Bachelor Thesis

Simulating COVID-19 contact tracing queue using compartment models

Student
Marleen Hielkema (4478460)

Supervisor
Prof.dr.ir. C. Vuik

July 6, 2021

Simulating COVID-19 contact tracing queue using compartment models

by

M. Hielkema

to obtain the degree of Bachelor of Science, Applied Mathematics

at the Delft University of Technology,

to be defended on Friday July 9, 2021 at 15:00.

Student number:	4478460	
Project duration:	April 19, 2021 – July 9, 2021	
Thesis committee:	Prof. dr. ir. C. Vuik,	TU Delft, supervisor
	Dr. ir. M. Keijzer,	TU Delft

TU Delft

Preface

This bachelor thesis is written as completion of the Bachelor Applied Mathematics at Delft University of Technology under supervision of Prof.dr.ir. C. Vuik on behalf of the department Numerical Analysis at EEMCS.

The application of Numerical Analysis in real life has interested me throughout the entirety of my bachelor, it always showed me how important mathematics are in every day life. This project is therefore a very pleasant ending of my studies, especially since I got the opportunity to apply this field to one of the most unusual appearances in my lifetime: a pandemic. It has been very interesting to investigate how some mathematical models can be used to forecast and describe patterns such that authorities can make better decisions.

To describe the content of this research briefly, the research is set up to improve the classical compartment models in such a way that it can simulate COVID-19. This improved model is then used to estimate the amount of contact tracers needed to keep the pandemic under control in a specific country.

I would like to thank my supervisor Prof.dr.ir. C. Vuik for his help and support throughout the past period of time. Our clear communication and efficiency made it very pleasant to work together, even though we haven't yet had the pleasure of meeting each other in real life. I enjoyed the freedom throughout my research and when I had a question, I could always count on a quick response.

At last I would like to thank Dr.ir. M. Keijzer for taking seat in my assessment committee.

M. Hielkema
Delft, July 2021

Abstract

This research is mainly about optimising compartment models for COVID-19 and then using them for different applications that can be used to consult authorities.

A compartment model describes the dynamics of a disease by implementing differential equations for the different states that belong to that disease. This is done using parameters which can be estimated from actual data or can be extracted from researches done by external authorities. We begin by extending the standard compartment model to a model which is applicable to COVID-19, this extension is based on characteristics of the virus. After this extension we are going to counteract on the assumption that susceptible and infected individuals are heterogeneously in contact with each other, since social distancing and quarantine prevent these two compartments from interacting. Also, we use specific time intervals and optimise the mortality rate.

After these improvements we create an external model which estimates a contact tracing queue such that authorities can forecast how many contact tracers are needed to keep the pandemic under control. Then using the implementation of this queue, another queue is made which is dependent on its own length to forecast what happens if the government is not in contact with individuals who possibly are infected with COVID-19. Then we extend the model in such a way that vaccinated individuals can be assigned to a specific compartment, and at last we create a queue as described above with this extended model.

Contents

1

Introduction

During the current pandemic, it has been important that the authorities of a country are in contact with civilians that are possibly infected by the virus. The main problem for the government is the availability of contact tracers and, in a later stadium of the virus, the vaccine availability. Especially in the beginning of the pandemic and during its peaks, it was crucial that individuals got a call from contact tracers for a few reasons. First of all, the contact tracers told the (possibly) infected individuals to stay at home and not make physical contact with others. Secondly, the contact tracers made sure they could contact the people that are possibly infected by the virus due to contact with this infected individual. At last, the contact tracers monitor the infected individuals to make sure they are well and to ask if they have not been outside that day.

A lot of contact tracers are needed to investigate a proper amount of people such that the pandemic is researched in a good way and the infected inhabitants are under sight of the government. But how many do we need in the near future? In this report we create a model to forecast the dynamics of COVID-19 by using compartment models which are optimised such that they fit to the characteristics of the virus and the behaviour of the inhabitants of a certain country. This model can in turn be used to make an estimation of the amount of contact tracers needed by creating hypothetical contact tracing queues. The model for the contact tracing queue is based on an article which is briefly explained in Chapter 2.

Since February, vaccination became more common, and from June on everyone could make an appointment if they wanted. This resulted in much less infected, exposed and deceased individuals each day. To continue the research, we created a new model in which the vaccination rate is implemented, such that the decrease in infected, exposed and deceased is taken into account when creating other hypothetical contact tracing queues.

To sum up this thesis the following research questions and sub-questions are stated.

1. How can the classical compartment model be adjusted such that it is applicable to COVID-19?
   (a) How can the characteristics of the virus be implemented?
   (b) How can the behaviour of the inhabitants (social distancing and quarantine) be implemented into the model?
2. How many contact tracers are needed in the future to keep the pandemic under control?
3. How many individuals have to be vaccinated to get the desired decrease of infected individuals?

Please note that throughout the course of my thesis more data became available. Until Chapter 8 data is used up until February 29, and from Chapter 9 on data until June 21 is used.

All data used in this report, unless mentioned otherwise, is from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and can be found at `https://github.com/CSSEGISandData/COVID-19`. This data is updated every day and made public for everyone to analyse and use for their research.

2

Research article

The article on which a big part of this thesis is based on, is "Modelling Resource Demands and Constraints for COVID-19 Intervention Strategies" by Bagal et al. [1]. This section contains a summary in own words of the article. Later in this report we will elaborate on the adjustments made to the methods used in this article that make it applicable to our research.

2.1. Introduction

The main goal of the article is to simulate a queue for people that very recently have been in contact with an individual who tested positive for COVID-19. When someone who has been in contact with a COVID-19 positive person and the contact tracing is done in time, the information gained about their behaviour and contacts is usable. When this is not done in time, you will not have representative information and are to late to keep control of their situation. Therefore, the queue mentioned earlier is important, since it indicates whether the effect of the interventions at the time are investigated properly or not.

The timeliness of contact tracing is constrained by the length of the infectious period, the turn-around time for testing and result reporting and the ability to successfully reach and interview patients and their contacts.

2.2. Contact tracing queue

The contact tracing queue consists of the following two elements.

1. The flow in, defined as the new cases that need to be investigated
2. The flow out, defined as the new cases that are investigated

Both functions are time dependent and form the queue as follows

$$\frac{dC}{dt} = [\text{flow}_{in}] - [\text{flow}_{out}] \tag{2.1}$$

We define the function $C(t)$ as the new cases that are in need of investigation, this implies that $\frac{dC}{dt}$ defines change in new cases which need to be investigated and can be seen as the contact tracing queue.

