critical point $(\gamma \frac{P_{tot}}{S}, \frac{k_1}{k_2})$ in the Jacobi matrix yields

$$\begin{bmatrix} 0 & (\gamma \frac{P_{tot}}{S} - \beta_{low}) (k_2 (\gamma \frac{P_{tot}}{S} - \beta_{high})) \\ \frac{k_1}{k_2} \frac{S}{P_{tot}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -b \\ c & 0 \end{bmatrix}. \quad (87)$$

The linearised system is of the form $\mathbf{x}' = A\mathbf{x}$. Then the solution to the DE is equal to

$$\mathbf{x}(t) = e^{At} \mathbf{c}, \quad (88)$$

where

$$A = \begin{bmatrix} 0 & -b \\ c & 0 \end{bmatrix}. \quad (89)$$

Computing the matrix exponential e^A is done as follows. We can diagonalize the matrix A such that

$$A = PDP^{-1}, \quad (90)$$

where

$$P = \begin{bmatrix} -i\frac{\sqrt{b}}{\sqrt{c}} & i\frac{\sqrt{b}}{\sqrt{c}} \\ 1 & 1 \end{bmatrix}, \quad (91)$$

$$D = \begin{bmatrix} -i\sqrt{b}\sqrt{c} & 0 \\ 0 & i\sqrt{b}\sqrt{c} \end{bmatrix}, \quad (92)$$

$$P^{-1} = \begin{bmatrix} i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \\ -i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \end{bmatrix}. \quad (93)$$

Then it follows that

$$A^n = P D^n P^{-1} = P \begin{bmatrix} (-i\sqrt{b}\sqrt{c})^n & 0 \\ 0 & (i\sqrt{b}\sqrt{c})^n \end{bmatrix} P^{-1} \quad (94)$$

for $n = 1, 2, \dots$. The matrix exponential is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}. \quad (95)$$

Then with $A = PDP^{-1}$, it follows that

$$\begin{aligned} e^{At} &= \sum_{k=0}^{\infty} \frac{(At)^k}{k!} = \sum_{k=0}^{\infty} \frac{(PDP^{-1})^k t^k}{k!} = \sum_{k=0}^{\infty} \frac{PD^k t^k P^{-1}}{k!} = P \left(\sum_{k=0}^{\infty} \frac{(Dt)^k}{k!} \right) P^{-1}, \\ &= P \left(\sum_{k=0}^{\infty} \frac{1}{k!} \begin{bmatrix} (-i\sqrt{b}\sqrt{c})^k t^k & 0 \\ 0 & (i\sqrt{b}\sqrt{c})^k t^k \end{bmatrix} \right) P^{-1}, \\ &= P \left(\begin{bmatrix} e^{(-i\sqrt{b}\sqrt{c})t} & 0 \\ 0 & e^{(i\sqrt{b}\sqrt{c})t} \end{bmatrix} \right) P^{-1}. \end{aligned} \quad (96)$$

So we have that

$$\begin{aligned}
e^{At} &= P \left(\begin{bmatrix} e^{(-i\sqrt{b}\sqrt{c})t} & 0 \\ 0 & e^{(i\sqrt{b}\sqrt{c})t} \end{bmatrix} \right) P^{-1} = \begin{bmatrix} -i\frac{\sqrt{b}}{\sqrt{c}} & i\frac{\sqrt{b}}{\sqrt{c}} \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} e^{(-i\sqrt{b}\sqrt{c})t} & 0 \\ 0 & e^{(i\sqrt{b}\sqrt{c})t} \end{bmatrix} \cdot \begin{bmatrix} i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \\ -i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \end{bmatrix}, \\
&= \begin{bmatrix} -i\frac{\sqrt{b}}{\sqrt{c}} & i\frac{\sqrt{b}}{\sqrt{c}} \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos(\sqrt{bct}) - i\sin(\sqrt{bct}) & 0 \\ 0 & \cos(\sqrt{bct}) + i\sin(\sqrt{bct}) \end{bmatrix} \cdot \begin{bmatrix} i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \\ -i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \end{bmatrix}, \\
&= \begin{bmatrix} -i\frac{\sqrt{b}}{\sqrt{c}} \left(\cos(\sqrt{bct}) - i\sin(\sqrt{bct}) \right) & i\frac{\sqrt{b}}{\sqrt{c}} \left(\cos(\sqrt{bct}) + i\sin(\sqrt{bct}) \right) \\ \cos(\sqrt{bct}) - i\sin(\sqrt{bct}) & \cos(\sqrt{bct}) + i\sin(\sqrt{bct}) \end{bmatrix} \cdot \begin{bmatrix} i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \\ -i\frac{\sqrt{c}}{2\sqrt{b}} & 1/2 \end{bmatrix}, \\
&= \begin{bmatrix} \cos(\sqrt{bct}) & -\frac{\sqrt{b}}{\sqrt{c}} \sin(\sqrt{bct}) \\ \frac{\sqrt{c}}{\sqrt{b}} \sin(\sqrt{bct}) & \cos(\sqrt{bct}) \end{bmatrix}.
\end{aligned} \tag{97}$$

Thus, the analytical solution is given by

$$\begin{aligned}
\mathbf{x}(t) &= e^{At} \mathbf{c} = \begin{bmatrix} \cos(\sqrt{bct}) & -\frac{\sqrt{b}}{\sqrt{c}} \sin(\sqrt{bct}) \\ \frac{\sqrt{c}}{\sqrt{b}} \sin(\sqrt{bct}) & \cos(\sqrt{bct}) \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \\
&= c_1 \begin{bmatrix} \cos(\sqrt{bct}) \\ \frac{\sqrt{c}}{\sqrt{b}} \sin(\sqrt{bct}) \end{bmatrix} + c_2 \begin{bmatrix} -\frac{\sqrt{b}}{\sqrt{c}} \sin(\sqrt{bct}) \\ \cos(\sqrt{bct}) \end{bmatrix},
\end{aligned} \tag{98}$$

where c_1 and c_2 are constants.

5.2.4 Period of the analytical solution

The period of the solutions can be calculated by setting

$$\sqrt{cb} = \sqrt{\frac{k_1 S}{k_2 P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\beta_{high} - \gamma \frac{P_{tot}}{S} \right) \right)} = \frac{2\pi}{T_{period}}. \tag{99}$$

Now, if we set $\beta_{high} = 0.4, \beta_{low} = 0.1$ (because then $\beta_{low} < \gamma$) and we assume at the start of the pandemic, $P_{tot} \approx S$, we have that

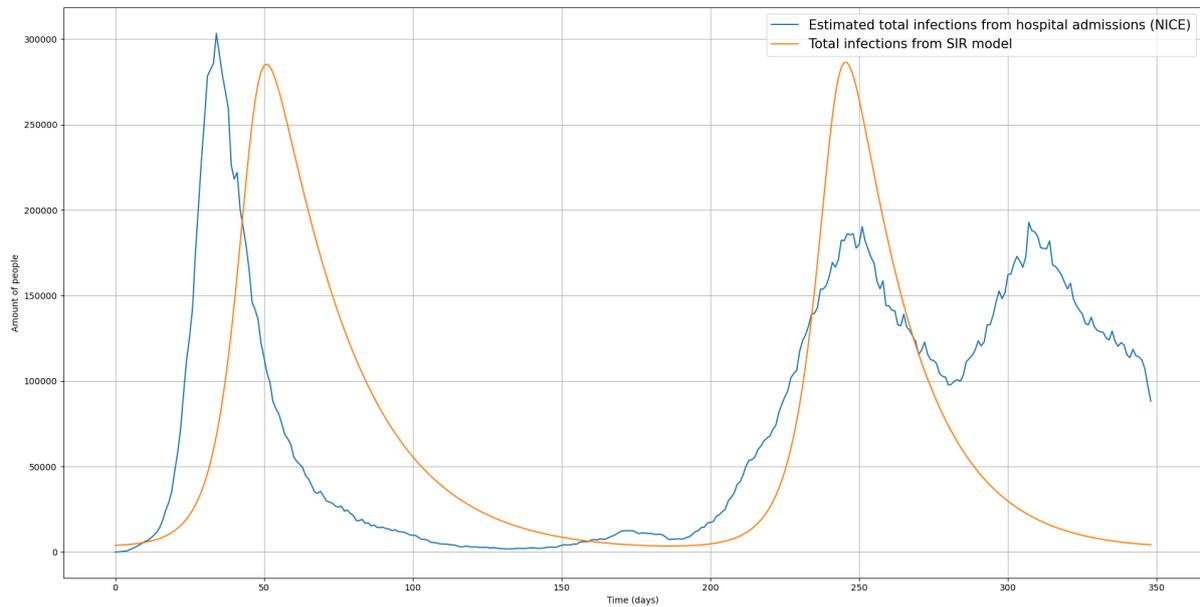
$$\begin{aligned}
\sqrt{cb} &= \sqrt{\frac{k_1 S}{k_2 P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\beta_{high} - \gamma \frac{P_{tot}}{S} \right) \right)} = \frac{2\pi}{T_{period}}, \\
&\Rightarrow \frac{k_1 S}{k_2 P_{tot}} \left(\gamma \frac{P_{tot}}{S} - \beta_{low} \right) \left(k_2 \left(\beta_{high} - \gamma \frac{P_{tot}}{S} \right) \right) = \left(\frac{2\pi}{T_{period}} \right)^2, \\
&\approx \frac{k_1}{k_2} \left(\frac{1}{8} - 0.1 \right) \left(k_2 \left(0.4 - \frac{1}{8} \right) \right) = \left(\frac{2\pi}{T_{period}} \right)^2, \\
&\approx k_1 \left(\frac{1}{8} - 0.1 \right) \left(0.4 - \frac{1}{8} \right) = \left(\frac{2\pi}{T_{period}} \right)^2, \\
&\Rightarrow k_1 \approx \frac{1}{0.025} \frac{1}{0.275} \left(\frac{2\pi}{T_{period}} \right)^2.
\end{aligned} \tag{100}$$

We note that k_1 determines more or less the time period between each infection wave, and k_2 determines how fast the number of infected people rises. Note also that in this case, we have that k_2 is a 'free' parameter, which should be fitted according to the data.

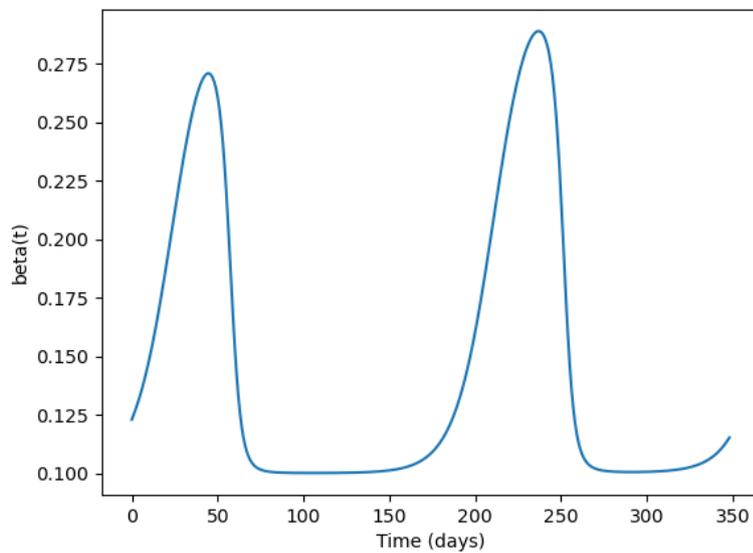
5.2.5 Results of the second adjustment of the SIR model

We will show a few results with the second adjusted version of the SIR model. The initial conditions in Figure 17 will be the same as in Figure 16, and we set $\beta_{low} = 0.1$ because then $\beta_{low} < \gamma$, which corresponds to case 2.

With the adjusted version, we see in Figure 17b that the $\beta(t)$ function is no longer smaller than β_{low} , which is what we wanted. Furthermore in Figure 17a, from a qualitative point of view, we were able to choose parameters in order to get two waves of infections that are around the same time as the waves from the the estimated data.

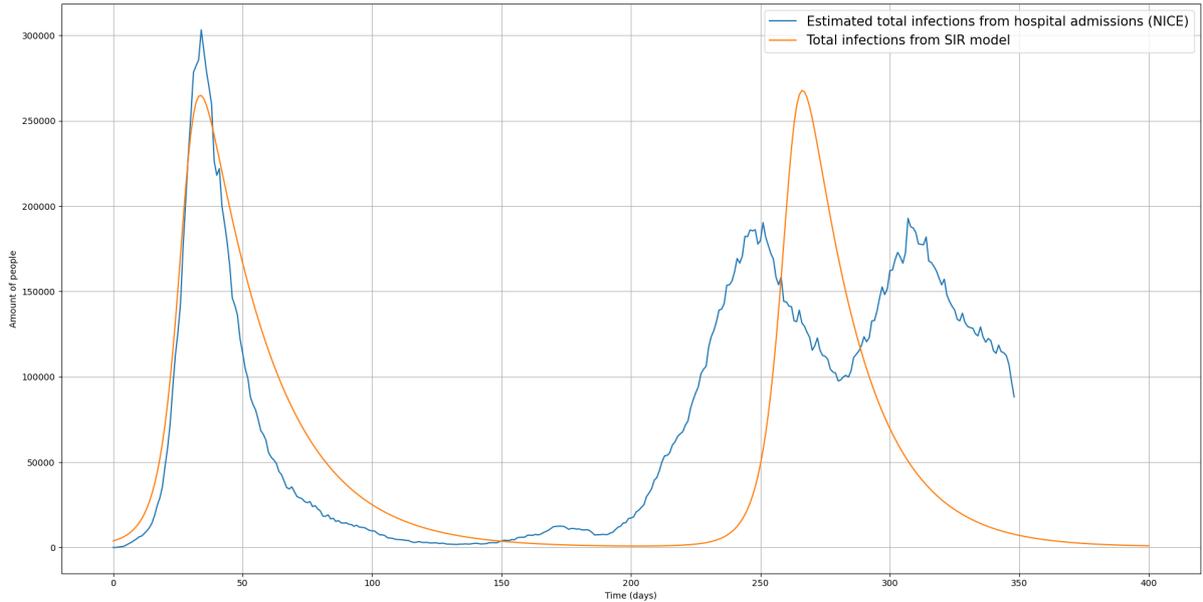


(a) Two waves of infections from the second adjustment of the SIR model (orange) that are around the same time as the two waves from the estimated data.

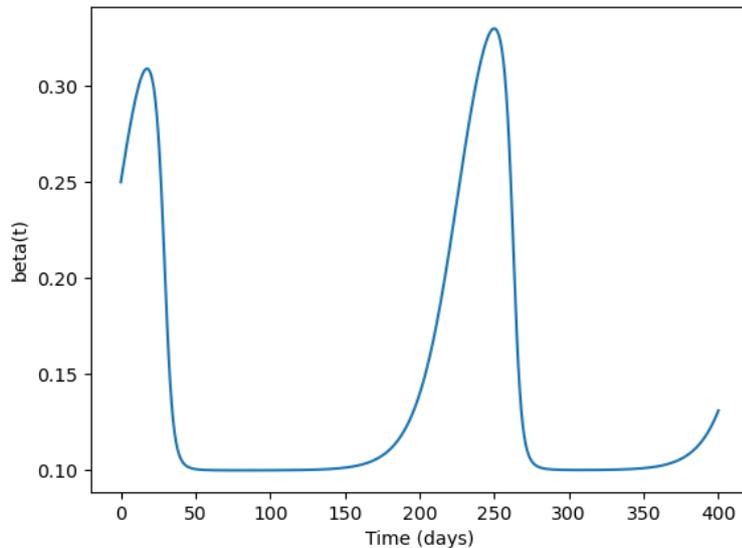


(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.31, \beta_{low} = 0.1, \beta(t_0) = 0.16, k_1 = 0.4$ and $k_2 = 100$. Note that the two spikes occur a few days earlier than the spikes of the model.

Figure 17: Plot of an output of the second adjustment of the SIR model and its $\beta(t)$ -function. Here, the model produces two infections waves and the $\beta(t)$ -function is non-negative.



(a) Second adjustment of the SIR model with different initial values. Here, two waves of infections are produced by the SIR model (orange), but now the first wave of the SIR model is in line with the first wave of the data. The second wave of the SIR model is shifted more towards the middle of the second wave of the data.



(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.4$, $\beta_{low} = 0.1$, $\beta(t_0) = 0.25$, $k_1 = 0.239$ and $k_2 = 83$.

Figure 18: Another plot of an output of the adjusted Lotka-Volterra SIR model and its $\beta(t)$ -function. Now with other initial values.

In Figure 18, we chose to start with different initial values for β_{high} and $\beta(t_0)$. We set $\beta(t_0) = 0.25$, because here we considered the entire data set, so that t_0 starts at February 17th 2020. Thus, with equation (49), we have that $\mathcal{R}_0 = 2$, which corresponds more or less to what the RIVM calculated around that time. For β_{high} , we also increase the value to 0.4 because the basic reproductive number may be higher than what the RIVM had calculated.

With the change in initial values, we see in Figure 18a that the first wave from the SIR model is in line with the first wave from the estimated data. The second wave from the SIR model also occurs around the same time as the second wave from the estimated data. However, the second wave from the estimated data is much wider.

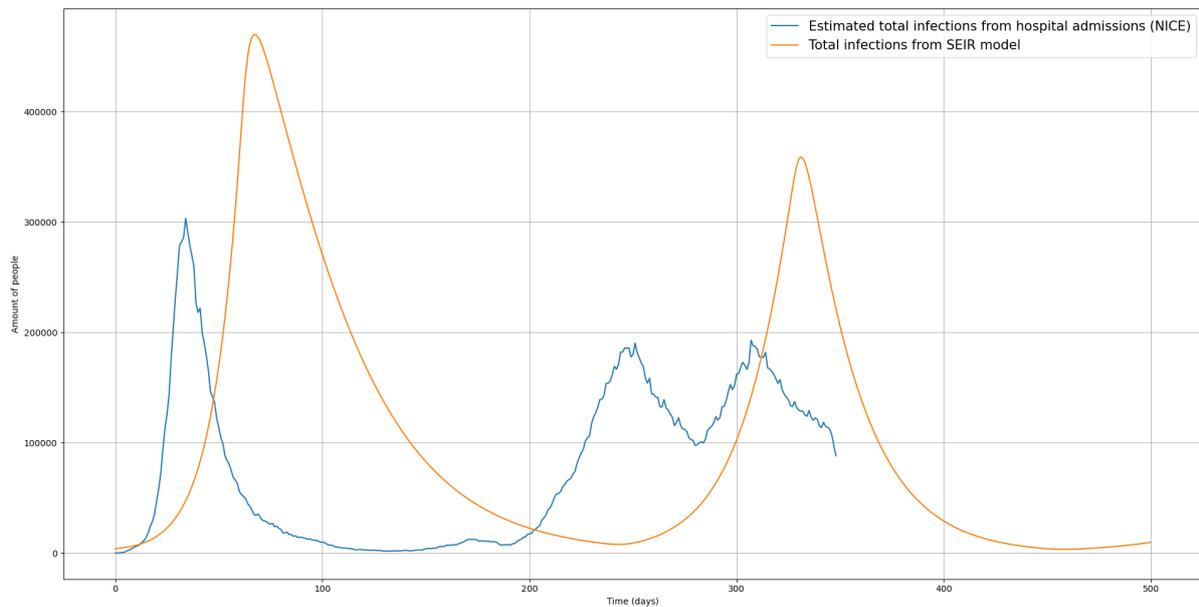
5.3 Lotka-Volterra and the SEIR model

So far, we applied the Lotka-Volterra equations on the SIR model, which gave us a few qualitative good results. Now, we will return to the SEIR model again, where the SEIR equations are given by (35) with

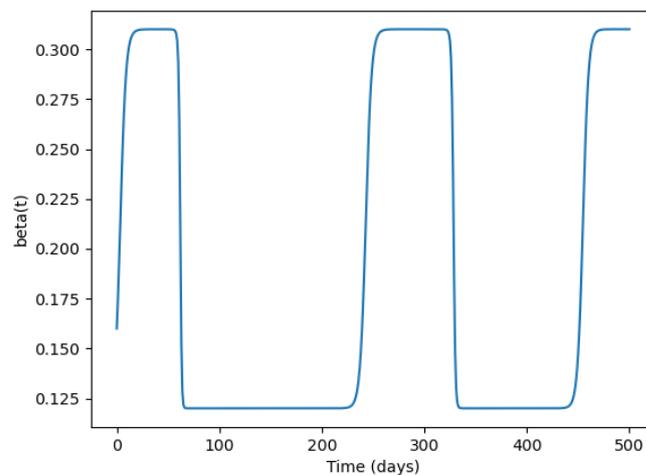
$$\frac{d\beta}{dt} = (\beta - \beta_{low}) \cdot (-k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I). \quad (101)$$

This time, we will not analyse the stability of the system. Instead, we will immediately show the results of the model output. In Figure 19 and Figure 20, we used the values $\beta_{high} = 0.31$ and $\beta(t_0) = 0.16$. In Figure 21 and Figure 22 we used $\beta_{high} = 0.4$ and $\beta(t_0) = 0.25$. These values were used in the results of the second adjustment of the SIR model as well.

5.3.1 Results with parameters $\beta_{high} = 0.31$ & $\beta(t_0) = 0.16$



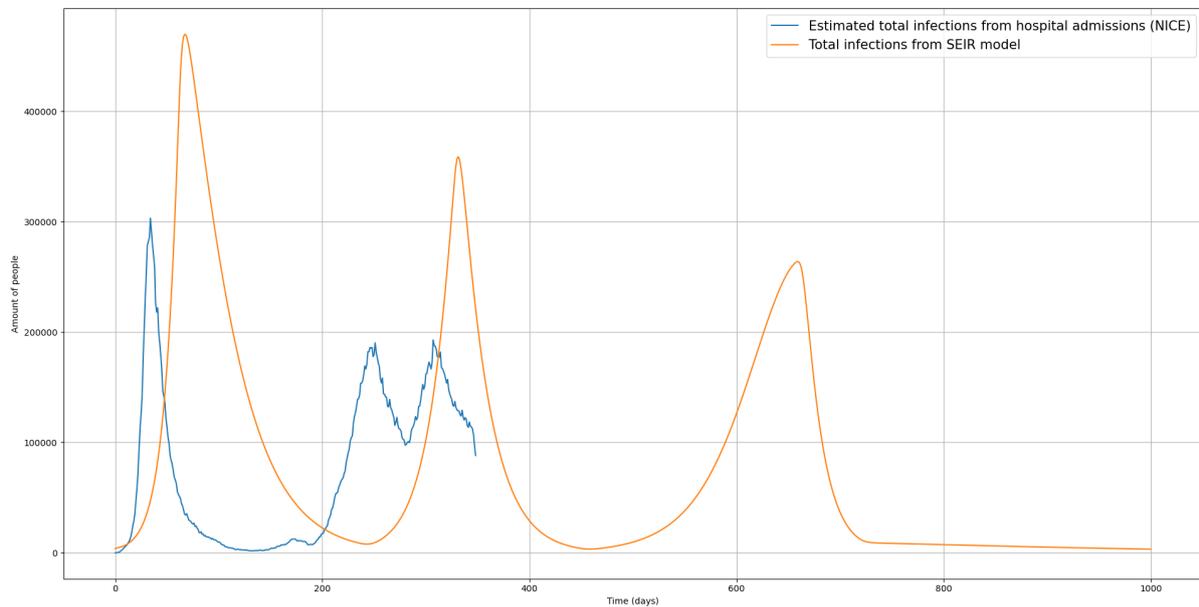
(a) Output of the adjusted SEIR model. Periodicity of the adjusted SEIR model is observed by having two waves of infections. However, the waves of the model are much larger and wider compared to the waves of the estimated data. Also, the waves of the model occur at later moments.



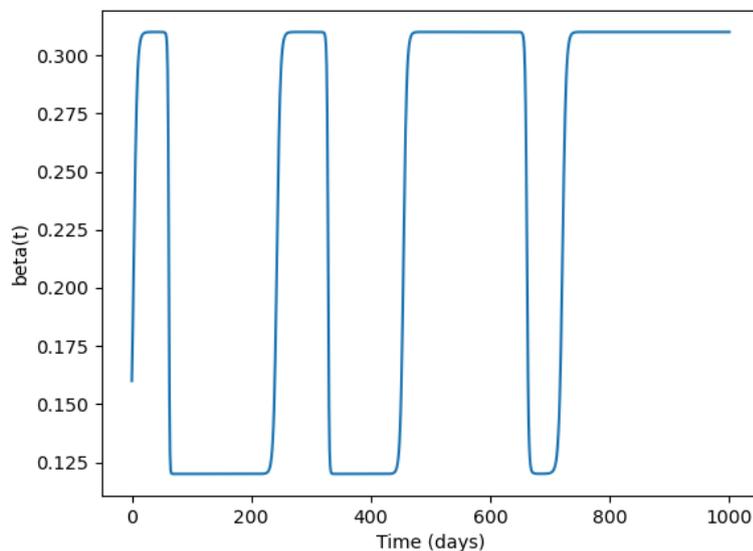
(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.31, \beta_{low} = 0.12, \beta(t_0) = 0.16, k_1 = 2.05$ and $k_2 = 375$.

Figure 19: Results of the adjusted SEIR model and its $\beta(t)$ -function.

With our first parameters, we observe in Figure 19a that the model produces two waves of infections, where the first wave is larger than the second wave. However, the peaks of the model are much higher and wider (especially the first wave), compared to the estimated data. This means that the model overestimates the number of infected people. Moreover, we see that the two waves of the model occur at a later moment, which is also not completely in line with the estimated data.



(a) Same output of the SEIR model as in Figure 19 if the model runs for a longer period of time. We observe the occurrence of a third wave and the epidemic eventually dies out if no outside-effects are present.

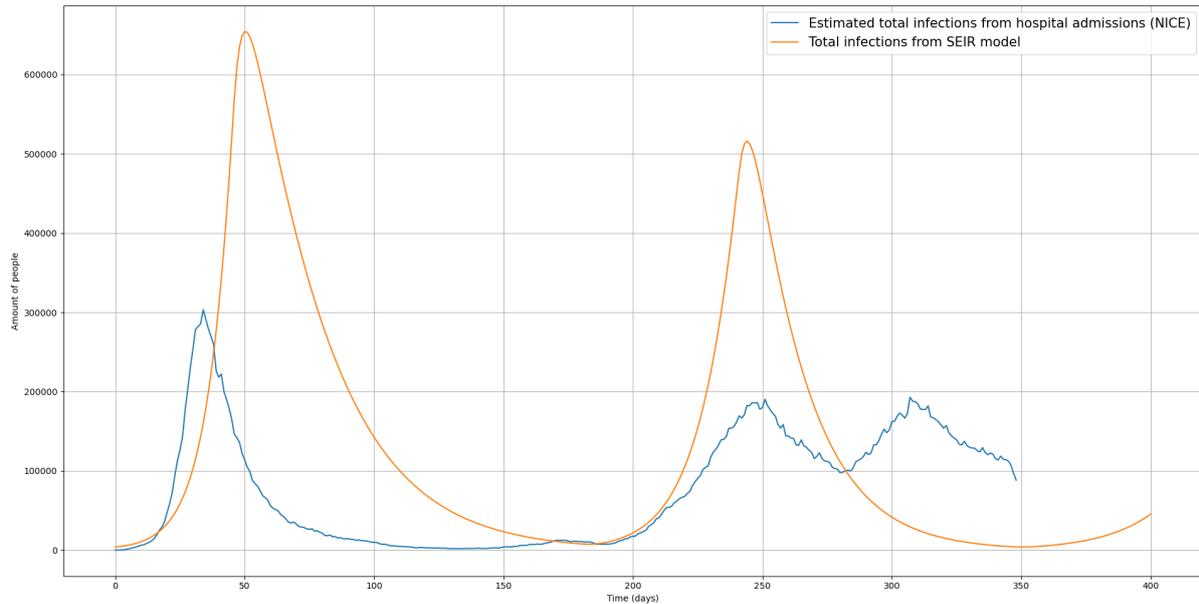


(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.31, \beta_{low} = 0.12, \beta(t_0) = 0.16, k_1 = 2.05$ and $k_2 = 375$.

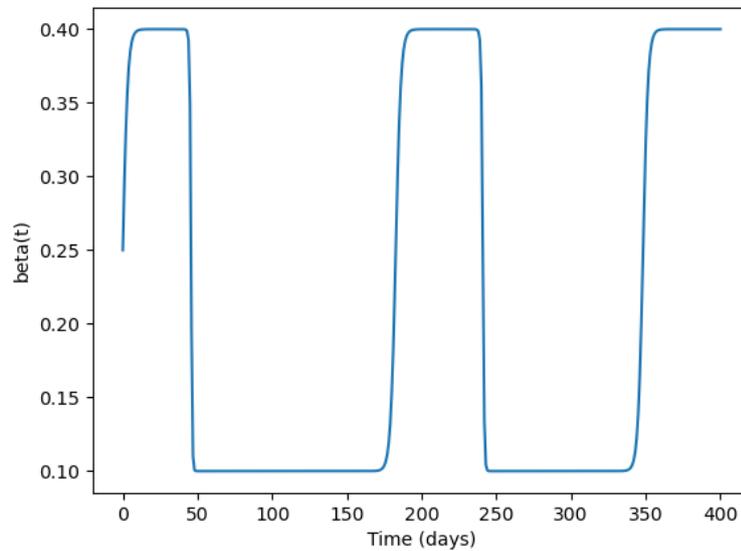
Figure 20: Results of the SEIR model that runs for a longer period of time. The initial values are the same as in Figure 19.

If we run the model for a longer period of time, we see in Figure 20a that the model produces a third wave a bit after 600 days, which means around September-October 2021. After that, the epidemic dies out because of the change in phase space behaviour. Note that we did not include any outside-effects such as (new) variants of the virus that may be more contagious, or inclusion of vaccination programs.

5.3.2 Results with parameters $\beta_{high} = 0.4$ & $\beta(t_0) = 0.25$



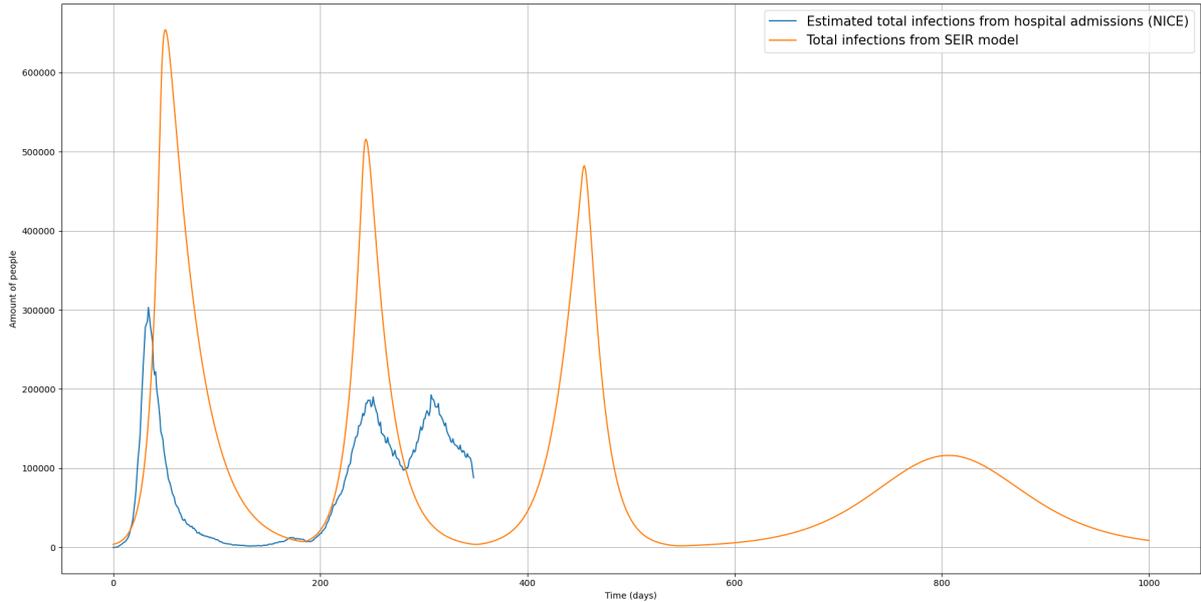
(a) The adjusted SEIR model with different parameters. Compared to Figure 19a, we see that the waves of infections in this figure is shifted more towards left, meaning that the waves of infections start earlier. Same as in Figure 19a, the waves of infections are much higher than the waves of the estimated data.



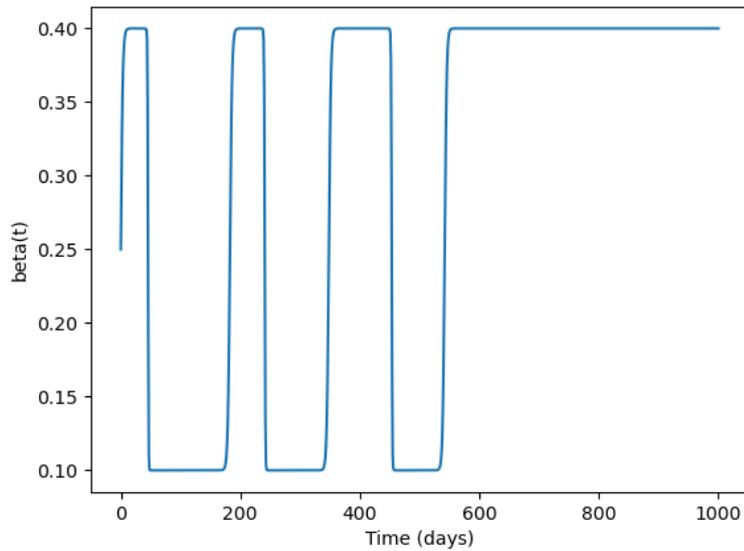
(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.4, \beta_{low} = 0.1, \beta(t_0) = 0.25, k_1 = 2.05$ and $k_2 = 305$.

Figure 21: Result of the adjusted SEIR model and its $\beta(t)$ -function with different parameters.

With the parameters $\beta_{high} = 0.4$ and $\beta(t_0) = 0.25$, we see in Figure 21a that the waves of infections of the model occur earlier than in Figure 19a. This means that these waves are now more in line with waves from the estimated data. However again, the model overestimates the number of infected people.



(a) Same output of the adjusted SEIR model as in Figure 21a, but now the model runs for a longer period of time. We observe again that the epidemic eventually dies out, but now with the addition of a third and fourth wave.



(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.4, \beta_{low} = 0.1, \beta(t_0) = 0.25, k_1 = 2.05$ and $k_2 = 305$.

Figure 22: Same output as in Figure 21, but now the model runs for a longer period of time.

Running the model for a longer period of time again, we observe in Figure 22a that the epidemic eventually dies out as well. Though again, the model overestimates the number of infections. We note also that the model produces a third wave after day 400 (April 2021) and a fourth wave around day 800 (May 2022). The time period between the third and fourth wave are larger than the time period between the other waves. Perhaps this suggests that the majority of the population has become immune to the virus after the third wave.

5.4 Conclusion & Discussion

In this section, we have seen how we can apply the Lotka-Volterra equations into the SIR and SEIR model. The basic Lotka-Volterra model has periodic solutions, and we wanted to use that periodicity

to model the infection waves that have occurred in the Netherlands. This can be seen in the results of both the SIR and SEIR models. From a qualitative point of view, we were able to match the waves of infections of the S(E)IR model with the estimated data. However, the actual number of infections did not match completely with the estimated data, meaning that the model with Lotka-Volterra overestimates the number of infections. Nevertheless, this new equation for $\beta(t)$ is not too complicated, while also being able to model infection waves through periodicity. For a quantitative match of the model with the estimated data, It would be interesting for future research if it is possible to fine-tune the function $\beta(t)$ such that the function becomes country-specific. If COVID-19 is also affected by seasonal events (such as warmer/sunnier climates), this may also be incorporated in the $\beta(t)$ -function.

6 SEIR model with gamma distributed stages

The standard SEIR model implicitly assumes that both the incubation time period and the infectious time period are exponentially distributed with mean $1/\alpha$ and $1/\gamma$ respectively. However, assuming this distribution, it means in the case of the infectious time period that the probability of recovery within a given time interval is constant, regardless of the time since infection, which is unrealistic [7, 8]. In reality we see that the chance of recovery in a given interval is initially small but increases over time. Therefore the standard SEIR model needs to be adjusted with use of a different distribution. One of the candidates that can be used instead is the gamma distribution. In this section we will discuss the SEIR model where the exposed stage and the infectious stage are gamma distributed.

First, we will discuss how to adjust the standard SEIR model in Section 6.1. After that, we will show a few numerical solutions in Sections 6.2 and 6.3, where we use both the ODE for $\beta(t)$ from Section 4 and the Lotka-Volterra equations for $\beta(t)$ from Section 5.

6.1 The model equations

In these papers [7, 8, 9], the authors proposed a modification to the standard SEIR model with use of a gamma distribution. To incorporate the gamma distribution, we subdivide the exposed compartment E and the infected compartment I into m and n stages respectively; this is also called the method of stages. Then, we assume that the time spent in each substage of E is exponentially distributed with mean $1/\alpha$. Consequently, the total time spent in the m substages of E is the sum of m independent identical exponential distributions. From probability theory, we know that this corresponds to a gamma distribution. Specifically, the probability density function of this gamma distribution is given by

$$f_E(t) = \frac{(\alpha m)^m}{\Gamma(m)} t^{m-1} e^{-\alpha m t}, \quad (102)$$

where $\Gamma(m)$ is the gamma function. Note that the mean of the gamma distribution equals $m/(\alpha m) = 1/\alpha$, which is the same mean for the incubation time period when the exponential distribution is used in the standard SEIR model.

Similarly for the infectious time period, we have a gamma distribution for the total time spent in the n substages of I , which is given by

$$f_I(t) = \frac{(\gamma n)^n}{\Gamma(n)} t^{n-1} e^{-\gamma n t}. \quad (103)$$

We note again that the mean of this distribution equals $n/(\gamma n) = 1/\gamma$, which is the same mean for the infectious time period when the exponential distribution is used in the standard SEIR model.

With the same means $1/\alpha$ and $1/\gamma$ for both the standard SEIR model and this adjusted SEIR model, it enables us to compare these models, where the only difference is in their distributions of incubation time periods and infectious time periods.