Both the $[\text{flow}_{in}]$ and the $[\text{flow}_{out}]$ will be approximated using a compartment model combined with mathematical functions, these approximations will be elaborated on in this chapter.

2.3. Compartment model

The compartment model we are using is an extension of the SEIR-model, more compartments are added to make it applicable for the research in this chapter. A visualisation can be found in Figure 2.1 and the corresponding compartments are defined in Table 2.1.

Another important extension made is the quarantined compartments which are noted with subscript one. Two assumptions are made regarding quarantined individuals: They are assumed not to have contact with anyone, so they will not cause more contact tracing work, and when someone is tested positive they will go into quarantine immediately.

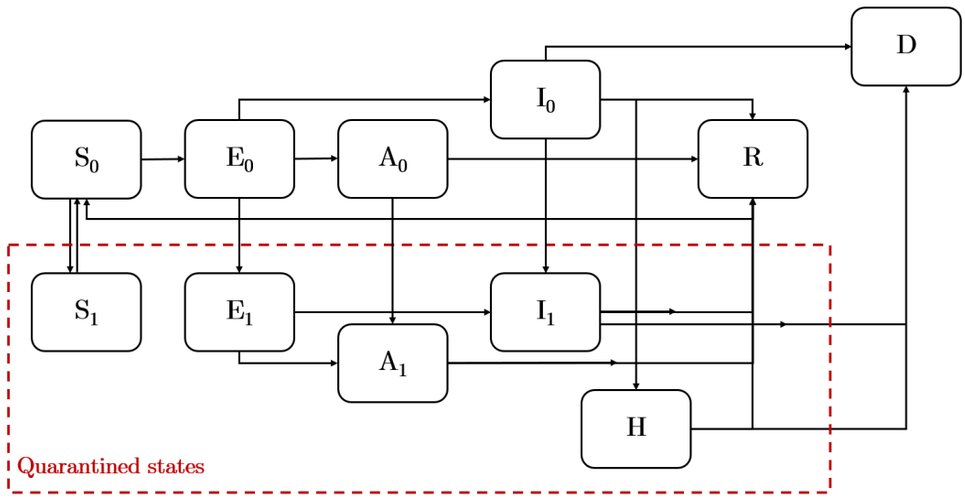


Figure 2.1: Extended SEIR model visualisation

	Meaning	Explanation
S	Susceptible	Can get ill
E	Exposed	Is infected but not infectious yet
A	Asymptomatic	Is infectious but does not show symptoms of the disease
I	Infective	Is infectious including symptoms of the disease
H	Hospitalised	Is infectious and hospitalised due to heavy symptoms
R	Recovered	Is not ill or infectious, immune
D	Deceased	Died of the disease

Table 2.1: SEIR-model compartments explained

2.4. Flow in

The flow in is retrieved by adding all positive tested COVID-19 cases, then multiplying this number by the average number of contacts. There are three ways someone finds out that they are infected by COVID-19.

The first way is when someone is testing randomly. When people do not feel well or they want to visit their family and want to test out of precaution are 'random testers'. The $q_{r\cdot}$-functions are time dependent rates of random testing for both the asymptomatic and infective people. They can differ since the chance of someone testing while having symptoms is bigger than when they do not. People who are testing for COVID-19 and belong to the compartment asymptomatic not in isolation (A_0) or infective not in isolation (I_0) can find out they are positive and will be assigned to the respective compartment in isolation as stated in the transmission equations below.

$$q_{rA}(t)A_0(t) \rightarrow A_1(t)$$
$$q_{rI}(t)I_0(t) \rightarrow I_1(t)$$

The second way someone finds out they are positive is when they will test because they came into contact with someone who was infective at that time. These people belong to the same compartments as before, with the exposed in isolation (E_1) compartment as an addition. The $q_{t\cdot}$-functions are time dependent rates of people who go testing triggered by contact. Again, transmission equations are set up to represent this.

$$q_{tA}(t)A_0(t) \rightarrow A_1(t)$$
$$q_{tI}(t)I_0(t) \rightarrow I_1(t)$$
$$q_{tE}(t)E_1(t) \rightarrow \{A_1(t), I_1(t)\}$$

The final way someone finds out the are positive is a small group of people who were missed by the non-pharmaceutical interventions and require hospitalisation.

$$\tau_{IH}(t)I_0(t) \rightarrow H(t)$$

Here, τ_{IH} is the inverse of the expected amount of time for which an infected individual is symptomatic before hospitalisation. The reason we look at the inverse is obvious when you look at it in another way.

$$I_0(t) \rightarrow H(t)\tau_{IH}^{-1}(t)$$
$$I_0(t) \rightarrow H(t) * (\text{symptomatic time before } H(t))$$

We use that the amount of individuals in compartment $I_0(t)$ that are eventually being admitted to the hospital, is approximately equal to the amount of people in $H(t)$ multiplied by the expected amount of days they have symptoms before being hospitalised.

When adding all these compartments on the left side of the arrows, we have estimated the amount of people that tested positive on time t. With this information we can calculate the amount of people that need to be called by the contact tracers. This can be calculated by multiplying all these people by the average number of contacts per case. This number is represented as $K(\kappa, T_S, \phi_k)$ and depends on the average number of contacts a day (κ), the average number of days for which an individual is infectious before going into isolation (T_S), and the likelihood that the individual will recall his/her contacts (ϕ_k).

This results in the following expression, it models the rate of increase for the contact tracers' backlog at day t.

$$[\text{flow}_{in}] := K(\kappa, T_S, \phi_k)\big[q_{tE}(t)E_1(t) + \big(q_{rA}(t) + q_{tA}(t)\big)A_0(t) + \big(q_{rI}(t) + q_{tI}(t) + \tau_{IH}(t)\big)I_0(t)\big]$$

2.5. Flow out

The flow out is the work that is executed on contact tracing for the newly arrived individuals in the queue. It comes down to a formulation that multiplies the portion of new cases of all the cases by time that is available for these new cases.

First of all, we estimate the total work required to monitor known cases and to investigate new cases. We do this by defining variables for the time that it costs to monitor a contact each day (w_m) and time necessary to investigate new contacts each day (w_c). Again, we use the compartment model to represent this.

$$\text{Total work required to monitor known cases} = w_m(A_1(t) + I_1(t))$$
$$\text{Total work required to investigate new cases} = w_c C(t)$$

Remember $C(t)$ is defined as in Equation 2.1. Now we can compute the portion new cases of all cases which will be used later.

$$\frac{w_c C(t)}{w_m(A_1(t) + I_1(t)) + w_c C(t)}$$

The amount of work that is available is the amount of work a contact tracer prosecutes multiplied by the amount of contact tracers.

$$q_w N_{trace}$$

Where q_w is the fraction of a day that consists of the work hours of each tracer. And N_{trace} is the number of available contact tracers.