Using this method of stages, the modified SEIR model is given by the following equations:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{SI}{P_{tot}}, \\
\frac{dE_1}{dt} &= \beta \frac{SI}{P_{tot}} - m\alpha E_1, \\
\frac{dE_2}{dt} &= m\alpha E_1 - m\alpha E_2, \\
&\vdots \\
\frac{dE_m}{dt} &= m\alpha E_{m-1} - m\alpha E_m, \\
\frac{dI_1}{dt} &= m\alpha E_n - n\gamma I_1, \\
\frac{dI_2}{dt} &= n\gamma I_1 - n\gamma I_2, \\
&\vdots \\
\frac{dI_n}{dt} &= n\gamma I_{n-1} - m\gamma I_n, \\
\frac{dR}{dt} &= n\gamma I_n.
\end{aligned} \tag{104}$$

Note that when $m = 1$ and $n = 1$, we have the standard SEIR model again. Next, we have that the total number of exposed E and infected persons I are given by $E = \sum_{k=1}^m E_k$ and $I = \sum_{i=1}^n I_i$. Furthermore, we will use either

$$\frac{d\beta}{dt} = k_1(\beta_{high} - \beta) - k_2 \frac{\beta I}{P_{tot}}$$

or

$$\frac{d\beta}{dt} = (\beta - \beta_{low}) \cdot (-k_1(\beta - \beta_{high}) + k_2(\beta - \beta_{high})I)$$

to model the time-dependent $\beta(t)$ -function.

We will also point out that the amount of stages does not need to have a biological meaning of the infection itself. The subdivision of these stages is merely a mathematical device used to consider non-exponential incubation/infectious time periods [8]. Although in some cases it might be possible to give these stages a biological meaning, such as early-late incubation stage, or early-late infection stage.

6.2 Results with first ODE for $\beta(t)$

We use the same initial values from the previous section:

$$\begin{aligned}
P_{tot} &= 17 \cdot 10^6, \\
S(t_0) &= 16 \cdot 10^6, \\
E(t_0) &= 4510, \\
I(t_0) &= 3854, \\
R(t_0) &= 0.988 \cdot 10^6, \\
\beta(t_0) &= 0.16, \\
\beta_{high} &= 0.31.
\end{aligned}$$

Because we have m and n stages for the exposed and infected compartments respectively, we will distribute the initial values $E(t_0)$ and $I(t_0)$ over their respective stages homogeneously (if possible, otherwise rounding the numbers to make sure we have integer valued numbers). So we have

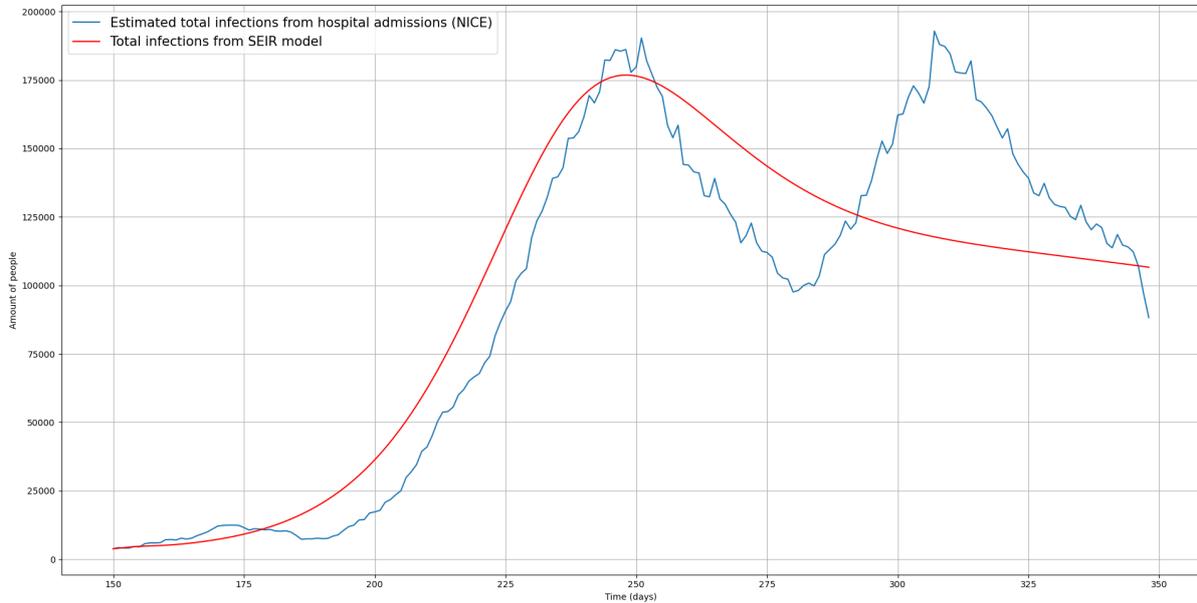
$$E_k(t_0) = \frac{E(t_0)}{m}, \quad I_i(t_0) = \frac{I(t_0)}{n}, \tag{105}$$

for $k = 1, 2, \dots, m$ and $i = 1, 2, \dots, n$. The parameters we want to fit are $k_1, k_2, \beta_{high}, \gamma, \alpha$ and $\beta(t_0)$. We will study two cases: 5 exposed and 5 infected stages in Section 6.2.1 and 10 exposed and 10 infected stages in Section 6.2.2.

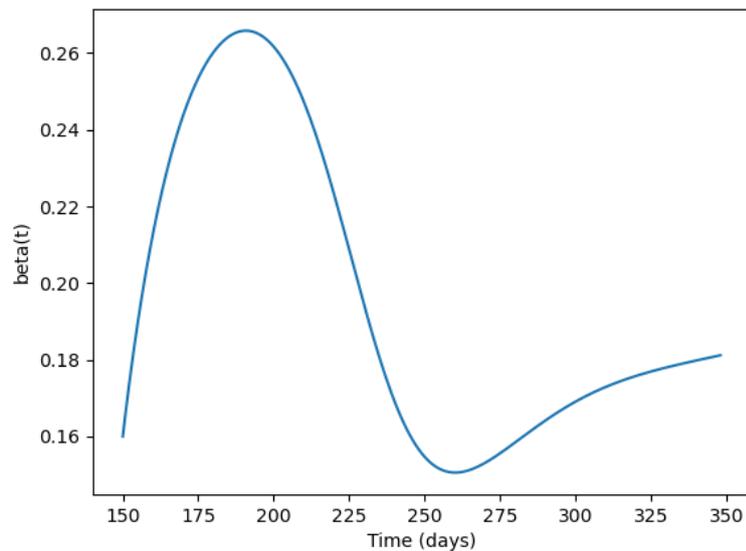
6.2.1 5 Exposed and 5 Infected stages

In Figures 23a and 24a, we will show the total number of infected people as a function of time after fitting with $n, m = 5$ stages for the exposed and infected compartments. We observe only marginal difference with the preliminary results obtained in Section 4. We also observe that varying β_{high} and $\beta(t_0)$ only leads to marginal differences as well.

Parameters $\beta_{high} = 0.31, \beta(t_0) = 0.16$ (from Section 4).



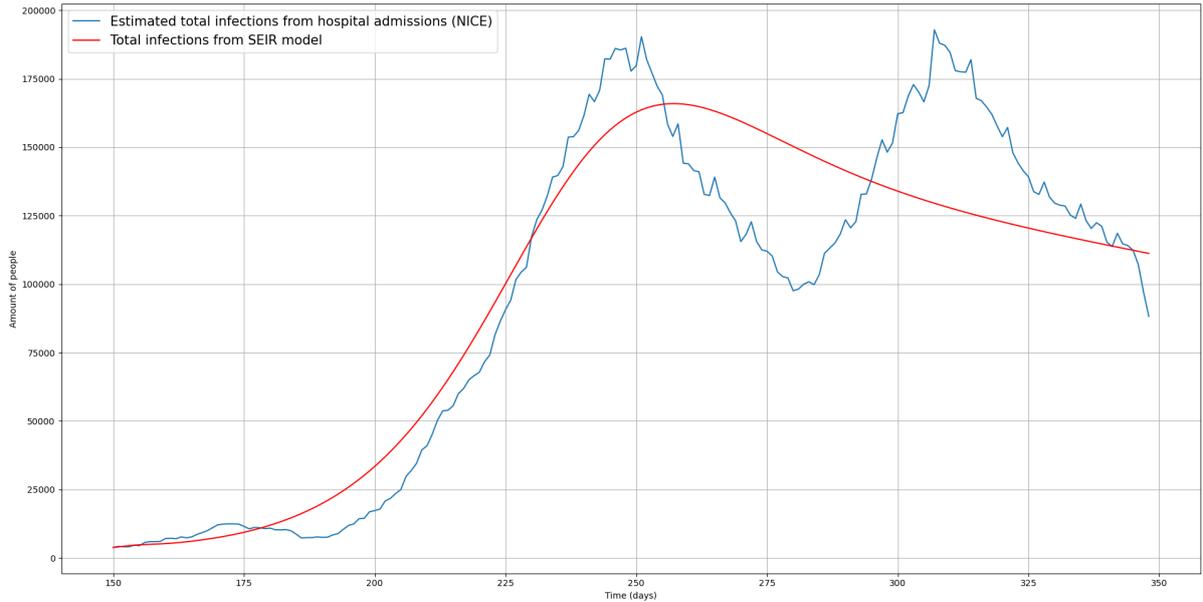
(a) Model solution (red) of the SEIR model with 5 exposed and 5 infected stages, where the first ODE for $\beta(t)$ is used. The estimated data (blue) is also shown for reference. We observe that the difference between the results of the standard SEIR model in Section 4 and the SEIR model with gamma distributed stages is small.



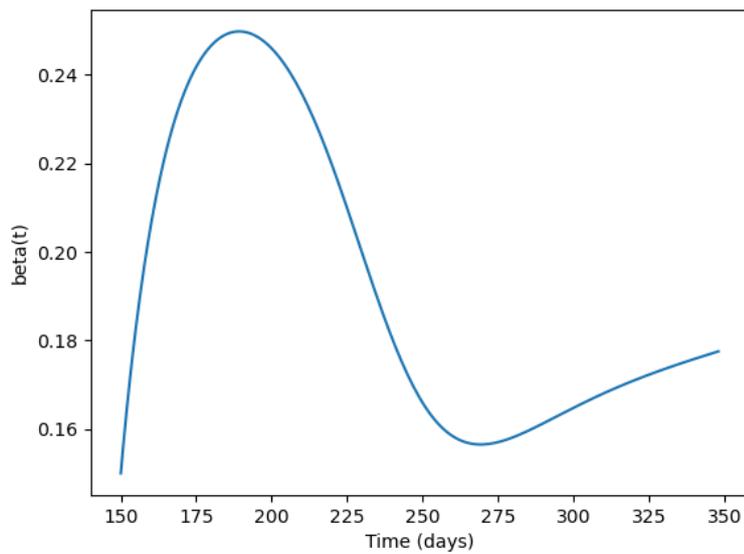
(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.30, \beta(t_0) = 0.16, k_1 = 0.050$ and $k_2 = 5.0$.

Figure 23: Result of the SEIR model with 5 exposed and 5 infected stages. Also the $\beta(t)$ -function is shown with the fitted parameters.

Parameters $\beta_{high} = 0.28, \beta(t_0) = 0.15$ (from Section 4).



(a) Model solution of the SEIR model with 5 exposed and 5 infected stages, where the first ODE for $\beta(t)$ is used. Now, different parameters are used as well. We observe that varying the parameter values leads to a minor difference in the results.



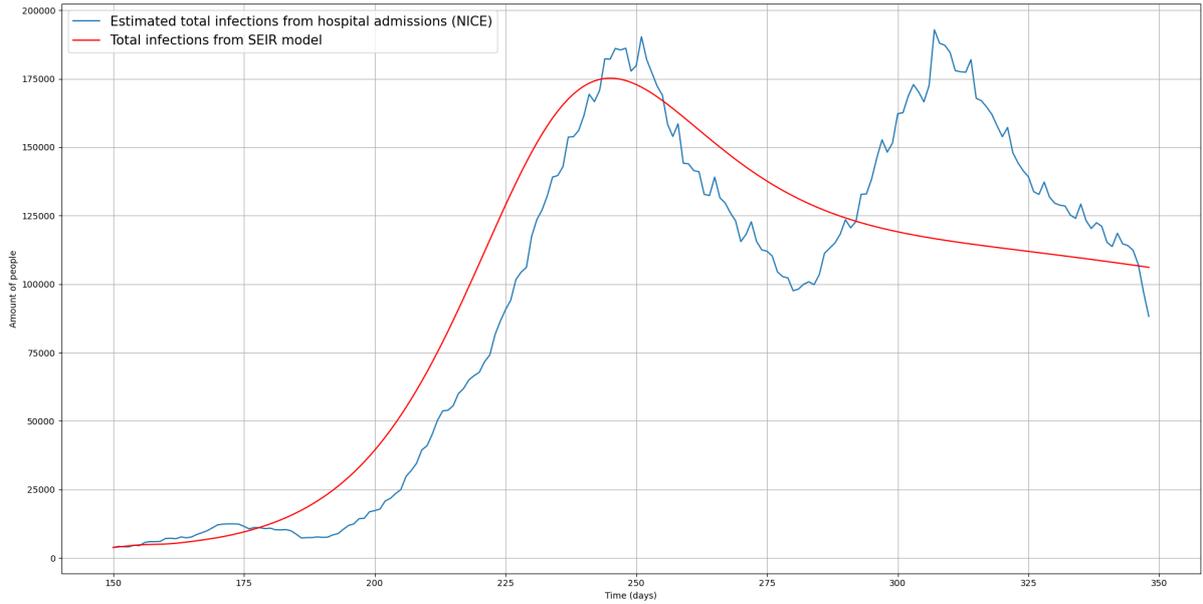
(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.27$, $\beta(t_0) = 0.15$, $k_1 = 0.066$ and $k_2 = 5.0$.

Figure 24: Results of the SEIR model with 5 exposed and 5 infected stages. Here, different parameter values are used.

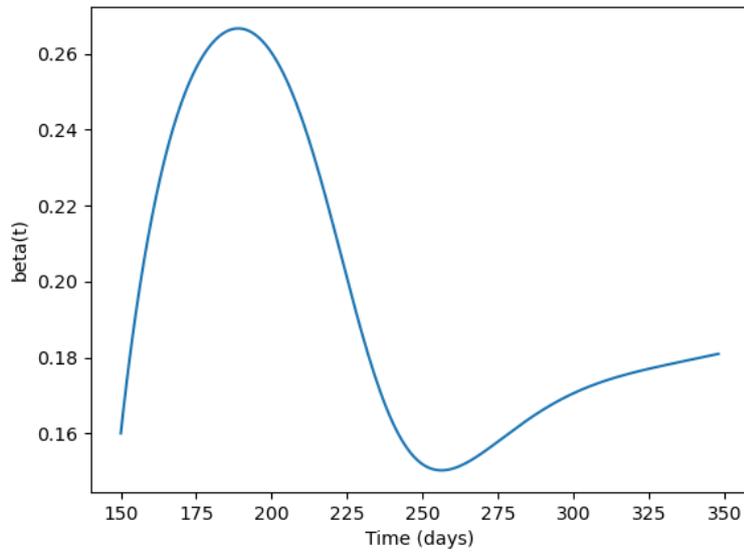
6.2.2 10 Exposed and 10 Infected stages

In Figures 25a and 26a, we will show the total number of infected people as a function of time after fitting with $n, m = 10$ stages for the exposed and infected compartments. Again we observe only marginal difference with the preliminary results obtained in Section 4 and we also observe that varying β_{high} and $\beta(t_0)$ leads to minor differences as well. Finally, we observe that varying the number of stages seems to have minimal effect.

Parameters $\beta_{high} = 0.31$, $\beta(t_0) = 0.16$



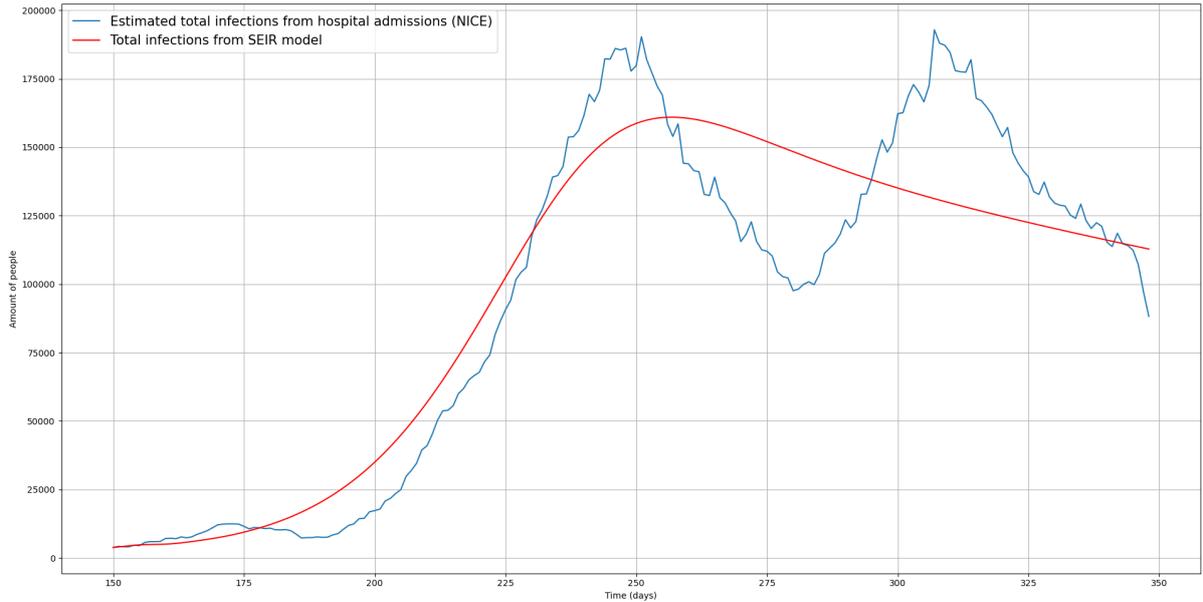
(a) Model solution (red) of the SEIR model with 10 exposed and 10 infected stages, where the first ODE for $\beta(t)$ is used. The estimated data (blue) is also shown for reference. We observe that the difference between the results of the standard SEIR model in Section 4 and the SEIR model with gamma distributed stages is small.



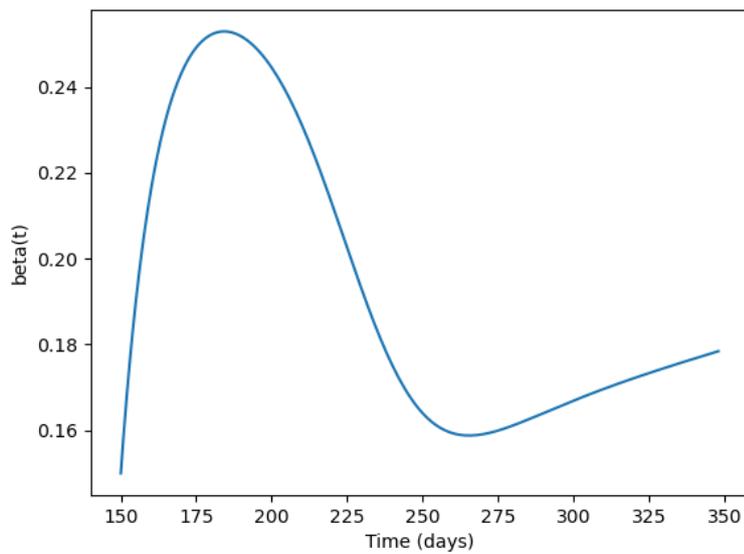
(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.3, \beta(t_0) = 0.16, k_1 = 0.053$ and $k_2 = 5.41$.

Figure 25: Result of the SEIR model with 10 exposed and 10 infected stages. Also the $\beta(t)$ -function is shown with the fitted parameters.