The last part of the formulation of the flow out is implementing a bounded exponential function which provides a smooth approximation for the relationship between $\text{work}_{applied}$ and work_{demand}. When the $\text{work}_{applied} > \text{work}_{demand}$, the contact tracers can work on the backlog. In Figure 2.2 you can see the relation between $\text{work}_{applied}$ and work_{demand}. Here work_{demand} is equal to $w_m(A_1(t) + I_1(t)) + w_c C(t)$.

As a result we obtain the expression

$$[\text{flow}_{out}] := \frac{C(t)}{w_m(A_1(t) + I_1(t)) + w_c C(t)}[q_w N_{trace}(1 - \exp(-\alpha(w_m(A_1(t) + I_1(t)) + w_c C(t))))]$$

The α can be adjusted per situation, as shown in Figure 2.2, and can be fitted to approximate the linear function. The reason for this approximation of the linear function is that it has a lot of analytic benefits, since it is a smooth function.

Note that we have left out the w_c in the nominator of the fraction since we want to simulate the number of recently added people to the queue who have been investigated rather than the amount of hours.

The functions for the flow in and the flow out can be filled in in the queue as defined in Equation 2.1.

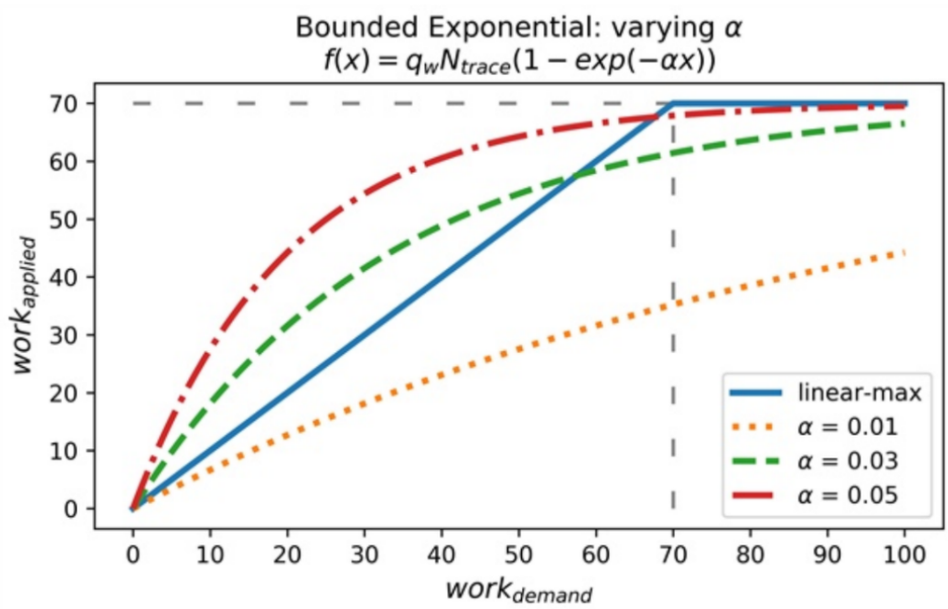


Figure 2.2: Bounded exponential as a smooth approximation of the linear relationship between work$_{applied}$ and work$_{demand}$

3

Modelling compartment models

A lot of research has already been done on compartment models and their dynamic structure, in this chapter one can find the information on these models which we are going to use in this thesis. We will take a look at the original SIR-model and its extension to the SEIR- and SEIRD-model to make the models applicable to COVID-19. For this section, we will use the data of Denmark as an example for calculations.

3.1. Assumptions of the model

The main assumption for the classical compartment model is that the total population is constant. This means that the same number of people are born as that have died that day, and that immigration and emigration are neglected. This assumption is a constraint for modelling diseases that exist for a long period of time in countries where the balance between the birth and death rate is not present.

The classical compartment models assume that here is a homogeneous mixing of the infected and susceptible people. This is a constraint for populations which act on measures such as social distancing and quarantine since it is obvious that the spread of a population is not homogeneous for that case.

Furthermore, when plotting a classical compartment model, the amount of susceptible people decreases towards zero. This is not the case for most viruses since the amount of time someone is immune is often not infinite.

3.2. SIR-model

The original compartment model is the SIR-model, containing the compartments $Susceptible$, $Infectious$ and $Recovered$. To indicate that these compartments are dynamic and that the people belonging to the compartment vary over time, we indicate the departments as functions of time: $S(t)$, $I(t)$ and $R(t)$. The dynamics between these compartments are illustrated in Figure 3.1 with its transition parameters between the compartments, the defnition of these parameters can be found in in Table 3.1.

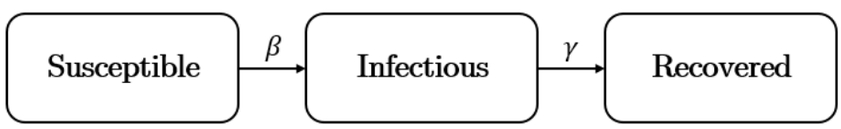


Figure 3.1: SIR-model and its transition parameters

Parameter	Unit	Explanation
N	-	Total population
β	-	Infection rate (average number of contacts per person per time multiplied by the probability of disease transmission in a contact)
γ^{-1}	Days	Average infectious time period

Table 3.1: Transition parameters of the SIR-model

The SIR-model dynamics can be described by the following differential equations. [2]

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{3.1}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \tag{3.2}$$

$$\frac{dR}{dt} = \gamma I \tag{3.3}$$

When adding all the equations, we find that $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$. It follows that $S(t) + I(t) + R(t) =$ constant $= N$, this number is the population size of the area you are forecasting the disease for.

We are interested in the values of parameters β and γ, we can use data of the countries we want to investigate to calculate them. Another interesting value we want to find is the R_0-value. This is the ratio of transmission and recovery rates: $R_0 = \beta/\gamma$. It represents the number of individuals that are infected by a single individual. The bigger the R_0-value, the faster the disease spreads. When the value is lower than one, the disease is not spreading.

3.2.1. Finding γ, β and R_0

We want to find the transition parameters β and γ, these can be found using the differential equations corresponding to the dynamics of the compartment model (Equations 3.1, 3.2 and 3.3). [2]

First, we will need an initial value for $S(t)$, this can be obtained intuitively. On the beginning of an epidemic, it holds that $S = N$ since no one has been infected yet. For simplicity we will denote $\beta - \gamma$ as m. We substitute this in Equation 3.2 and we obtain the following

$$\begin{aligned}
\frac{dI}{dt} &\sim \frac{\beta IS}{N} - \gamma I \\
\frac{dI}{dt} &\sim \beta I - \gamma I \\
\frac{dI}{dt} &\sim Im \\
I &= I_0 e^{mt} \\
\ln I &= mt + \ln I_0
\end{aligned}$$

m can be estimated using observed data of compartment $I(t)$ and a polynomial fit function in Python.

Next, we want to find γ, which is the inverse of the infectious time period. If we suppose that $I(t) = I_0$ (constant), we get the following

$$\frac{dR}{dt} = \gamma I_0 \tag{3.4}$$

$$R(t) = \gamma t I_0 \tag{3.5}$$

Now we can estimate γ by rewriting Equation 3.4 in terms of $R(t)$

$$\frac{R(t+dt)}{dt} = \gamma I \tag{3.6}$$

$$\gamma \approx \frac{R(t+1) - R(t)}{I(t)} \tag{3.7}$$

Where dt is the change in time which is set at 1 so we can easily retrieve data form $R(t)$ in the Python code. As the γ estimate for a period, we take the average of all the γ's in the last seven days of the period calculated using Equation 3.7.

Since we have an estimation for m and γ, we have an estimation for β now too because $m = \beta - \gamma$. And now that we have found β and γ, we can calculate the R_0-value.

3.2.2. SIR-models for Denmark

An overview of the exact dates of lock downs in Denmark is not available. Therefore, we decided to take a look at certain periods of time. We have data available for a period of 457 days which we split up in 6 periods of approximately 77 days. In Table 3.2 one can see the values for γ and β estimates in certain periods of time which are found using Python.

In Appendix A one can find the estimated SIR-models using the parameters from Table 3.2. We want to emphasise that the models that are made and shown in Figure A.1 forecast the following 500 days of the virus dynamics based on the data from a certain period. When making these forecasts, we can investigate whether the situation in a certain period is under control by looking at the values obtained in Table 3.2.

Period	Begin	End	β	γ^{-1}	R_0
0	22-01	07-04	0.1329	10.60	1.409
1	08-04	23-06	0.08355	13.62	1.138
2	24-06	09-08	0.1294	13.93	1.802
3	10-08	24-11	0.08433	13.92	1.165
4	25-11	10-02	0.05861	10.24	0.600
5	11-02	29-04	0.08846	15.04	1.330

Table 3.2: Period analysis of Denmark from January 22nd until April 29th

When analysing Table 3.2 together with Figure A.1, we can conclude that the pandemic is completely under control in period 4. We can see this from the low value of $R_0 = 0.600$ which is much lower than one, this means that the virus does not spread further. In Figures A.1b and A.1d one can see that the pandemic is also under control in periods 1 and 3, but since R_0 is not below 1 yet, one can not say it is completely gone.

3.3. SEIR-model

Since the incubation period is very important in the spread of COVID-19, the SIR-model is extended with the compartment $Exposed$. When an individual is assigned to this compartment, it means that the individual is infected by someone but does not show any symptoms yet. In this period, a person is able to spread the virus without him/her knowing it. The dynamics of the extension are visualised in Figure 3.2 and the definition of the transition parameters can be found in Table 3.3.

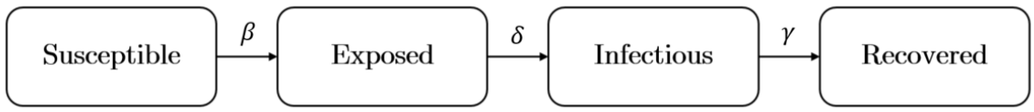


Figure 3.2: SEIR-model

Parameter	Unit	Meaning	Value
N	-	Total population	World-o-meter Found on `https://www.worldometers.info/world-population/`
β	-	Infection rate	Estimated from SIR-model
γ^{-1}	Days	Average infectious time period	Estimated from SIR-model
δ^{-1}	Days	Average incubation time	5

Table 3.3: Variables in SEIR-model

The differential Equations that describe the model are as follows

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N} \\
\frac{dE}{dt} &= \frac{\beta IS}{N} - \delta E \\
\frac{dI}{dt} &= \delta E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

Again, we have $S + E + I + R = N$, which is found the same way as for the SIR-model.

Since we have an extra compartment, we need extra data on initial values. Since we don't have any information on the amount of individuals in $E(t)$, we set this compartment equal to the array $I(t) \cdot R_0$, where R_0 is the reproduction number retrieved from the SIR-model in the corresponding period (Table 3.2).

We will not estimate the parameter δ the way we did for the parameters β and γ in the SIR-model. We use the β and γ estimated in the SIR-model (Table 3.2) and we set the δ as the medical information we have from the government which says the incubation time for COVID-19 is approximately five days. (RIVM `https://www.rivm.nl/coronavirus-covid-19/ziekte`, consulted May 5)

3.3.1. SEIR-model for Denmark

The SEIR-model is implemented and one can find the figures for the second period in Denmark in Appendix B in Figure B.1a. We decided to plot the models from period 2 since the corresponding SIR-model gives

a classic SIR-curve. In Figure B.1b only the infected and exposed individuals are plotted to give a better visualisation of their dynamics.

The biggest differences between the SEIR- and the SIR-model are that the peak of the infected people is lower and later. The lower peak can be explained from the fact that an incubation time is implemented which spreads out the infected individuals more. The later peak has to do with the incubation time too, since the people who are assigned to the $Infectious$-compartment in the SIR-model, will first be assigned to the $Exposed$-compartment in the SEIR-model.

3.4. SEIRD-model

Once we have implemented the SEIR-model, we find that it is also interesting to estimate the deceased people. This is another type of model since you can not be assigned to $Recovered$ and $Dead$ in one model as visualised in Figure 3.3. The parameters of the SEIRD-model are elaborated in Table 3.4.