Parameters $\beta_{high} = 0.28, \beta(t_0) = 0.15$



(a) Model solution of the SEIR model with 10 exposed and 10 infected stages, where the first ODE for $\beta(t)$ is used. Now, different parameters are used as well. We observe again that varying the parameter values leads to a minor difference in the results.

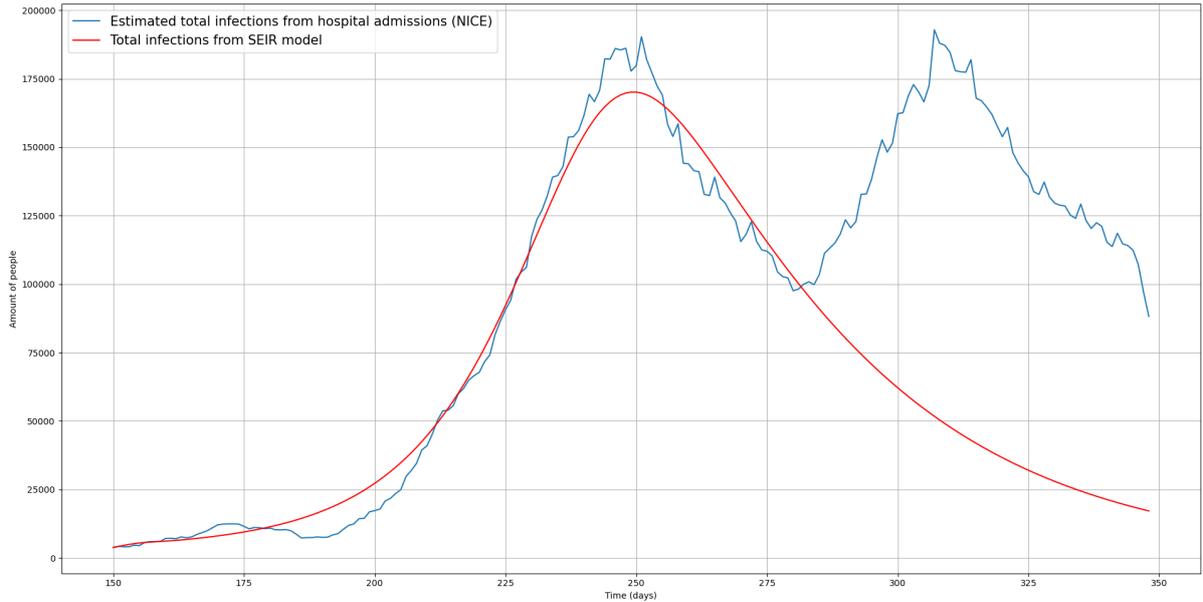


(b) Plot of $\beta(t)$ with parameters $\beta_{high} = 0.27, \beta(t_0) = 0.15, k_1 = 0.084$ and $k_2 = 6.25$.

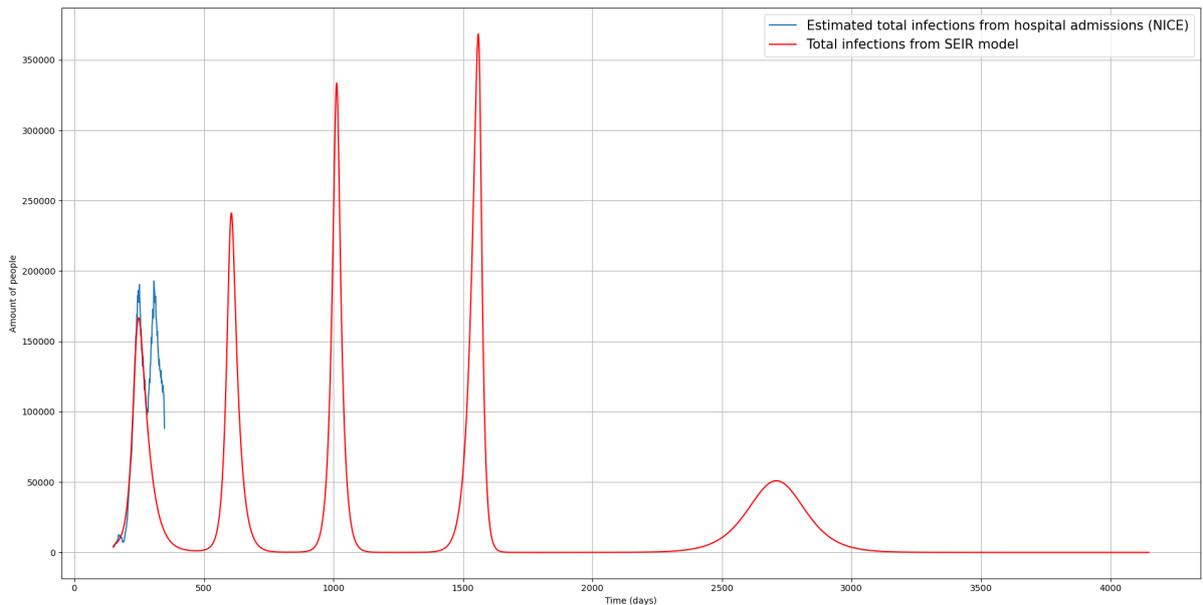
Figure 26: Results of the SEIR model with 5 exposed and 5 infected stages. Here, different parameter values are used.

6.3 Results with Lotka-Volterra

The Lotka-Volterra model is able to capture consecutive peaks which allows us to fit our model solutions to the entire estimated data set. So in this section we fit our model equations with 5 exposed and 5 infected stages and with 10 exposed and 10 infected stages to the data. Also in this case, we observe that inclusion of multiple stages does not alter the behaviour of the model solutions significantly (see Section 5).



(a) Plot after fitting with the Lotka-Volterra SEIR model. Here we choose 5 exposed and 5 infected stages with parameters $\beta_{high} = 0.3$, $\beta_{low} = 0.11$, $\beta(t_0) = 0.16$, $k_1 = 0.179$ and $k_2 = 75$. The estimated data starts from the second wave.



(b) Same plot when the model runs for a longer time. We observe from the model that the waves of infections increases, and eventually dies out. In general this is not a realistic scenario.

Figure 27: When the Lotka-Volterra SEIR model is used to fit only a part of the data set, it has to make sure that the behaviour of the model is also realistic for a longer period of time. We observe however that the behaviour is not realistic. Therefore we choose to fit the entire data set.

6.3.1 5 Exposed and 5 Infected stages - Lotka-Volterra

In Figure 28, we show the total number of infected people after fitting with 5 exposed and 5 infected stages. We observe that the model behaves well in the sense that after each consecutive peak, the epidemic slowly dies out. However, we observe also that the number of infected people in each peak of the model is higher than the estimated data set. Therefore, it seems that the inclusion of a multiple stages model only has a minor effect, because these results are similar to the results from Section 5.

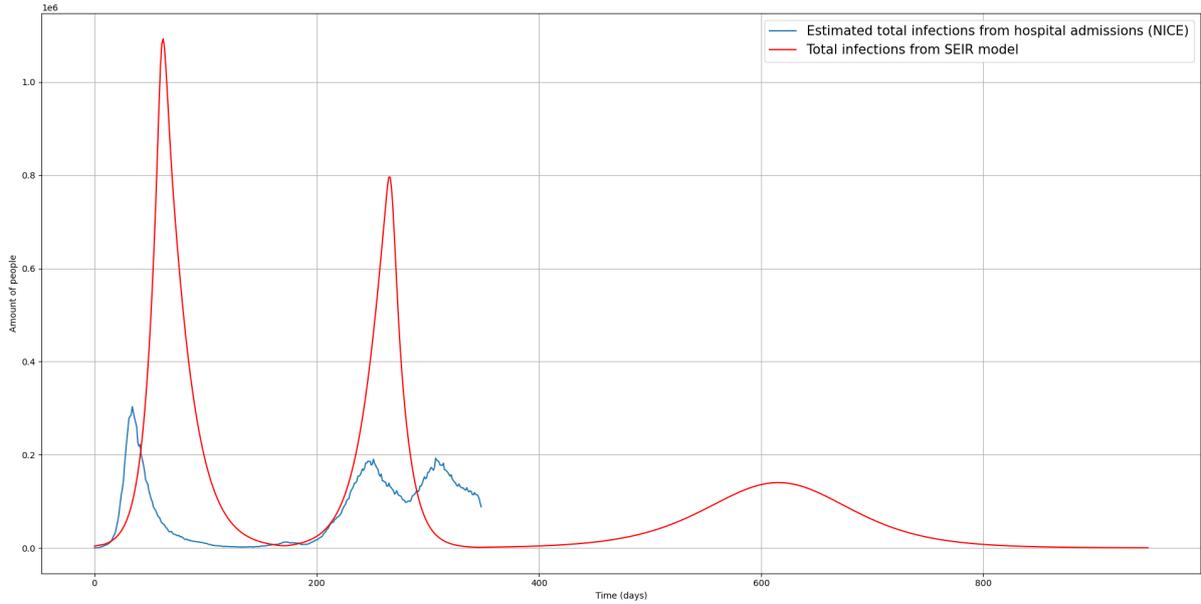


Figure 28: Result of the Lotka-Volterra SEIR model with 5 exposed and 5 infected stages. The parameters are $\beta_{high} = 0.31, \beta_{low} = 0.1, \beta(t_0) = 0.16, k_1 = 3.50$ and $k_2 = 375$. We observe that the number of infected people at each peak of the model is higher than the estimated data. We also observe that the peaks of the model decline, and eventually the epidemic dies out.

6.3.2 10 Exposed and 10 Infected stages - Lotka-Volterra

We see in Figure 29 the result of the Lotka-Volterra SEIR model with 10 exposed and 10 infected stages. We observe that the model solution behaves similarly to what we have seen in Figure 28. This implies that altering the number of exposed and/or infected stages does not have a significant effect on the model behaviour.

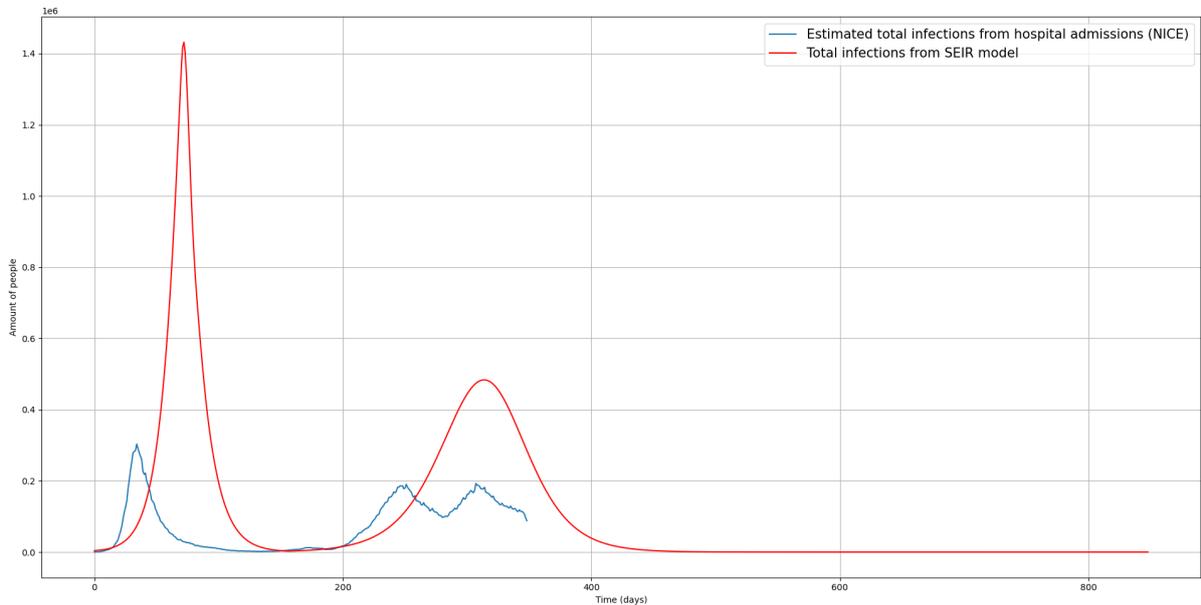


Figure 29: Lotka-Volterra SEIR model plot with parameters $\beta_{high} = 0.305, \beta_{low} = 0.097, \beta(t_0) = 0.16, k_1 = 4.608$ and $k_2 = 359.9$. The number of infected persons from the model is higher than the estimated data.

6.4 Conclusion & Discussion

In this section, we discussed why and how to implement gamma distributed stages in the standard SEIR model. After that, we showed multiple results of the adjusted SEIR model, in combination with use of our first ODE for $\beta(t)$ and the Lotka-Volterra. From these results, we made several observations. First, the differences in the results were minimal when the parameters were altered. Perhaps due to the small change in parameter values we did not observe major changes in the model behaviour. Though we cannot change the parameter values too much because they have to be in line with the COVID-19 statistics (such as β_{high} and $\beta(t_0)$ that are dependent on the reproduction number). Secondly, the inclusion of multiple exposed and/or infected stages only lead to minor differences with the preliminary results obtained in Section 4 and the Lotka-Volterra equations with standard SEIR model in Section 5. However, as we have mentioned at the beginning of this section, the assumption of an exponential distribution for the time period is unrealistic. Regardless of the minor differences in the results, we think that the assumption for an exponential distribution should not be made, and therefore other distributions should be considered (such as a gamma or normal distribution). Lastly, varying the number of stages seemed to have a minor effect as well. We chose either 5 exposed and 5 infected stages or 10 exposed and 10 infected stages. The choice for the number of stages was arbitrary, and although not necessary, we do not know yet if a specific number of stages will have a biological meaning that is in line with COVID-19. Perhaps if we had increased the number of stages, we would have seen a larger effect of the multiple stages model. We recommend future research into this topic.

7 The SEIR model and spatial heterogeneity

In the previous sections, we did not include spatial heterogeneity in the model. In this section, we will discuss how to adjust the standard SEIR model such that spatial heterogeneity is taken into account. Without this adjustment, the standard SEIR model implicitly assumes that the spread of the virus is more or less homogeneous in the Netherlands. However, we already mentioned in Section 2 that the virus did not spread homogeneously in the Netherlands. We expect that the virus spreads faster in densely populated areas, such as in cities and other urban areas, whereas in more rural areas, the spread is much slower. Therefore, it makes sense to include spatial heterogeneity in the model to take these observations into account.

Before we discuss the implementation of spatial heterogeneity, we need to have a look at the RIVM data again. The data can be grouped by the so-called 25 security regions (veiligheidsregio's) of the Netherlands. A map¹⁴ of all the security regions is shown in Figure 30.



Figure 30: The 25 security regions (Veiligheidsregio's) of the Netherlands.

Now the idea is to subdivide the Netherlands into different regions such that we are able to model the spread of the virus in and between these regions. A good way to take spatial heterogeneity into account

¹⁴<https://www.rijksoverheid.nl/onderwerpen/veiligheidsregios-en-crisisbeheersing/veiligheidsregios>

would be subdivision of the Netherlands in these 25 security regions. However, in order to make the model more tractable, we divide the Netherlands into 5 different regions: Noord (North), Oost (East), Zuid (South), West (West) and Zuidwest (South-west). These 5 regions consist of the following security regions:

| Region | Security Region | Population (millions) | Density (people/km ²) |
|-----------|------------------------------|-----------------------|-----------------------------------|
| North | 1,2,3 | 1.719 | 628 |
| East | 4,5,6,7,8,25 | 3.569 | 2356 |
| West | 9,10,11,12,13,14,15,16,17,18 | 7.508 | 14136 |
| South | 20,21,22,23,24 | 3.606 | 2804 |
| Southwest | 19 | 0.381 | 213 |

Table 3: Information about the 5 regions.

Region West contains most big cities and is typically an urban region, region North is a more rural region, and regions East and South contain numerous smaller cities and are a mixed- urban/rural region. Finally, region Southwest only contains Zeeland which has a small/negligible population.

In Figure 31, the number of hospital admissions for each region is shown. The data starts at February 27th 2020 and stops at February 3rd 2021. We observe that most admissions came from regions South (green) and West (red) during the first wave of infections, while the regions North (blue) and Southwest (purple) have the least number of admissions. During the second wave, the majority of the admissions came from region West.

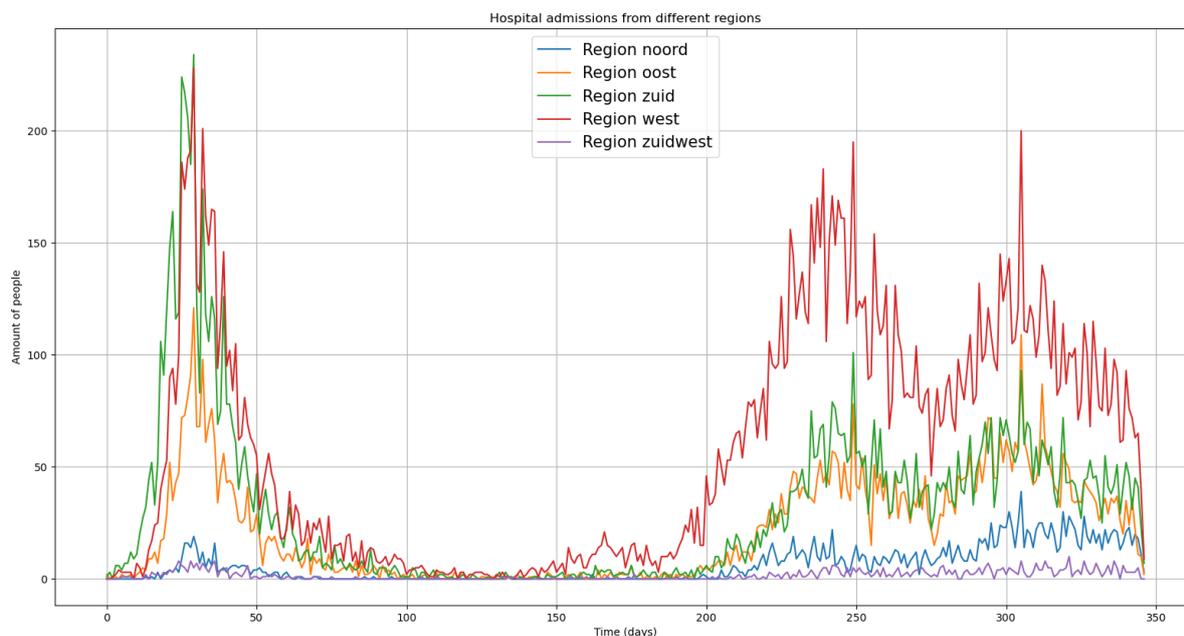


Figure 31: Hospital admission from different regions per day. The data starts from February 27th 2020 to February 3rd 2021. We observe a clear difference in the number of hospital admissions per region.

In Section 3, we estimated the total number of infected people in the Netherlands based on the hospital admissions data. The same methods can be applied to the different regions to estimate the total number of infected people per region per day. The estimation plots are shown in Figure 32. During the first wave, we see that the estimated total infections is fairly high in region West and South. During the second wave, region West accounts for the majority of the infections. Region West is a densely populated area where the largest cities of the Netherlands are there as well. So this should not come as a surprise. However, region South is less densely populated than the West, but it still has a high number of hospital admissions and total infections.