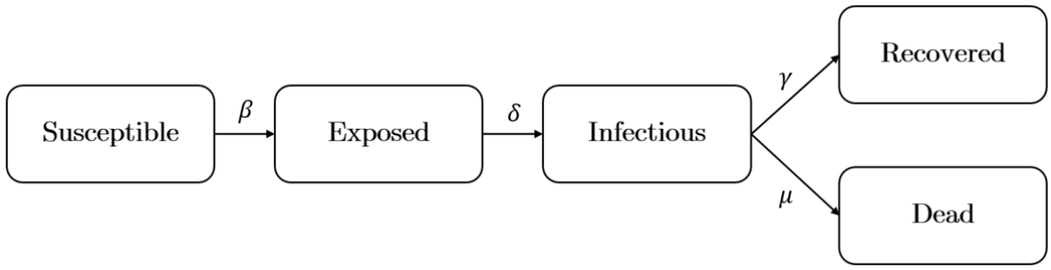


Figure 3.3: SEIRD-model

Parameter	Unit	Meaning	Value
N	-	Total population	World-o-meter*
β	-	Infection rate	Estimated from SIR-model
γ^{-1}	Days	Average infectious time period	Estimated from SIR-model
δ^{-1}	Days	Average incubation time	5
μ	-	Mortality rate	0.0103

Table 3.4: Variables in SEIRD-model. *`https://www.worldometers.info/world-population/`, consulted May 5

The differential equations corresponding to the dynamics of the compartments are as follows

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta IS}{N} \\
\frac{dE}{dt} &= \frac{\beta IS}{N} - \delta E \\
\frac{dI}{dt} &= \delta E - \gamma I - \mu I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \mu I
\end{aligned}$$

The only extra parameter is μ, which is the mortality rate of COVID-19. For now, we will estimate this rate by dividing the total death cases in Denmark by the total confirmed cases in Denmark until this day, this gives a mortality rate of $\mu = 0.0103$.

The reproduction number, R_0, for this model is equal to $\beta/(\gamma+\mu)$. [4]

3.4.1. SEIRD-model for Denmark

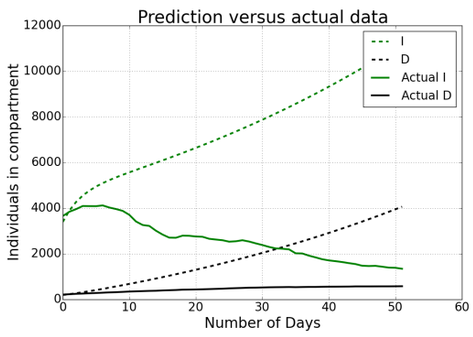
Since we have an extra compartment now, we need some extra information to simulate the virus: the amount of people that died up until a certain period of time. Luckily, the information on deceased people is posted up well in Denmark and we can retrieve this easily. The output of the model for period 2 in Denmark can be found in Appendix B in Figure B.1d.

The difference between the SEIR- and the SEIRD-model is that the peak of the infectives and exposed is lower in the SEIRD-model. The lower peak for the infectives is due to the fact that we now extract people from the infectives compartments and let them flow into the deceased compartment. The $Exposed$ compartment has a lower peak because less infectives imply less exposed individuals.

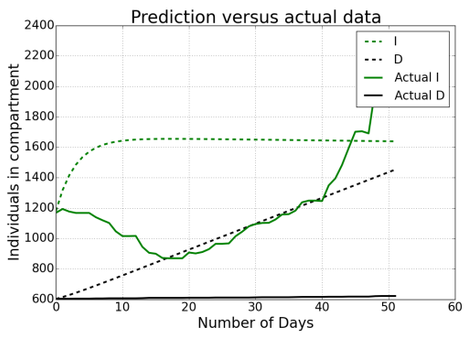
4

Improve estimation

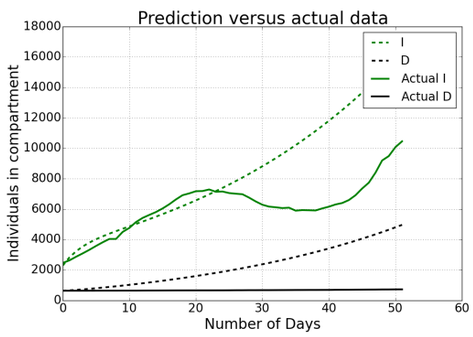
We are going to test the accuracy of our forecast for deceased and infectious individuals since those are the compartments we have observed data on and are interesting to look at. To test the accuracy of our forecast performance we can use observed data, this accuracy test is called sample testing. This means we are going to make a forecast with our model and compare it to the observed data. We can do this for periods zero to four since the fifth period forecasts unobserved data which means we are not able to compare it. The forecasts and the observed data plotted in the same graph can be found in Figure 4.1.



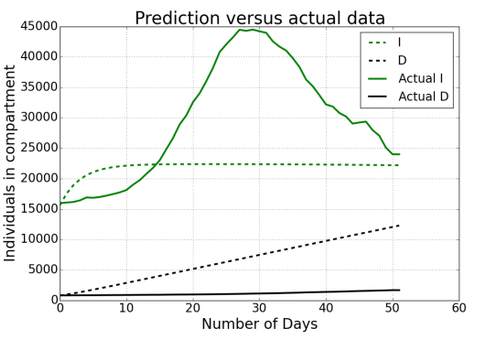
(a) SEIRD-model forecast post period 0 in Denmark



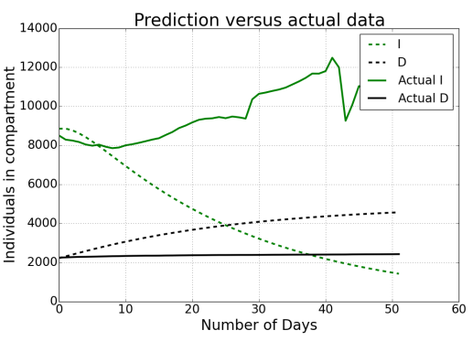
(b) SEIRD-model forecast post period 1 in Denmark



(c) SEIRD-model forecast post period 2 in Denmark



(d) SEIRD-model forecast post period 3 in Denmark



(e) SEIRD-model forecast post period 4 in Denmark

Figure 4.1: Forecasts using the SEIRD-model using the data of the respective period in Denmark

4.1. Accuracy testing

4.1.1. Least Squares Method

To improve our model in the following sections, we are going to use the Least Squares Method. The method defines a function which subtracts the estimated values from the actual data after which it minimises the sum of the squared subtractions by fitting the parameters.

$$\min. \sum_{t=1}^{T} \left(y_t - \hat{y}_t(x_t, \text{params})\right)^2$$

The output of this method is the parameters which ensure that the value of the sum is as low as possible, this indicates that using these parameters, the smallest error possible is obtained.

4.1.2. RMSE

To check the improvements of our estimation, we use the Root Mean Squared Error. This method measures the distance between the estimated and actual data if you would plot it on (x, y)-axes. If the RMSE is large, it means this distance is big which indicates a large error.

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2}$$

The tables in this chapter contain columns called 'Impr.', this is a column with the original RMSE from the SEIRD-model divided by the RMSE of the improved model. This is a number which indicates how much the adjustment of the model has improved the estimation.

4.2. Adjusting mortality rate

The first thing that stands out in Figure 4.1 is that the estimated amount of deceased individuals is way too high, which means our estimated μ should be lowered. The optimal μ for each period can be found using the Least Squares Method and can be found in Table 4.1. Since the values are not very far apart from each other in periods one to four, we decided to take the average of these as the new mortality rate. Period zero is not included in the mean since the information in this period is not reliable due to test unavailability in that time frame.

Period	μ
0	0.0012050
1	0.0001880
2	0.0001767
3	0.0005709
4	0.0008734
Mean	0.0004523

Table 4.1: Least Squares μ values per period in Denmark

We will include the result of these changes in the following section.

4.3. Adjusting time interval

Another way to improve our estimations is to take a better look at which time intervals we are retrieving data from to forecast. Remember that we took data from 22-01-2020 till 29-04-2021 and cut it equally in 6 periods. Since the SEIRD-model is not able to capture lockdown effects decreasing the amount of people in contact, we should estimate the parameters in our SEIRD-model using time intervals who do not have such peaks. From Figure 4.2 we selected periods that have a steady increase or decrease of infectives which might be a good data set to forecast with.

We use these periods to estimate the parameters γ & β for the SIR-model and act as if they are the same for the SEIRD-model, we use medical information on δ, and we set μ as calculated in Section 4.2. In Table 4.2 one can find the periods chosen to forecast with and the estimated parameter values, Figure 4.3 plots the corresponding SEIRD-model.

In Figures 4.3a and 4.3b we can see that the forecast is not very accurate, but in Figure 4.3c we can see that the forecast tends to the follow the trend of the actual data. One of the reasons for the high forecast in $Infectives$ in Figures 4.3a and 4.3b might be that the model assumes that everyone is in contact with each other, this is not the case due to social distancing and quarantine.

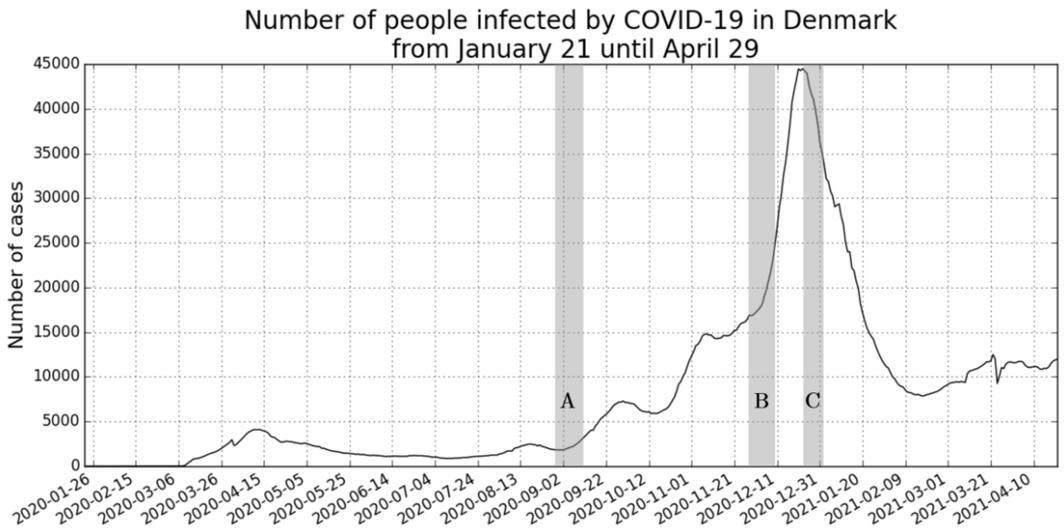
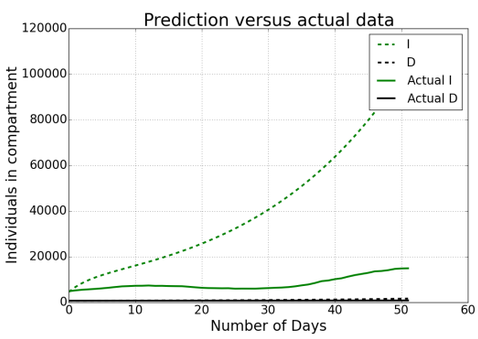


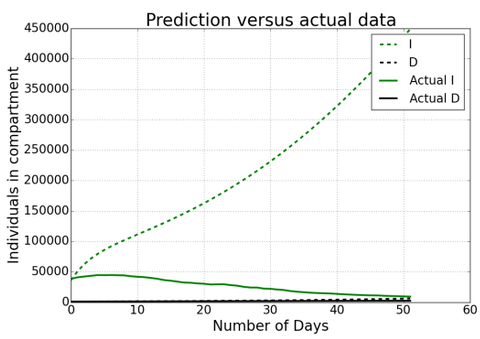
Figure 4.2: Infected people in Denmark, January 21 until April 29. Selected periods are highlighted and labelled.

Period	Begin	End	R_0	β	γ	μ	δ
A	31-08	17-09	2.60746	0.113365	0.0430249	0.0004523	0.2
B	26-11	10-12	2.29675	0.110959	0.0478591	0.0004523	0.2
C	23-12	02-01	0.63956	0.0622129	0.0968222	0.0004523	0.2

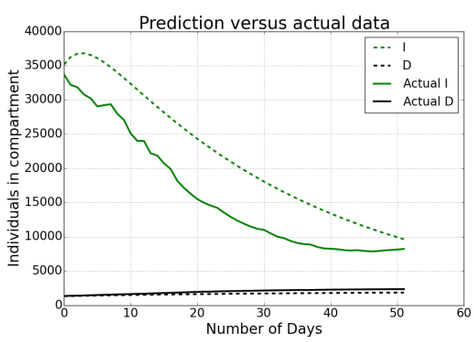
Table 4.2: SEIRD-model parameter estimations using assigned periods in Denmark



(a) SEIRD-model forecast post period A



(b) SEIRD-model forecast post period B



(c) SEIRD-model forecast post period C

Figure 4.3: Forecasts using specific periods found in Table 4.2

4.4. Adjusting relation between I and S

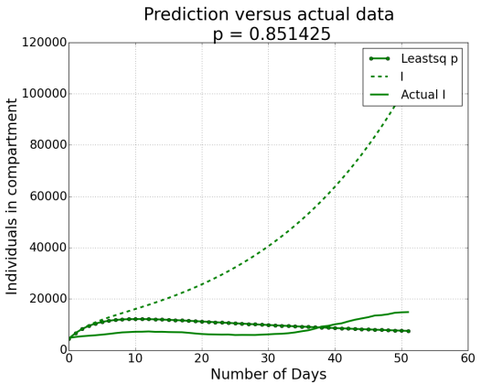
Classical compartment models as described in this report do not take social distancing and quarantine into account. When looking at differential equations of the SEIRD-model, one can see that βIS, the incidence rate, is a bi-linear term. This bi-linear term indicates a homogeneous spread of individuals in all the compartments, while this is not the case when people act on social distancing and quarantine.