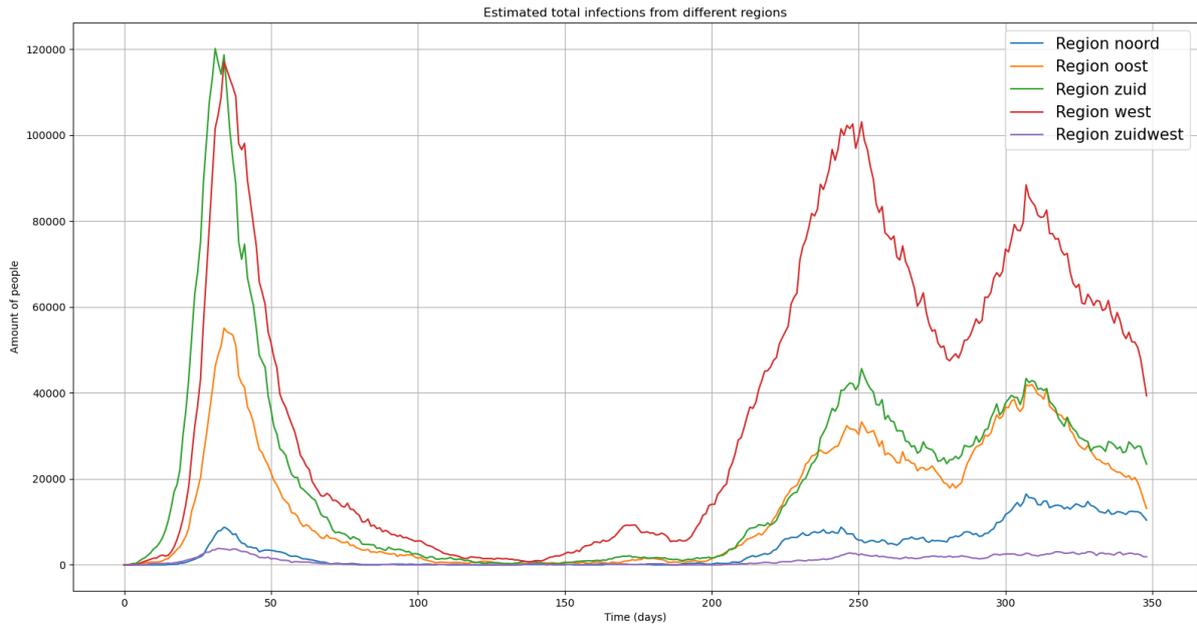


Figure 32: Estimated total infected persons per day for the different regions. The estimation methods are described in Section 3.

In Figure 33, we divided the estimated total number of infections of each region by the population number of that same region. With this, instead of the number of infected people, we see the fraction of the population number per region that was infected. We observe that region South has in proportion more infected people during the first wave. This is in line with what we mentioned in Section 2, i.e., the virus started in Noord-Brabant (which is in region South) and quickly spread through the rest of the Netherlands.

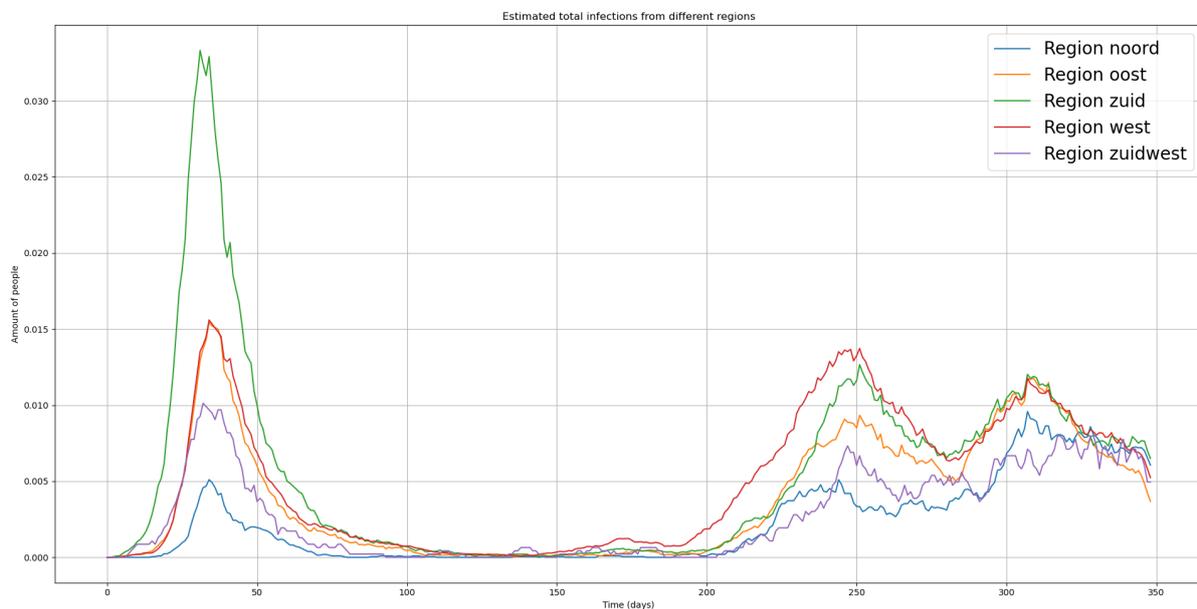


Figure 33: Estimated total infected persons per day in fractions for the different regions. Here we divided the absolute number of infections per region by the population of that region.

We note that it is not known from the data how many people are transported from one region to the other due to lack of Intensive Care Units (ICU's) in the hospitals. Therefore, it may be possible that the hospital admissions per region actually came from other regions.

Next, we will describe the implementation of spatial heterogeneity in the SEIR model in Section 7.1. After that, we will show a few numerical solutions in Section 7.2.

7.1 Model equations for the subregions

The standard SEIR model we used earlier can still be applied to the setting with several subregions. In general, the total population of each subregion is subdivided into four groups:

- Susceptibles S_m ,
- Exposed E_m ,
- Infected I_m ,
- Recovered /deceased R_m ,

where $m = 1, 2, \dots, N_r$, and N_r is the number of subregions.

After that, the standard SEIR model can be generalised as follows:

$$\begin{aligned}
\frac{dS_m}{dt} &= - \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}}, \\
\frac{dE_m}{dt} &= \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}} - \alpha E_m, \\
\frac{dI_m}{dt} &= \alpha E_m - \gamma I_m, \\
\frac{dR_m}{dt} &= \gamma I_m,
\end{aligned} \tag{106}$$

where the transmission coefficient $\beta_{m,n}(t)$ (off-diagonal) represents the transmission between an infected from region n to a susceptible of region m . If we have $\beta_{m,m}(t)$ (diagonal), it represents the transmission within region m .

We expect that the diagonal coefficients are larger than the off-diagonal coefficients, because the spread of the virus is much faster within a region than between different regions. Places such as school/work/supermarket/meeting with friends are mostly within a region, whereas for example visiting family is more likely to be in a different region. It is therefore important to model both the diagonal and off-diagonal coefficients differently.

7.1.1 Modelling intra-region transmission

The transmission within a region can be modelled as a chemical reaction using the law of mass action, where population densities play the role of the chemical species. We translate this as follows:

$$\frac{ds_m}{dt \text{ intra}} = -k_m s_m i_m, \tag{107}$$

where we define

$$s_m = \frac{S_m}{A_m}, i_m = \frac{I_m}{A_m} \quad [\text{number of people/km}^2], \tag{108}$$

and A_m is the area of region V_m in square kilometers. Similarly, we define

$$\rho_m = \frac{P_m}{A_m} \quad [\text{number of people/km}^2] \tag{109}$$

for the population density of region V_m . With these definitions, we can rewrite equation 107 as follows:

$$\begin{aligned}
\frac{ds_m}{dt \text{ intra}} &= \frac{d}{dt} \frac{S_m}{A_m \text{ intra}} = -k_m \frac{S_m}{A_m} \frac{I_m}{A_m}, \\
\Rightarrow \frac{dS_m}{dt \text{ intra}} &= -k_m \frac{P_m}{A_m} \frac{S_m I_m}{P_m} = -k_m \rho_m \frac{S_m I_m}{P_m},
\end{aligned} \tag{110}$$

where it follows that $\rho_m k_m = \beta_{m,m}$.

7.1.2 Modelling inter-region transmission

We will model the inter-region transmission of the virus as in Section 2:

$$\frac{dS_m}{dt}_{inter} = - \sum_{n=1, n \neq m}^{N_r} \beta_{m,n} \frac{S_m I_n}{P_n}, \quad (111)$$

where $\beta_{m,n}$ is the number of contacts of an individual living in region V_m with individuals in region V_n per day times the chance that such a contact leads to an infection.

7.1.3 ODE for $\beta_{m,n}(t)$

Just as described in Section 2 about the ODE for $\beta(t)$, we can model the temporal behaviour of the $\beta_{m,n}(t)$ in a similar manner. Here again, we have to consider both intra and inter-region spreading. For intra-region spreading, we have that $\beta_{m,m} = \rho_m k_m$. It follows that

$$\begin{aligned} \frac{d\beta_{m,m}(t)}{dt} &= k_+^{m,m}(\beta_{m,m;high} - \beta_{m,m}) - k_-^{m,m} I_m \beta_{m,m}, \\ \Rightarrow \rho_m \frac{dk_m}{dt} &= \rho_m k_+^{m,m}(k_{m,high} - k_m) - \rho_m k_-^{m,m} I_m k_m, \\ \Rightarrow \frac{dk_m}{dt} &= k_+^{m,m}(k_{m,high} - k_m) - k_-^{m,m} I_m k_m. \end{aligned} \quad (112)$$

Furthermore, if we assume the same internal spreading rates constants for all regions $k_+^{m,m} = k_+^{int}$, $k_-^{m,m} = k_-^{int}$, we find that

$$\frac{dk_m}{dt} = k_+^{int}(k_{m,high} - k_m) - k_-^{int} I_m k_m, \quad (113)$$

where k_+^{int} and k_-^{int} are two unknown fitting parameters.

Similarly for inter-region spreading, and assuming $k_+^{m,n} = k_+^{ext}$, $k_-^{m,n} = k_-^{ext}$, we have the equations

$$\frac{d\beta_{m,n}}{dt} = k_+^{ext}(\beta_{m,n;high} - \beta_{m,n}) - k_-^{ext} I_n \beta_{m,n}. \quad (114)$$

7.1.4 Summary of the equations model

Summarizing the equations we discussed in the previous sections, we have for our subdivisions model the following equations:

$$\begin{aligned} \frac{dS_m}{dt} &= - \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}}, \\ \frac{dE_m}{dt} &= \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}} - \alpha E_m, \\ \frac{dI_m}{dt} &= \alpha E_m - \gamma I_m, \\ \frac{dR_m}{dt} &= \gamma I_m, \end{aligned} \quad (115)$$

where

$$\frac{dk_m}{dt} = k_+^{int}(k_{m,high} - k_m) - k_-^{int} I_m k_m, \quad (116)$$

for intra-region spreading and

$$\frac{d\beta_{m,n}}{dt} = k_+^{ext}(\beta_{m,n;high} - \beta_{m,n}) - k_-^{ext} I_n \beta_{m,n}. \quad (117)$$

for inter-region spreading.

7.2 Results of the spatial SEIR model

Now we want to fit our model to the data again. From now on, we will combine the regions Southwest and North, because they have the lowest population density of all other regions. So in total, we will work with $N_r = 4$ regions. The parameters we want to fit are k_+^{int} , k_-^{int} , k_+^{ext} and k_-^{ext} . Furthermore, we are going to fit our model to the second wave of infections.

7.2.1 Initial values and parameters

For the intra-region spreading, we have the initial values $k_{m,high}$. We recall from Section 4 that we started with $\beta_{high} = 0.31$ and $\beta(t_0) = 0.16$. In a similar way, we use the same value setting $\beta_{m,m;high} = \rho_m k_{m;high} = 0.31$, which implies that

$$k_{m;high} = 0.31/\rho_m$$

And $\beta_{m,m}(t_0) = \rho_m k_m(t_0) = 0.16$ implies that

$$k_m(t_0) = 0.16/\rho_m.$$

For inter-region spreading, we choose

$$\beta_{m,n;high} = 0.031, \quad \beta_{m,n}(t_0) = 0.016,$$

which is 10 times less than intra-region spreading. Other initial values are shown in Table 4:

| Region | Infected | Recovered (millions) |
|-------------------|----------|----------------------|
| North + Southwest | 82 | 0.035 |
| East | 656 | 0.173 |
| South | 492 | 0.375 |
| West | 2624 | 0.404 |

Table 4: Initial values per region starting from July 16th 2020.

7.2.2 Numerical solutions of the spatial SEIR model

In Figures 34, 35, 36, 37, we show the total number of infected people per region as a function of time after fitting, and Figure 38 shows us the combined result of all the regions. The estimated data of infected people per region is also shown for reference. We obtained the following values for the parameters:

$$\begin{aligned} k_+^{int} &= 29.98, \\ k_-^{int} &= 9.6 \cdot 10^{-4}, \\ k_+^{ext} &= 28.04, \\ k_-^{ext} &= 0.177. \end{aligned}$$

From these figures, we observe that the numerical solutions of the model are overestimated (Figures 34, 35), underestimated (Figure 37) or in between (Figure 36) compared to the estimated data. Moreover, if all the regions are combined (Figure 38), we only observe minor differences with the preliminary results from Section 4.3. We remark also that altering the parameter values leads to marginal differences as well, so we will not show those results either. Lastly, in Figure 39 we show the transmission coefficients per region as a function of time.

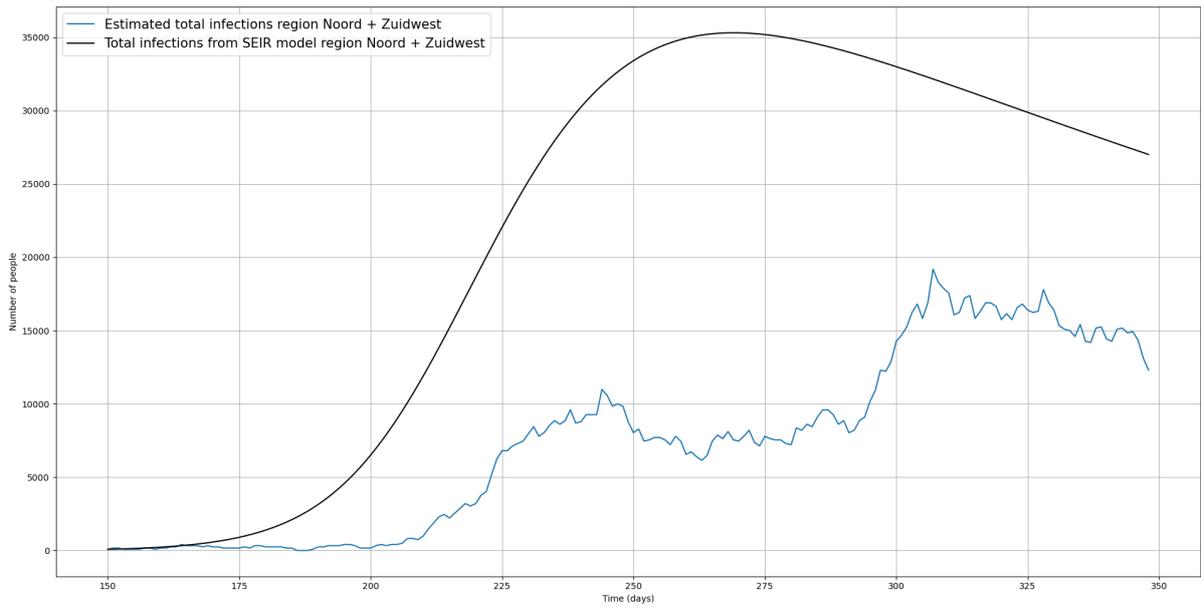


Figure 34: Result from the spatial SEIR model of the region North + Southwest and its estimated number of infected people for reference. Clearly, the model overestimates the total number of infected people in this region.

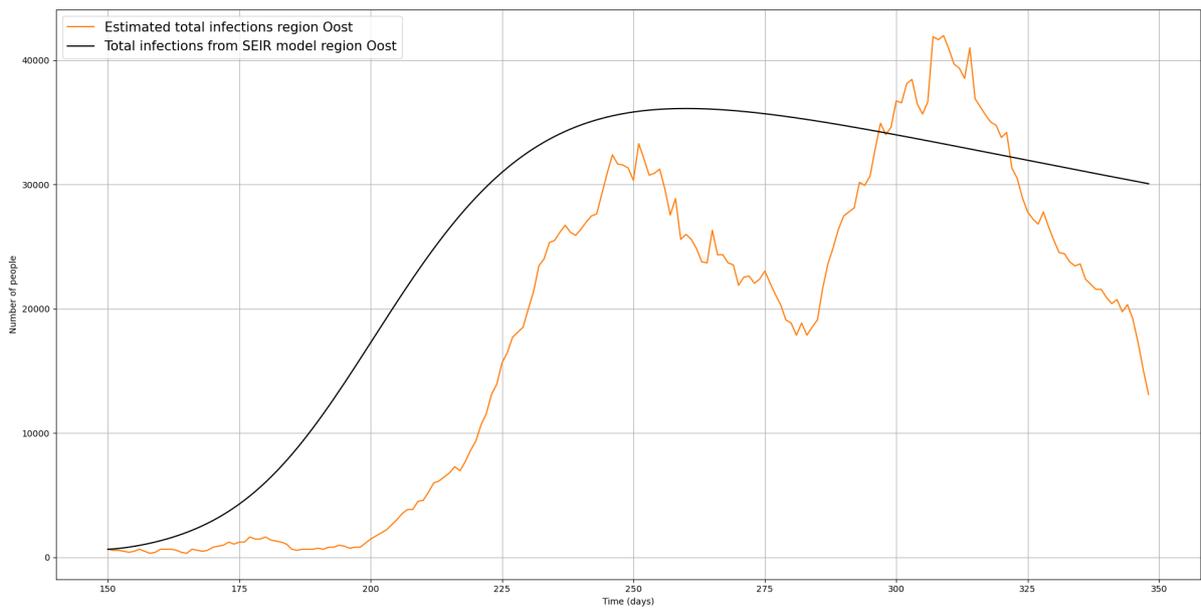


Figure 35: Result from the spatial SEIR model of the region East and its estimated number of infected people for reference. Here we observe also an overestimation of the spatial SEIR model.

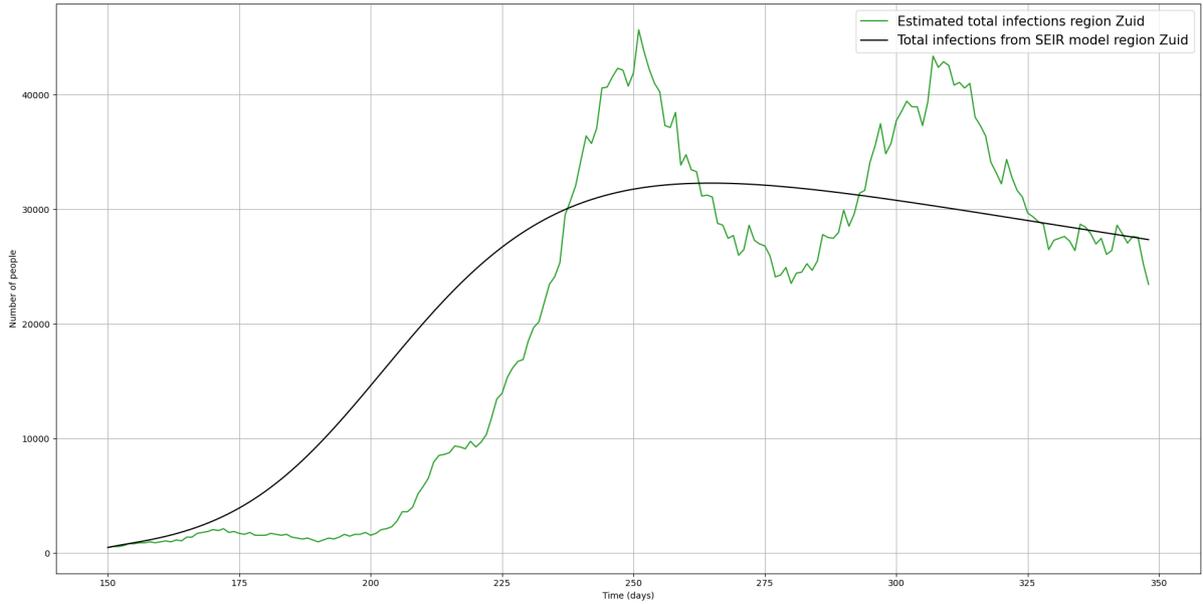


Figure 36: Result from the spatial SEIR model of the region South and its estimated number of infected people for reference. This time, the model output is not overestimated nor underestimated. However, the fitting is not too impressive either.