To reproduce a heterogeneous spread of compartments $Infective$ and $Susceptible$, one can adjust

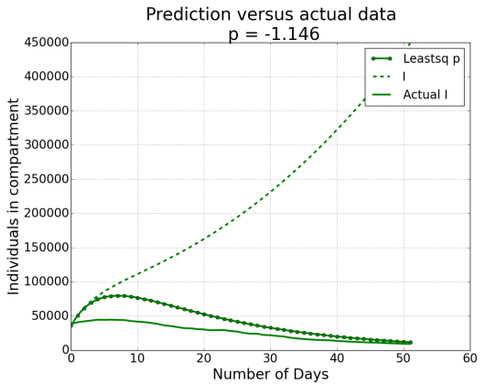
the classical SEIRD-model differential equations as follows [5]

$$\frac{dS}{dt} = -\frac{\beta I^p S}{N} \tag{4.1}$$
$$\frac{dE}{dt} = \frac{\beta I^p S}{N} - \delta E \tag{4.2}$$
$$\frac{dI}{dt} = \delta E - \gamma I - \mu I \tag{4.3}$$
$$\frac{dR}{dt} = \gamma I \tag{4.4}$$
$$\frac{dD}{dt} = \mu I \tag{4.5}$$

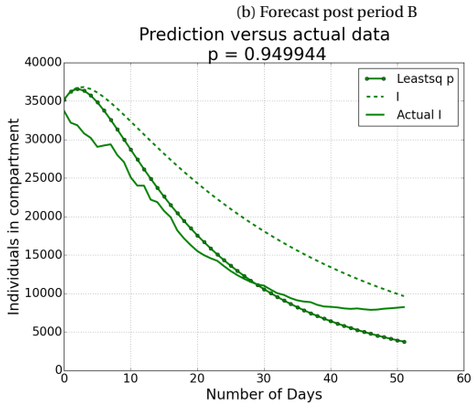
The p represents a power which is equal to one in the classical model, the parameter can be fitted using actual data and the Least Squares method. To show that this can result in a much better estimation, we plotted the models with the least squares fitted values for p in Figure 4.4. In Table 4.3 one can find the improvement in terms of RMSE-values of the estimations by fitting the p-value.



(a) Forecast post period A



(b) Forecast post period B



(c) Forecast post period C

Figure 4.4: Forecasts made with the SEIRD-model. The forecasts are fitted to observed data to show that a different value for p (as defined in Equations 4.1 and 4.2) can make a big difference for the accuracy of the estimation.

	Period	Begin	End	Original RMSE	p RMSE	Impr.
Denmark	A	31-08-20	17-09-20	40886.99	4346.84	9.41
	B	26-11-20	10-12-20	228880.03	20173.20	11.35
	C	23-12-20	02-01-21	6402.40	2748.08	2.34

Table 4.3: RMSE values of the fitted p-value model. Respective plots of the amount of infectives can be found in Figure 4.4.

4.5. Adjust reach of the model

In this section we want to counteract on the assumption that all individuals in a country are in contact with each other, since travelling between cities significantly reduces in the times of a lockdown. We adjust the

original SEIRD-model as follows [5]

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \cdot \frac{n_0}{N} \tag{4.6}$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} \cdot \frac{n_0}{N} - \delta E \tag{4.7}$$

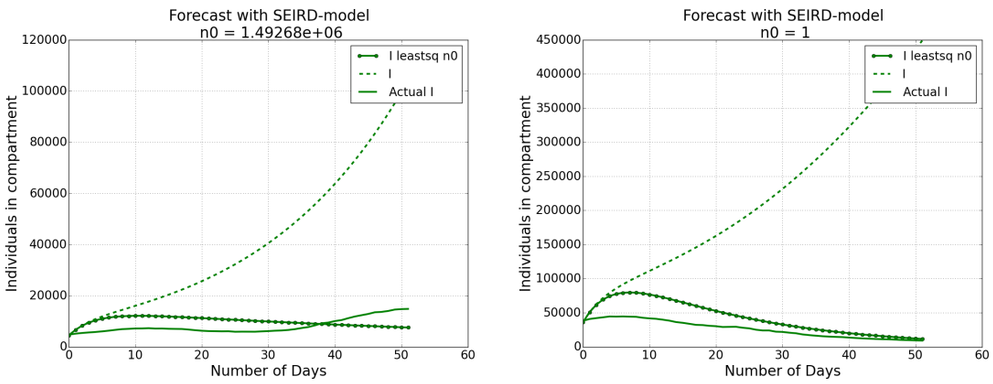
$$\frac{dI}{dt} = \delta E - \gamma I - \mu I \tag{4.8}$$

$$\frac{dR}{dt} = \gamma I \tag{4.9}$$

$$\frac{dD}{dt} = \mu I \tag{4.10}$$

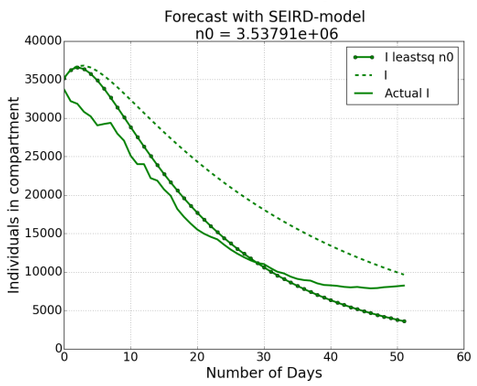
The n_0-value indicates the average size of networks in which people have contact with each other, this size decreases in times of lockdown. Using the Least Squares Method and the actual data, the following values for n_0 are found and the model is used to create forecasts. When the optimal n_0-value is negative, it is set to one and when it exceeds the number of inhabitants it is set to the population size N.