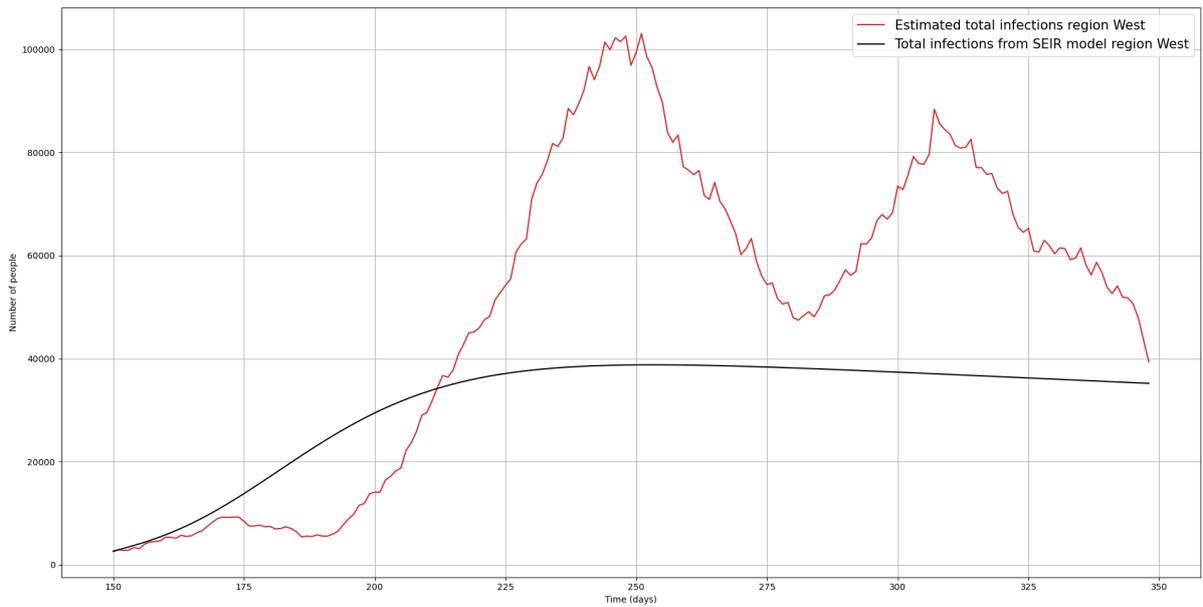


Figure 37: Result from the spatial SEIR model of the region West and its estimated number of infected people for reference. We observe that the model output is underestimated.

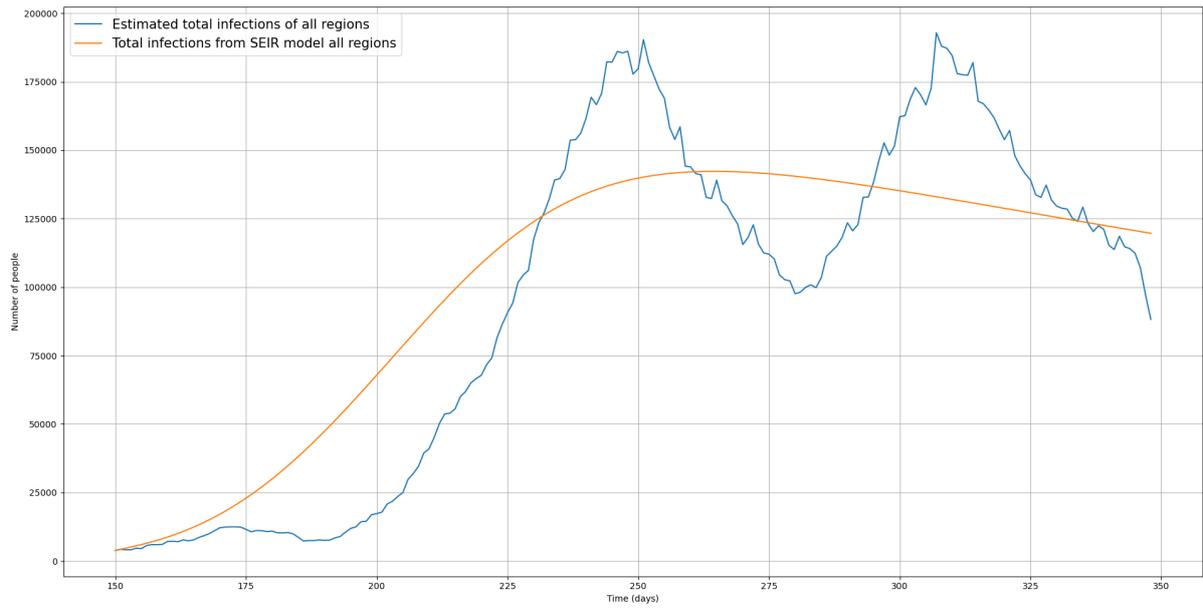
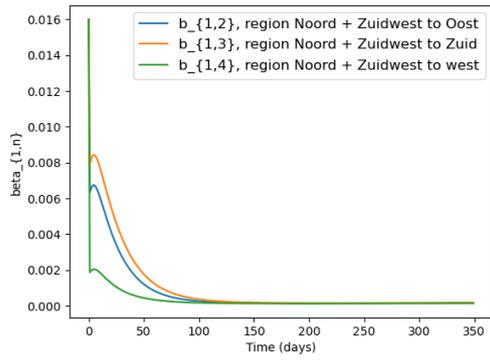
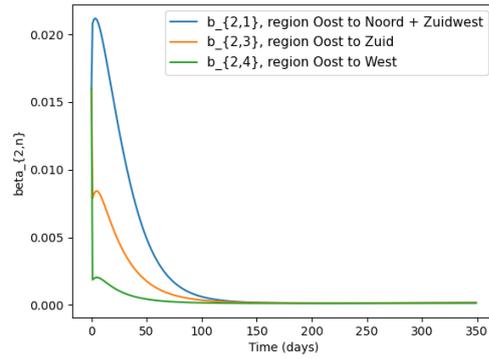


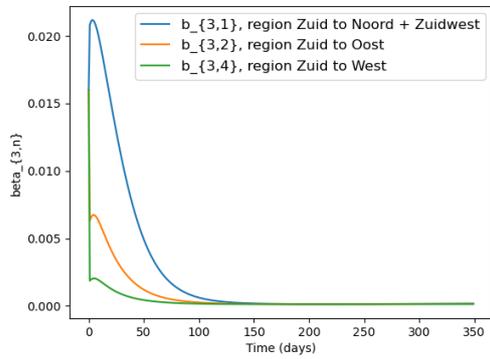
Figure 38: Result of the model when all the regions are combined.



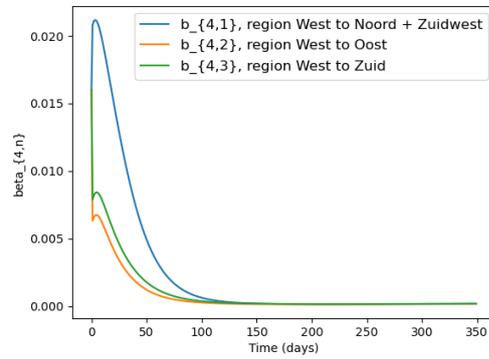
(a) Transmission coefficients of region North + Southwest.



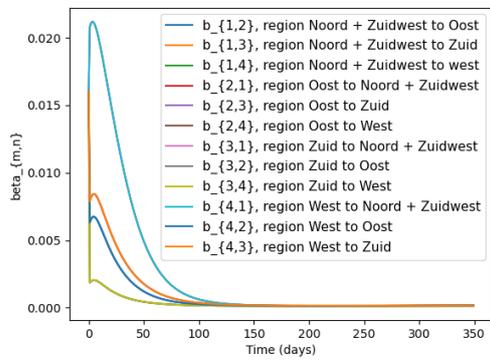
(b) Transmission coefficients of region East.



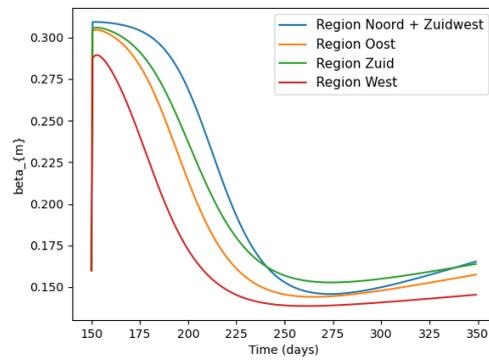
(c) Transmission coefficients of region South.



(d) Transmission coefficients of region West.



(e) Transmission coefficients of all regions in one figure.



(f) Plots of the the intra-region transmission coefficients $\beta_m(t)$.

Figure 39: Plots of all the inter-region transmission coefficients per region and the intra-region transmission coefficients.

7.2.3 Spatial SEIR model with Lotka-Volterra

We want to note that it is also possible to implement the spatial SEIR model inspired by the Lotka-Volterra equations. In a similar manner as explained in Section 5, we have the following set of DE's:

$$\begin{aligned}
\frac{dS_m}{dt} &= - \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}}, \\
\frac{dE_m}{dt} &= \sum_{n=1}^{N_r} \beta_{m,n}(t) \frac{S_m I_n}{P_{n,tot}} - \alpha E_m, \\
\frac{dI_m}{dt} &= \alpha E_m - \gamma I_m, \\
\frac{dR_m}{dt} &= \gamma I_m,
\end{aligned} \tag{118}$$

where

$$\frac{dk_m}{dt} = (k_m - k_{m,low}) \cdot (-c_+^{int}(k_m - k_{m,high}) + c_-^{int}(k_m - k_{m,high})I_m), \tag{119}$$

for intra-region spreading with $c_+^{int} = \rho_m \cdot k_+^{int}$ and $c_-^{int} = \rho_m \cdot k_-^{int}$.

Furthermore,

$$\frac{d\beta_{m,n}}{dt} = (\beta_{m,n} - \beta_{m,n;low}) \cdot (k_+^{ext}(\beta_{m,n} - \beta_{m,n;high}) + k_-^{ext}(\beta_{m,n} - \beta_{m,n;high})I_n). \tag{120}$$

for inter-region spreading. Unfortunately, we were not able to see fitting results. During the fitting, we either received numerical errors or the program was unable to find (optimal) fitting parameters within a certain amount of time, which also lead to errors. Consequently, we do not know the results and behaviour of the spatial SEIR model with Lotka-Volterra. Therefore, this model is open for future research.

7.3 Conclusion & Discussion

In this section, we discussed the idea and the implementation of spatial heterogeneity into the standard SEIR model. In the implementation, we made a distinction between inter-region and intra-region spreading of the virus. After that, we showed the numerical solutions of the model for each region. Unfortunately, we observed that the numerical results of the model did not lead to new insights. The results after fitting were either overestimated, underestimated, or in between; and the model could not capture the behaviour of the estimated data. For the spatial SEIR model with Lotka-Volterra, we do not know its results and behaviour either. We can think of a few reasons for this. First, the number of regions we had for our spatial SEIR model was $N_r = 4$. Perhaps if we had increased the number of regions (to the number of security regions), then the numerical solutions would have been more accurate. However, this means also an increase in parameters that need to be fitted. Due to time limitations, we were not able to consider more regions, so this problem is still open for further research. Secondly, the fitting routine we used is an in-built Python function (from a package). For simplicity and efficiency we only considered this function, but there are other fitting routines available. This may have affected the numerical solutions. Nevertheless, we think it is important to include spatial heterogeneity, because the spread of the virus is not likely to be homogeneous in a country.

8 Discussion

In this thesis, we focused on the first two steps to obtain an accurate and clear overview of the progress and behaviour of the virus in the Netherlands. These steps were: finding good estimates for the number of infected people as function of time and developing a model that reproduces these numbers (as reliably as possible). Throughout the thesis, the following major topics were discussed: in Section 3 we provided estimation methods to calculate the total number of infected people in the Netherlands. After that, we presented three modifications to the standard SEIR model in order to solve its (over)simplifications. These modifications were the modelling of $\beta(t)$ as a function of time using a linear ODE and the Lotka-Volterra equations in Sections 4 and 5, the use of gamma distributed stages for the SEIR model in Section 6 and the inclusion of spatial heterogeneity in Section 7.

In our estimations, we used both the daily hospital admissions data and the daily reported cases. We have seen that with hospital admissions, we obtained an estimation that was comparable to the estimations of the RIVM and Sanquin. For the reported cases however we observed that it underestimated the number of total infections, which was to be expected because not everyone tested themselves and initially not enough tests were available. In our method, we made a few assumptions to simplify the calculations. The most important assumptions were the infectious time period and the time between symptom onset and hospital admission. First we assumed a time period of 14 days between symptom onset and hospital admission, and 8 days for the infectious time. After the first results, we changed our assumption to 8 days for the time between symptom onset and hospital admission in order to better match the data. Although our estimation was comparable to the estimations of the RIVM and Sanquin, it would be interesting to see if we could improve our estimation using probability density functions ($\rho_H(a)$ and $\rho_R(a)$) instead of a point estimation (e.g. 8 days).

We presented two ways to model the infection rate $\beta(t)$ as a function of time. The first one is to model the infection rate using a relatively simple linear ODE, where two fitting parameters are involved. We have seen that this model (independent of the choice of the parameters) can only model a single wave, which means that this type of a model is unable to completely match with the historical data (second wave). Therefore, future predictions were not shown either. Nevertheless the numerical results from the model were still decent. Moreover, we observed in our results that the infection rate changed over time, which suggests the incorporation of an ODE for the infection rate $\beta(t)$ into a compartmental model. It would be interesting to see if we would obtain more accurate results when more fitting parameters are involved whilst retaining the linear structure.

However, in order to be able to model multiple consecutive waves, a different phase-space behaviour is required and non-linear terms need to be incorporated. We know that the Lotka-Volterra equations exhibit the desired behaviour, so we developed a system of ODEs inspired by the Lotka-Volterra equations. With these equations, the models (SIR and SEIR) did match with the historical data from a qualitative point of view, where in this case the historical data started from the beginning of the epidemic in the Netherlands. However, it overestimated the number of infected people. The overestimation was more noticeable when the SEIR model was used. The results suggest that we can model multiple waves of infections by the periodicity of the Lotka-Volterra equations. Increasing the accuracy of our Lotka-Volterra inspired model with for example the inclusion of seasonal events would be an interesting topic for future research.

In Section 6 we explained furthermore the idea and implementation of gamma distributed stages, where we assumed that the incubation time period and the infectious time period were gamma distributed. From the results, we observed that the overall behaviour of this model was comparable to the results without gamma distributed stages in Sections 4 and 5. Even with variation in the number of stages and parameters, it seemed to have minor effects on the behaviour of the model. One can ask then if it is necessary to use gamma distributed (or other non-exponential distributed) stages in this case. From a theoretical and biological point of view, we think that gamma distributed stages should be incorporated because the assumption of an exponential distribution is not in line with reality. From a practical point of view however, omitting the gamma distributed stages leads to less equations. The number of equations can become quite large if one wants to take spatial heterogeneity into account as well. This may also have an effect on the calculation speed when we want to obtain numerical results. Therefore, we recommend to explore more into this topic in the future.

Lastly, we explained how a compartmental model can take spatial heterogeneity into account in Section 7. Unfortunately, the numerical results were not in line with our expectations. The model was not able to match with the historical data accurately. We note that we only divided the Netherlands into four subregions based on the security regions, whereas there are 25 security regions in total. It would be interesting to see if the results will be affected when all 25 security regions are considered in the model. We will summarize our findings in the final section.

9 Conclusion & Outlook

In order to contain the virus and elevate the pressure on the healthcare system, good and effective measures are needed. These can only be achieved if one has an accurate and clear overview of the progress and behaviour of the virus and the expected impact of a proposed measure on this behaviour. We mentioned three steps in order to obtain this overview.

The first step is to acquire and combine data from hospital admissions and serological data such that we can estimate the total number of infected people. We can then compare the estimation with the number of positive cases to obtain a more reliable picture of the epidemic.

The second step is to have a good model that accurately describes the progress and behaviour of the epidemic. This means that the model is able to match with historical data and also predicts the future developments reasonably well.

The third step is to have an overview of the benefits and costs of each measure. In this way, we can decide whether the benefits of a particular measure outweigh the costs that come along. In this thesis, we focused on the first two steps.

We showed that it is possible to estimate the total number of infected people using hospital admissions data and serological data from Sanquin. We observed that our estimation were comparable to the estimations of the RIVM and Sanquin. Furthermore, we have also used the data from the reported cases per day to estimate the number of infected people. With this estimation and with enough test capacity, we found that on average only 39.3% of the total number of cases were detected. So the majority of the infections are not detected, which is due to not wanting to test, not having symptoms (and thus no reason to test) or other factors. Lastly, we found that on average 1.2% of the total number of infected people is admitted to the hospital and that 18.6% of the hospitalized patients is admitted to the ICU.

In the first two sections, we mentioned a few oversimplifications of compartmental models such as the standard SEIR model. These were for example the lack of spatial heterogeneity, only one wave of infections can be seen and the use of a constant value for the infection rate in some cases. In this thesis, we provided three modifications to the standard SEIR model in order to overcome these problems.

For our first modification, we first modelled the infection rate $\beta(t)$ as a function of time using a linear ODE and secondly we developed a system of ODEs inspired by the Lotka-Volterra equations.

With the linear ODE, we observed that the model was unable to match with the historical data because it cannot generate multiple consecutive waves. However, we did observe the change over time of the infection rate, which suggests the incorporation of an ODE for the infection rate. It remains to be investigated if we would obtain more accurate results when more fitting parameters are involved whilst retaining the linear structure.

We observed from our Lotka-Volterra inspired model that the model was able to match qualitatively with the historical data, because we are able to model multiple waves of infections over a longer period of time. However, the model overestimated the number of infected people. The inclusion of other effects like seasonal events into our model would be an interesting topic for future work to increase the accuracy of our model.

In a standard SEIR model, an exponential distribution is used to model the transitions from E to I and from I to R . From a biological viewpoint however, it seems more logical to use a different distribution like a gamma distribution instead. For the second modification, we implemented gamma distributed stages into the model for the incubation and infectious time period. We observed however only minor effects on the behaviour of the model compared to the results when no gamma distribution was assumed. Consequently, it remains to be seen whether or not the incorporation of gamma distributed stages is necessary. It is however important to check this when modelling different epidemics in future work.

In the Netherlands the epidemic started in the south of the Netherlands in the provinces Noord-Brabant and Limburg. So we expected that it might be important to include spatial heterogeneity into the SEIR model. The model was not able to completely match with historical data, however we had only divided the Netherlands into four subregions due to time limitations. For future work, we recommend to divide the Netherlands into 25 subregions based on the security regions to see if improvements are achieved. If not, it could also very well be that the Netherlands are homogeneous enough after all and that the

Netherlands can be treated as a single region.

In the case of COVID-19 it is observed that the disease is more serious amongst older groups. However we observed that in the different age groups (≥ 12) the number of infections is not very different. So to model the overall behaviour of the epidemic, it might be possible to neglect age groups. However if one wants to mitigate the effects of the epidemic, it is important to counter the spread amongst the older people. So for this purpose, it is significant to include age groups in order to assess which measures are suitable to protect the older people. To model the different age groups, one can make use of contact matrices where the contact rates between the different age groups are shown. However, using contact matrices brings many parameters to the table. Although having many parameters is not too difficult, it can however easily lead to notation errors and messy equations. So for future work, the inclusion of age groups is dependent on the main purpose the researchers have.

All in all, we want to emphasize to use our Lotka-Volterra model when future developments of the epidemic are of interest, especially for a longer period of time. Other topics described in this section can be used to optimize and fine-tune the model.

References

- [1] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong. Epidemic analysis of covid-19 in china by dynamical modeling, 2020.
- [2] Alberto Godio, Francesca Pace, and Andrea Vergnano. SEIR modeling of the italian epidemic of SARS-CoV-2 using computational swarm intelligence. *International Journal of Environmental Research and Public Health*, 17(10):3535, May 2020.
- [3] Nikhil Anand, A. Sabarinath, S. Geetha, and S. Somanath. Predicting the spread of COVID-19 using SIR model augmented to incorporate quarantine and testing. *Transactions of the Indian National Academy of Engineering*, 5(2):141–148, June 2020.
- [4] Santosh Ansumali, Shaurya Kaushal, Alope Kumar, Meher K. Prakash, and M. Vidyasagar. Modelling a pandemic with asymptomatic patients, impact of lockdown and herd immunity, with applications to SARS-CoV-2. *Annual Reviews in Control*, 50:432–447, 2020.
- [5] Qianying Lin, Shi Zhao, Daozhou Gao, Yijun Lou, Shu Yang, Salihu S. Musa, Maggie H. Wang, Yongli Cai, Weiming Wang, Lin Yang, and Daihai He. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in wuhan, china with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93:211–216, April 2020.
- [6] Shaurya Kaushal, Abhineet Singh Rajput, Soumyadeep Bhattacharya, M. Vidyasagar, Alope Kumar, Meher K. Prakash, and Santosh Ansumali. Estimating hidden asymptomatics, herd immunity threshold and lockdown effects using a covid-19 specific model, 2020.
- [7] A. L. Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, May 2001.
- [8] Alun L. Lloyd. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60(1):59–71, August 2001.
- [9] Zhilan Feng, Dashun Xu, and Haiyun Zhao. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bulletin of Mathematical Biology*, 69(5):1511–1536, January 2007.
- [10] Dankmar Böhning, Irene Rocchetti, Antonello Maruotti, and Heinz Holling. Estimating the undetected infections in the covid-19 outbreak by harnessing capture–recapture methods. *International Journal of Infectious Diseases*, 97:197–201, August 2020.
- [11] Xi He and Eric H. Y. Lau et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675, April 2020.
- [12] Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Shi Zhao, Maggie H. Wang, and Zuyao Yang. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *BMJ*, March 2020.
- [13] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, March 2020.
- [14] Ashleigh R. Tuite, David N. Fisman, and Amy L. Greer. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of ontario, canada. *Canadian Medical Association Journal*, 192(19):E497–E505, April 2020.
- [15] José Lourenço, Robert Paton, Craig Thompson, Paul Klenerman, and Sunetra Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the sars-cov-2 epidemic. *medRxiv*, 2020.
- [16] Hongjun Zhu, Yan Li, Xuelian Jin, Jiangping Huang, Xin Liu, Ying Qian, and Jindong Tan. Transmission dynamics and control methodology of COVID-19: A modeling study. *Applied Mathematical Modelling*, 89:1983–1998, January 2021.

- [17] Nicholas G. Davies, , Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M. Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, June 2020.
- [18] Nicholas G Davies et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, July 2020.
- [19] Elena Loli Piccolomini and Fabiana Zama. Monitoring italian COVID-19 spread by an adaptive SEIRD model. *medRxiv*, April 2020.
- [20] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, Kirsty O'Brien, Patrick Wall, Kieran A Walsh, and Simon J More. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*, 10(8):e039856, August 2020.
- [21] Anding Liu, Wenjie Wang, Xuecheng Zhao, Xiaoxi Zhou, Dongliang Yang, Mengji Lu, and Yongman Lv. Disappearance of antibodies to SARS-CoV-2 in a -COVID-19 patient after recovery. *Clinical Microbiology and Infection*, 26(12):1703–1705, December 2020.
- [22] N. Ahmad Aziz, Victor M. Corman, Antje K. C. Echterhoff, Marcel A. Müller, Anja Richter, Antonio Schmandke, Marie Luisa Schmidt, Thomas H. Schmidt, Folgerdiena M. de Vries, Christian Drosten, and Monique M. B. Breteler. Seroprevalence and correlates of SARS-CoV-2 neutralizing antibodies from a population-based study in bonn, germany. *Nature Communications*, 12(1), April 2021.
- [23] Christel Faes, Steven Abrams, Dominique Van Beckhoven, Geert Meyfroidt, Erika Vlieghe, and Niel Hens and. Time between symptom onset, hospitalisation and recovery or death: Statistical analysis of belgian COVID-19 patients. *International Journal of Environmental Research and Public Health*, 17(20):7560, October 2020.
- [24] Pierre-Yves Boëlle, Tristan Delory, Xavier Maynadier, Cécile Janssen, Renaud Piarroux, Marie Pichenot, Xavier Lemaire, Nicolas Baclet, Pierre Weyrich, Hugues Melliez, Agnès Meybeck, Jean-Philippe Lanoix, and Olivier Robineau. Trajectories of hospitalization in COVID-19 patients: An observational study in france. *Journal of Clinical Medicine*, 9(10):3148, September 2020.
- [25] Martin Braun. *Differential Equations and Their Applications*. Springer New York, 1993.
- [26] J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, November 2006.

A Reproduction Factor

A parameter that is often used to describe the progression of an epidemic is the so called reproduction factor $\mathcal{R}(t)$. This factor describes how many secondary infections are caused by an infected individual, i.e., an $\mathcal{R}(t) = 2$ implies that every 100 infected individual causes 200 new infections (and the epidemic spreads) whereas an $\mathcal{R}(t) = 0.5$ implies that every 100 infected individuals only cause 50 new infections (and the epidemic dies out). Note that the transition between spreading and dying out occurs at $\mathcal{R}(t) = 1$.

In the Netherlands, the RIVM calculated the reproduction factor using methods described by Wallinga (2006) [26]. At first sight, it may not be clear what the underlying assumptions are in this paper and how he derived certain formulas. So in Appendix B we determined how the reproduction factor is calculated from scratch such that we can compare our derivations with Wallinga.

A.1 Calculating the reproduction factor

Wallinga defines the following terms:

$r_c(t)$ = *per capita* change in number of new cases per unit of time,

$r(t)$ = change in number of new cases per unit of time,

with the relation

$$r_c(t) = \frac{r(t)}{I(t)}. \quad (121)$$

T_c = mean generation interval,

defined as the mean duration between time of infection of a secondary infectee and the time between time of infection of its primary infector.

b_1 = the rate of leaving the exposed stage

and

b_2 = the rate of leaving the infectious stage.

After that, we have the relation

$$T_c = 1/b_1 + 1/b_2. \quad (122)$$

Wallinga stated that epidemic models such as SIR and SEIR implicitly specifies a generation interval distribution. The distribution for the SIR model is an exponential distribution with mean $T_c = 1/b$, and for the SEIR model it is a convolution of two exponential distributions with mean $T_c = 1/b_1 + 1/b_2$.

Then, we can calculate the reproduction factor by the formula

$$\mathcal{R}_t = (1 + r_c(t)/b_1) \cdot (1 + r_c(t)/b_2), \quad \text{where } r_c(t) > \min(-b_1, -b_2). \quad (123)$$

In our case, we have that $b_1 = 1/5$, and $b_2 = 1/8$.

A.1.1 Data from SEIR model

First, we will use the SEIR model output from Section 4.3 for the total infected persons per day $I(t)$, which can be seen in Figure 15a and the recovered persons per day $R(t)$ to calculate $r(t)$.

Since we take the unit of time to be in days, $r(t)$ is simply looking at the difference (derivative) between the number of new cases for each consecutive day. And the number of new cases per day can be calculated from

$$I^{new}(t) = I(t+1) - I(t) + R(t+1) - R(t), \quad (124)$$

which we saw earlier. A plot of $r(t)$ can be seen in Figure 40.

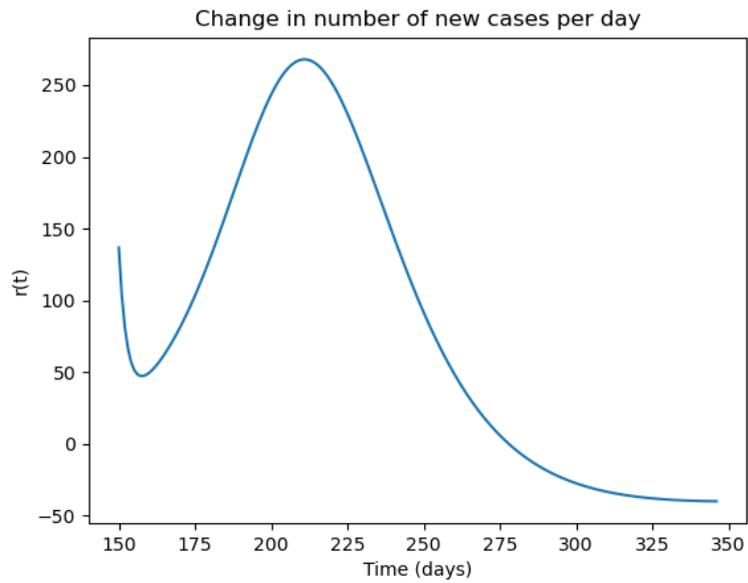


Figure 40: Change in the number of new cases per day.

Then, we can plot the graph of the reproduction factor \mathcal{R}_t in Figure 41.

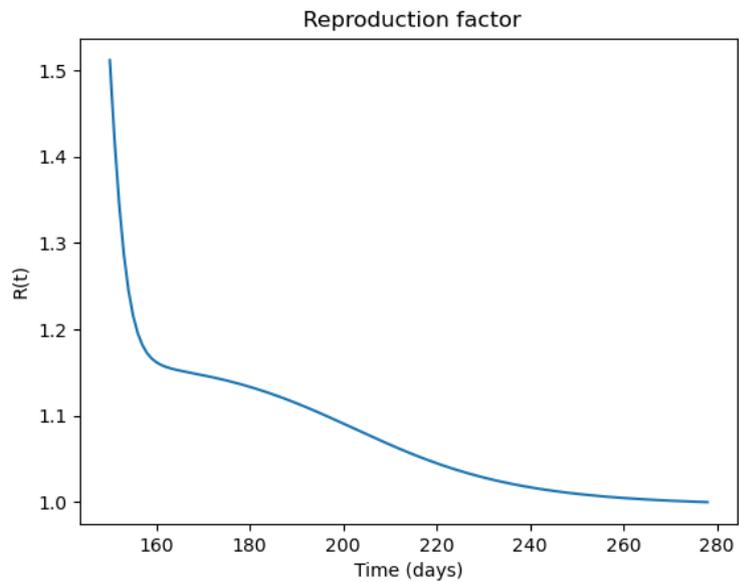


Figure 41: Plot of the reproduction factor \mathcal{R}_t

A.1.2 Data from estimated total infections

We will also calculate the reproduction factor using the data from our estimation in Figure 8. From equation (124), calculating the number of new cases per day can be done. This is shown in the next figure.

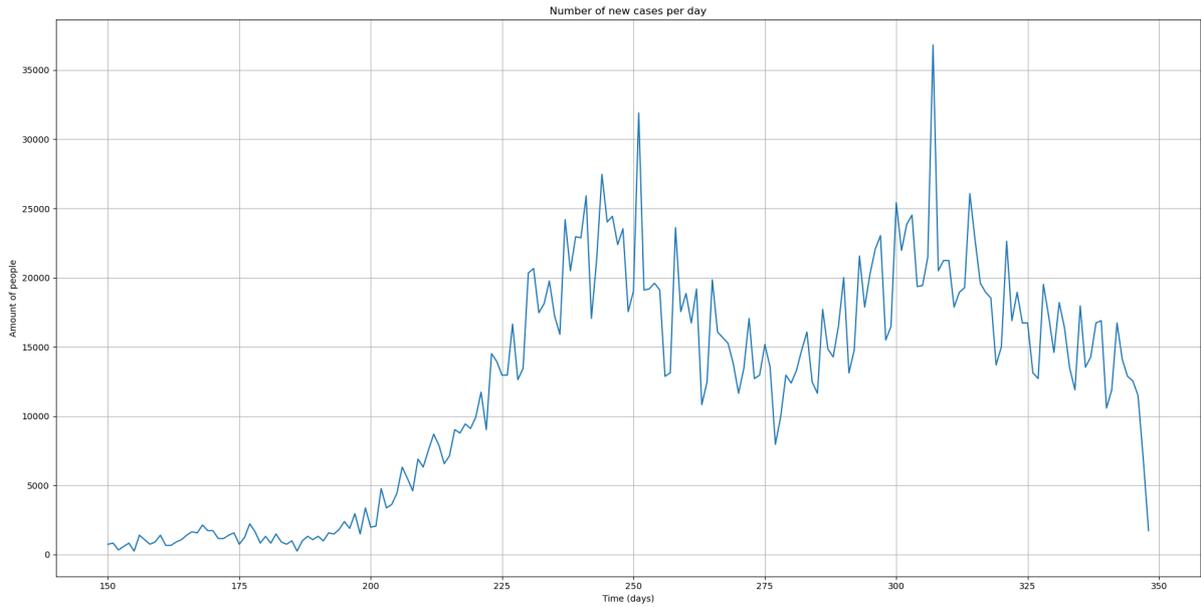


Figure 42: Number of new cases per day during the second wave using equation 124.

It is important to note that the number of new cases per day is not a smooth curve. This can cause some trouble when calculating the $r(t)$, which is in essence calculating the derivative of the number of new cases per day $I^{new}(t)$. Nevertheless, we can still calculate the $\mathcal{R}(t)$ from it, but it might not be a good representation. Therefore, it is better to smooth the curve first, which we will do so by calculating the moving mean (or rolling mean) with different window sizes. We chose 3, 5 and 10 days for our window sizes (arbitrary).

Then, the results of reproduction factor for these windows sizes and the non-smoothed curve are shown in Figures 43, 44, 45 and 46, with the estimation from the RIVM for comparison.

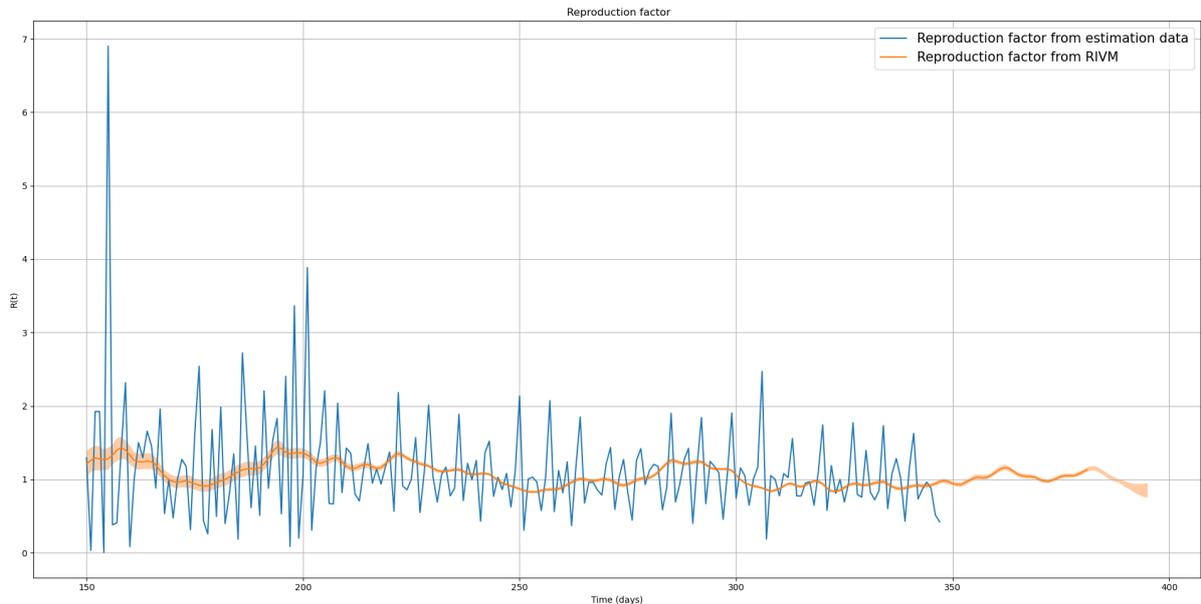


Figure 43: Plot of $\mathcal{R}(t)$ without smoothing.

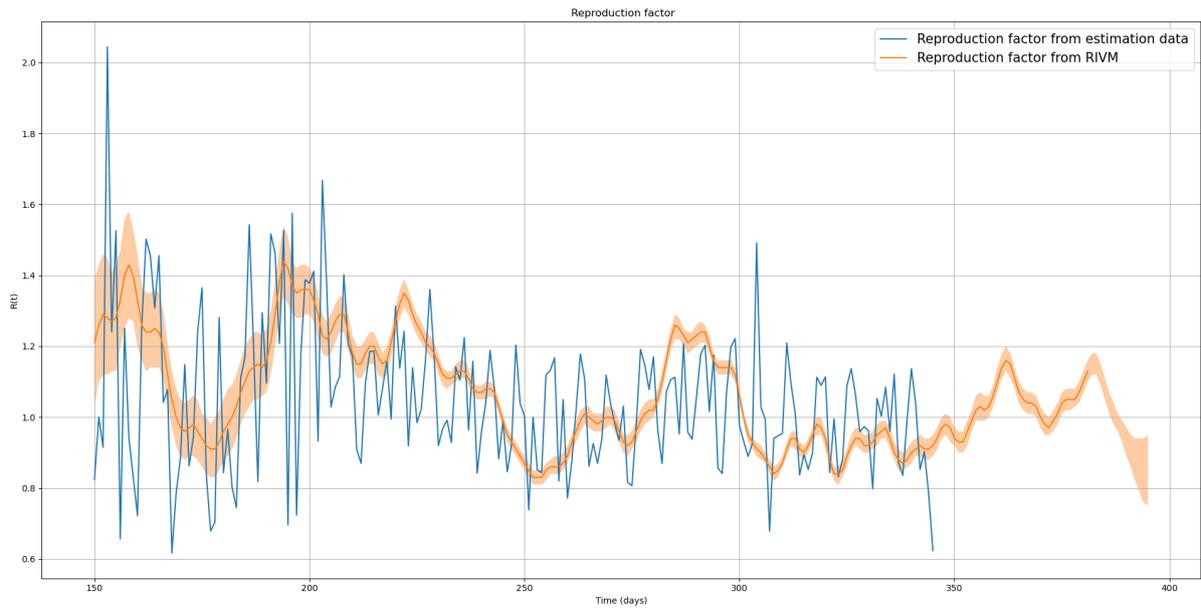


Figure 44: Plot of $\mathcal{R}(t)$ with window size of 3 days.

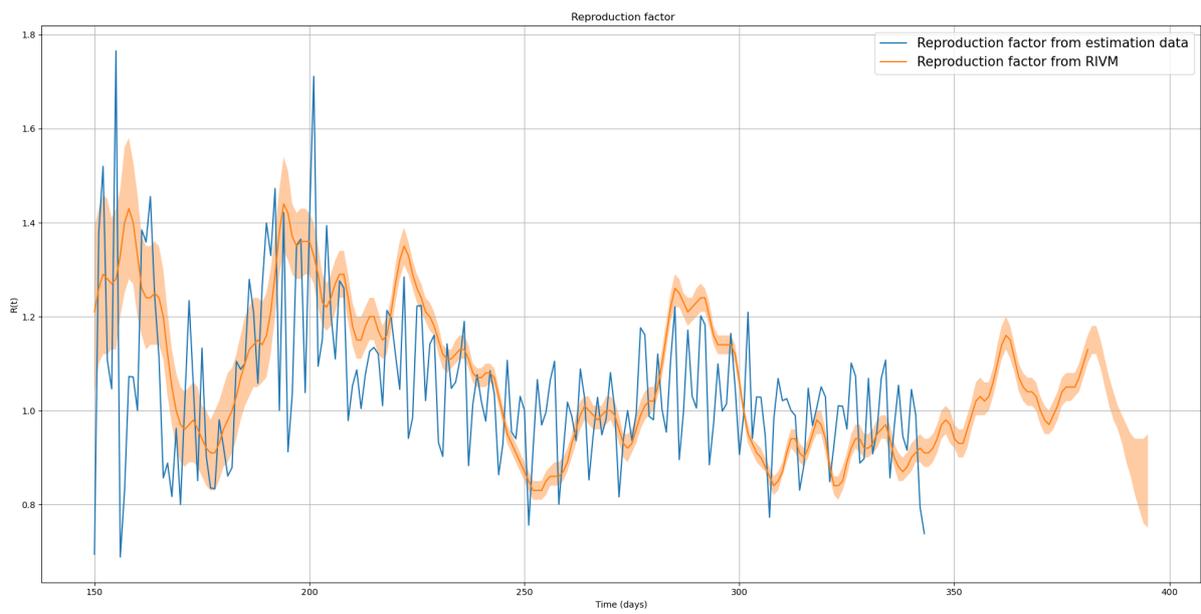


Figure 45: Plot of $\mathcal{R}(t)$ with window size of 5 days

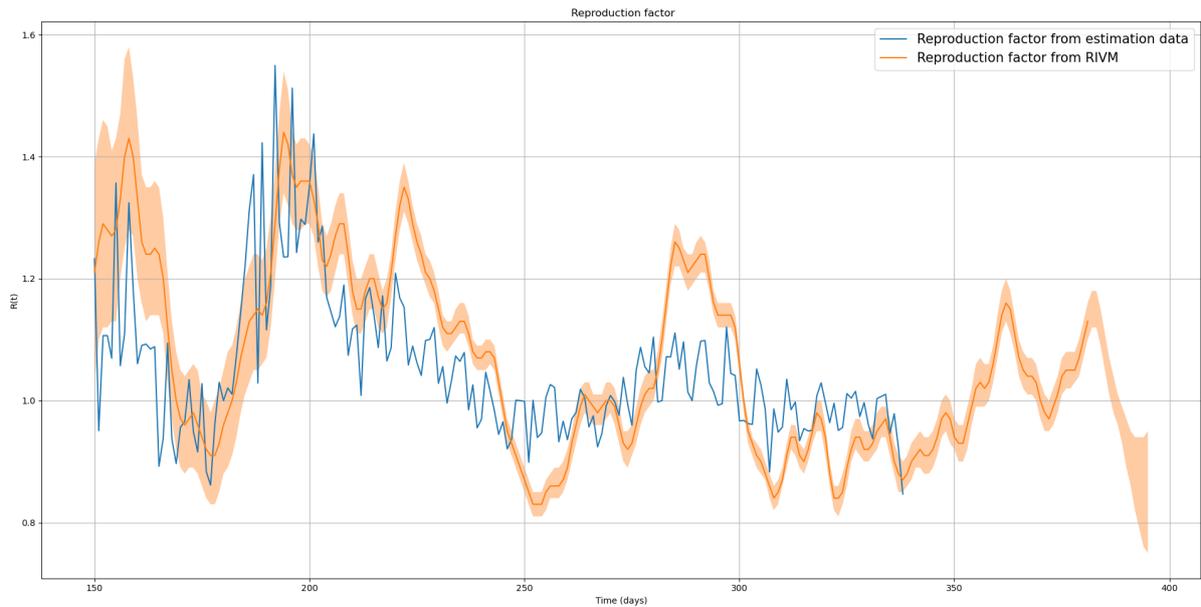


Figure 46: Plot of $\mathcal{R}(t)$ with window size of 10 days

It is also possible to smooth the data by fitting a polynomial curve. In Python, we can import the package *numpy.polynomial* to do this for us. We can choose the degree of the polynomial and in this case, we chose a polynomial of degree 8 to fit the polynomial to the data set. The result of the reproduction factor after the polynomial fit can be seen in Figure 47.

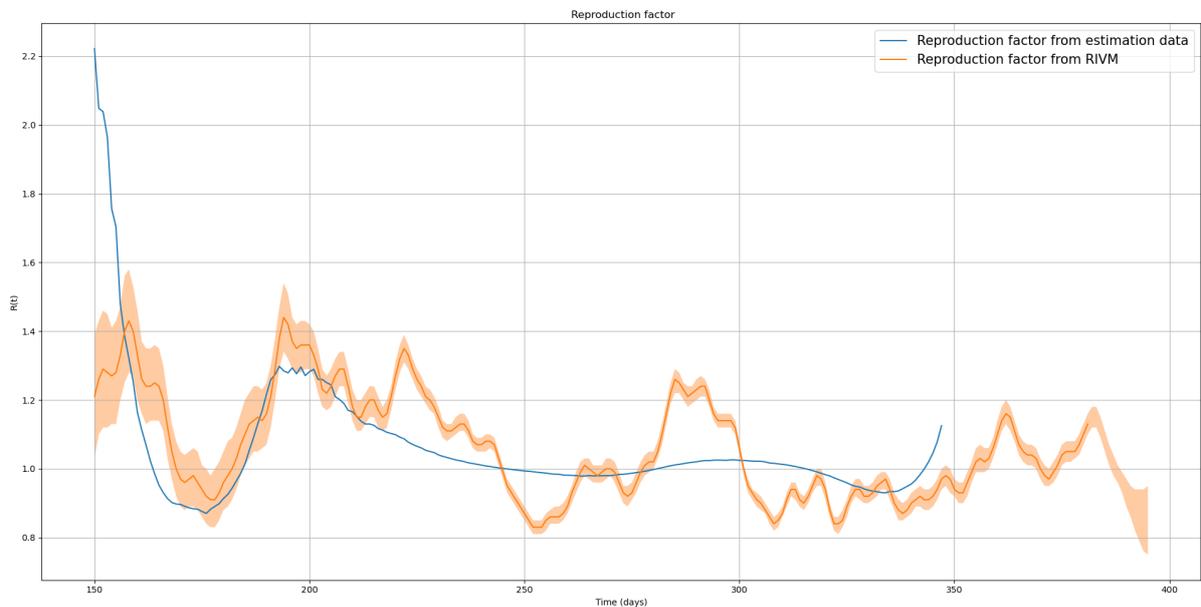


Figure 47: Plot of $\mathcal{R}(t)$ after smoothing by a polynomial of degree 8.

B Determining Reproduction Factor

In this section we will determine $\mathcal{R}(t)$ from a given function $I(t)$. This turns out to be surprisingly difficult and it even seems to result - for an ongoing epidemic - in an ill-posed problem. We will see that a number of assumptions have to be made to simplify the calculations.

We want to note that during the epidemic, the reproduction factor is commonly used to describe the progress of the virus. However, we think that calculating the reproduction factor is actually not necessary in order to have a good and accurate overview of the virus. So we think that the focus should be on the development of a good and accurate model.

B.1 Continuum model

An individual goes from state S after an infection to state E (exposed). In the stage E the individual is assumed to be unable to cause secondary infections. After incubation time the individual moves to stage I (infectious), in which the individual may cause secondary infections. After the infectious time the individual moves to stage R (recovered).

We will use the following basic definitions/notation:

$$E(t) = \text{total number of exposed individuals at time } t. \quad (125)$$

$$E^{new}(t) = \text{number of newly exposed individuals between } t \text{ and } t + \Delta t. \quad (126)$$

$$I(t) = \text{total number of infectious individuals at time } t. \quad (127)$$

$$I^{new}(t) = \text{number of newly infectious individuals between } t \text{ and } t + \Delta t. \quad (128)$$

$$R(t) = \text{total number of recovered individuals at time } t. \quad (129)$$

$$R^{new}(t) = \text{number of newly recovered individuals between } t \text{ and } t + \Delta t. \quad (130)$$

Similar definitions can be made for $S(t)$ and $S^{new}(t)$. These expressions however will not be used in the remainder of this section.

We can find an expression for $I^{new}(t)$ in terms of I and R as follows:

$$I(t + \Delta t) + R(t + \Delta t) = I(t) + R(t) + I^{new}(t) \Rightarrow I^{new}(t) = I(t + \Delta t) + R(t + \Delta t) - I(t) - R(t), \quad (131)$$

which means that we can obtain $I^{new}(t)$ from epidemic data on the total number of infected and recovered individuals.

We also have the probability density for the incubation times $\rho_{EI}(a)$, describing the probability of going from stage E to stage I , a days after the exposure took place. This yields an expression for the new infections $I^{new}(t)$ as follows:

$$I^{new}(t) = \int_{a=0}^{\infty} E^{new}(t)(t-a)\rho_{EI}(a)da. \quad (132)$$

Furthermore, we have the reproduction factor $\mathcal{R}(t)$ describing the number of secondary infections caused by an individual infected between t and $t + \Delta t$

$$\mathcal{R}(t) = \text{number of (new) infections caused per individual in } I^{new}(t). \quad (133)$$

Furthermore we have the probability density for the infectious period $\rho(a)$, describing the chance that a single secondary infection caused by an infected individual, takes place a days after the primary infection. Using these definitions we find the number of newly exposed individuals

$$E^{new}(\tau) = \int_{a'=0}^{\infty} I^{new}(\tau - a')\mathcal{R}(\tau - a')\rho(a')da' \quad (134)$$

and using equation (132) we obtain an integral equation for $\mathcal{R}(t)$

$$I^{new}(t) = \int_{a=0}^{\infty} \int_{a'=0}^{\infty} I^{new}(t-a-a')\mathcal{R}(t-a-a')\rho(a')\rho_{EI}(a)da'da. \quad (135)$$

Equation (135) can be simplified using the following transformation

$$u = a + a', v = a'. \quad (136)$$

Note that the Jacobian equals one, which means that we have (note the change in integral boundaries)

$$I^{new}(t) = \int_{u=0}^{\infty} \int_{v=0}^{\infty} I^{new}(t-u)\mathcal{R}(t-u)\rho(v)\rho_{EI}(u-v)dvdu. \quad (137)$$

Defining the convolution (which only depends on the known/assumed probability densities)

$$\tilde{\rho}(u) = \int_{v=0}^u \rho(v)\rho_{EI}(u-v)dv, \quad (138)$$

we finally find

$$I^{new}(t) = \int_{u=0}^{\infty} I^{new}(t-u)\mathcal{R}(t-u)\tilde{\rho}(u)du. \quad (139)$$

The problem of determining $\mathcal{R}(t)$ boils down to solving equation (139) using the observed infection number I^{new} and the probability densities (138).

B.2 Simplifying assumptions

Using equation (139) we can find an explicit expression for $\mathcal{R}(t)$, provided a few assumptions are made.

Assumption 1 The function $\tilde{\rho}(u)$ is the product of a polynomial and an exponential function:

$$\tilde{\rho}(u) = p_n e^{-\alpha u}, \quad (140)$$

where $p_n(u)$ is a polynomial of order n . Note that this assumption is satisfied if $\rho(a')$ and $\rho_{EI}(a)$ have an exponential/gamma distribution.

Assumption 2 The epidemic has been going on for a long time, i.e.,

$$t \gg T_{reg} = T_{inc} + T_{inf}, \quad (141)$$

where T_{inc} is the average incubation time and T_{inf} is the average time it takes to cause a secondary infection when an individual is infectious.

Note that Assumption 2 implies that

$$\tilde{\rho}(u) \approx 0 \text{ for } u > t, \quad (142)$$

because $\tilde{\rho}(u)$ expresses the chance of a secondary infection after u days. This chance is now (almost) zero, because t is much larger than the regeneration time. Note furthermore that we can approximate equation (139) by

$$I^{new}(t) = \int_{u=0}^t I^{new}(t-u)\mathcal{R}(t-u)\tilde{\rho}(u)du. \quad (143)$$

We assume that $t = 0$ is before the start of the epidemic, i.e., I^{new} and its derivatives are zero at $t = 0$.

Assumption 3A

$$I^{new}(0) = 0, \frac{dI^{new}}{dt}(0), \dots \quad (144)$$

Furthermore we assume that I^{new} satisfies a differential equation

Assumption 3B

$$\frac{dI^{new}}{dt} = r_t(t)I^{new}(t), t > 0, \quad (145)$$

i.e., we define the logarithmic derivative $r_t(t)$ of $I^{new}(t)$:

$$r_t(t) = \frac{d}{dt}(\ln(I^{new}(t))). \quad (146)$$

Note that Assumptions 3A and 3B combined that I^{new} is discontinuous at $t = 0$. Apart from that, this is more a definition of $r_t(t)$ and not really an assumption.

Finally we need to assume probability densities $\rho(a')$ and $\rho_{EI}(a)$ in order to find an explicit expression for $\mathcal{R}(t)$. We will take exponential distributions because

1. Exponential distributions yield the easiest expression for $\mathcal{R}(t)$.
2. The approach can easily be generalized to gamma distributions (only a bit more technical).
3. In the literature (e.g. Wallinga) this seems the preferred choice, so we can compare.

Assumption 4 Both $\rho_{EI}(a)$ and $\rho(a')$ satisfy an exponential distribution with parameters b_1 and b_2 respectively:

$$\rho_{EI}(a) = b_1 e^{-b_1 a}, \rho(a') = b_2 e^{-b_2 a'}. \quad (147)$$

Note that the choice of an exponential distributions is unphysical; for real application a generalization to a gamma distribution seems highly advisable.

B.3 Computations

First we use Assumption 4 to compute $\tilde{\rho}(u)$ as follows

$$\begin{aligned} \tilde{\rho}(u) &= \int_0^u b_1 e^{-b_1(u-v)} b_2 e^{-b_2 v} dv, \\ &= b_1 b_2 e^{-b_1 u} \int_0^u e^{(b_1-b_2)v} dv, \\ &= \frac{b_1 b_2}{b_1 - b_2} e^{-b_1 u} (-1 + e^{(b_1-b_2)u}), \\ &= \frac{b_1 b_2}{b_1 - b_2} (-e^{-b_1 u} + e^{-b_2 u}). \end{aligned} \quad (148)$$

We then note, as a consequence of Assumption 2, that we can rewrite equation (143) as a convolution

$$I^{new}(t) = \int_0^t f(t-u) \tilde{\rho}(u) du, \quad (149)$$

where we defined for convenience

$$f(\tau) = I^{new}(\tau) \mathcal{R}(\tau). \quad (150)$$

We take the Laplace transform on the LHS and the RHS of equation (149) and we use that the Laplace transform of a convolution is the product of the Laplace transforms:

$$\hat{I}^{new}(s) = \hat{f}(s) \hat{\rho}(s). \quad (151)$$

Computation of the Laplace transform of $\tilde{\rho}(u)$ is standard (due to Assumption 4, or more general, due to Assumption 1). Taking the Laplace transform of equation (148) we find

$$\hat{\rho}(s) = \frac{b_1 b_2}{b_1 - b_2} \left(\frac{-1}{s + b_1} + \frac{1}{s + b_2} \right) = \frac{b_1 b_2}{(s + b_1)(s + b_2)}. \quad (152)$$

Substitution of equation (152) in equation (151) and solving for $\hat{f}(s)$ yields

$$\hat{f}(s) = \frac{\hat{I}^{new}(s)(s + b_1)(s + b_2)}{b_1 b_2}. \quad (153)$$

We now assume Assumption 3A to ensure convergence of the inverse Laplace transform. Note that we have

$$\mathcal{L}\left\{\frac{dI^{new}(t)}{dt}\right\} = s\hat{I}^{new}(s) - I^{new}(0) = s\hat{I}^{new}(s) \quad (154)$$

and similarly

$$\mathcal{L}\left\{\frac{d^2 I^{new}(t)}{dt^2}\right\} = s^2 \hat{I}^{new}(s), \quad (155)$$

which means that we can rewrite the RHS of equation (153) as follows

$$\frac{\hat{I}^{new}(s)(s+b_1)(s+b_2)}{b_1 b_2} = \frac{1}{b_1 b_2} \left(\mathcal{L}\left\{\frac{d^2 I^{new}(t)}{dt^2}\right\} + (b_1 + b_2) \mathcal{L}\left\{\frac{dI^{new}(t)}{dt}\right\} + b_1 b_2 \mathcal{L}\{I^{new}(t)\} \right), \quad (156)$$

which means that we find

$$\begin{aligned} \mathcal{L}\{f(t)\} &= \mathcal{L}\{I^{new}(t)\mathcal{R}(t)\} \\ &= \frac{1}{b_1 b_2} \left(\mathcal{L}\left\{\frac{d^2 I^{new}(t)}{dt^2}\right\} + (b_1 + b_2) \mathcal{L}\left\{\frac{dI^{new}(t)}{dt}\right\} + b_1 b_2 \mathcal{L}\{I^{new}(t)\} \right), \end{aligned} \quad (157)$$

and taking the inverse Laplace transform on both sides and using linearity we find

$$I^{new}(t)\mathcal{R}(t) = \frac{1}{b_1 b_2} \left(\frac{d^2 I^{new}(t)}{dt^2} + (b_1 + b_2) \frac{dI^{new}(t)}{dt} + b_1 b_2 I^{new}(t) \right), \quad (158)$$

i.e.,

$$\mathcal{R}(t) = 1 + \frac{b_1 + b_2}{b_1 b_2} \frac{\frac{dI^{new}(t)}{dt}}{I^{new}(t)} + \frac{1}{b_1 b_2} \frac{\frac{d^2 I^{new}(t)}{dt^2}}{I^{new}(t)}. \quad (159)$$

We finally use Assumption 3B to simplify equation (159)

$$\frac{dI^{new}}{dt} = r(t)I^{new}(t) \Rightarrow \frac{d^2 I^{new}}{dt^2} = r'(t)I^{new}(t) + r(t) \frac{dI^{new}}{dt} = (r'(t) + r(t)^2)I^{new}(t), \quad (160)$$

and we find

$$\mathcal{R}(t) = 1 + \frac{b_1 + b_2}{b_1 b_2} r(t) + \frac{1}{b_1 b_2} (r'(t) + r(t)^2) = \left(1 + \frac{r(t)}{b_1}\right) \left(1 + \frac{r(t)}{b_2}\right) + \frac{r'(t)}{b_1 b_2}, \quad (161)$$

which is - apart from the last term - the expression of Wallinga.