In Figure 4.5 one can find the forecasts using the model stated above and in Table 4.4 one can find the improvements in terms of RMSE-values.



(a) Forecast post period A

(b) Forecast post period B



(c) Forecast post period C

Figure 4.5: Adjusted forecasts made with the SEIRD-model. The forecasts are fitted to observed data to show that a different value for n_0 (as defined in Equations 4.6 and 4.7) can make a difference for the accuracy of the estimation.

In Table 4.4 one can find the improvement of the estimations by fitting the p-value.

	Period	Begin	End	Original RMSE	n_0 RMSE	Impr.
Denmark	A	31-08-20	17-09-20	40886.99	4358.76	9.38
	B	26-11-20	10-12-20	228880.03	20173.20	11.35
	C	23-12-20	02-01-21	6402.40	2821.74	2.27

Table 4.4: RMSE values of the fitted n_0-value model. Respective figures can be found in Figure 4.5.

5

Verify improvements

It is obvious that both the adjustments of the models create a better forecast, but up until now we have only taken a look at data for Denmark, which is not enough data to base our research on. This is why data of Austria, Germany, Italy and Poland is used to do similar calculations with.

First, we are going to look at the observed values for the infected individuals over the time of the pandemic (Appendix C) to choose periods of which we will use data to forecast. Then we will calculate the parameters γ and β with the SIR-model (Section 3.2.1), use medical information to determine the value of δ, and calculate the mean mortality rate μ to obtain the basic SEIRD-model. At last we will improve the model by fitting the p-value (as in Equations 4.1 and 4.2) or the n_0-value (as in Equations 4.6 and 4.7). An overview of the parameter values and the results for the improvements can be found in Table 5.2

From Table 5.2 one can see that the RMSE-values become much lower for both improved models, but the improvement rates are very similar for both models when the p- or n_0-value is adjusted. Since the n_0 has restrictions ($0 \leq n_0 \leq N$) and p does not, we decide to leave the n_0/N term out of our model. Another reason for this decision is is that n_0/N is a linear term, by adjusting the β in Equation 4.1, one can include n_0/N if necessary.

We have proven that the modifications provide a more accurate prediction. The question is: How are we going to apply this? We want to use the compartment models as a prediction of the future dynamics of a virus. We used the Least Squares Method in this chapter, which means we used observed data to determine the optimal parameter values which we can not do in reality. We have information on the optimal p-value for three periods per country, with this information we can estimate a general p-value for each country by taking the mean of the values noted in Table 5.2. We set the p-values which are a negative number equal to zero since this does not differe very much in RMSE (which will be elaborated on later). We obtain the following p-values which will be used for the models from here on.

Denmark:	0.6604
Austria:	0.9459
Germany:	0.9731
Italy:	0.6607
Poland:	0.6626

To verify the choice to set the p-values at zero when it is negative, one can find the difference between the RMSE's in Table (Table 5.1). We can conclude that this difference is insignificant.

Country	Period	Optimal p-value	Used p-value	RMSE optimal p-value	RMSE used p-value
Denmark	B	-1.146	0.6604	20173.20	21271.68
Italy	A	-0.5640	0.6607	942596.63	959165.10
Poland	A	-0.8086	0.6626	204232.96	208562.30

Table 5.1: RMSE-values for SEIRD-models using originally calculated p-values (negative) compared to p-value means calculated without these out liers.

		Begin	End	R_0 (SIR)	β	γ	μ	δ	RMSE (I)	p	p-RMSE	Impr.	n_0	n_0-RMSE	Impr.
Denmark	A	31-08-20	17-09-20	2.634	0.1133	0.04302	0.0004523	0.2	41531.41	0.8514	4346.84	9.41	1492683	4358.76	9.38
	B	26-11-20	10-12-20	2.318	0.1109	0.04785	0.0004523	0.2	243971.97	-1.146	20173.20	11.35	1	20173.20	11.35
	C	23-12-20	02-01-21	0.642	0.0622	0.09682	0.0004523	0.2	6919.99	0.9599	2748.08	2.34	3537912	2821.74	2.27
Austria	A	18-10-20	01-11-20	2.552	0.1519	0.05954	0.0010054	0.2	430956.98	0.8492	12554.46	34.32	1650681	12551.08	34.33
	B	22-11-20	01-12-20	0.626	0.0642	0.10260	0.0010054	0.2	6349.83	1.0050	6301.38	1.00	9053044	6349.83	1.0
	C	21-02-21	03-03-21	1.214	0.1013	0.08340	0.0010054	0.2	15822.40	0.9837	4828.43	3.27	7392494	4815.13	3.28
Germany	A	12-10-20	27-10-20	2.684	0.1104	0.04113	0.0009576	0.2	1166899.61	0.9196	75527.78	15.44	25987404	73896.30	15.79
	B	22-01-21	04-02-21	0.471	0.0395	0.08398	0.0009576	0.2	118044.72	1.058	41767.47	2.82	84027485	118044.72	1.0
	C	16-03-21	02-04-21	1.337	0.0739	0.05526	0.0009576	0.2	386515.69	0.9416	90063.80	4.29	39487705	90208.68	4.28
Italy	A	19-10-20	27-10-20	7.291	0.0966	0.01324	0.0004409	0.2	8381148.60	-0.5640	942596.63	8.89	1	941113.73	4.89
	B	07-12-20	17-12-20	0.629	0.0242	0.03857	0.0004409	0.2	121077.31	0.9876	111705.48	1.08	51149016	111807.92	1.08
	C	09-04-21	17-04-21	0.739	0.0275	0.03726	0.0004409	0.2	76080.60	0.9945	72012.66	1.05	56102163	72036.10	1.05
Poland	A	17-10-20	01-22-20	3.572	0.1083	0.03033	0.0007413	0.2	3040177.78	-1.183	204232.96	14.88	1	209182.19	14.88
	B	29-11-20	09-12-20	0.483	0.0295	0.06101	0.0007413	0.2	85796.76	1.038	42968.20	1.99	37809208	85796.76	1.0
	C	03-03-21	10-03-21	1.168	0.0449	0.03849	0.0007413	0.2	209091.28	0.950	63799.49	3.27	19773743	63773.88	3.27

Table 5.2: Period analysis of Denmark, Austria, Germany, Italy and Poland.

6

Compartment model for queue

We are now in a position to simulate the contact tracing queue using the compartment models we have created in the previous chapter. The people who will be investigated are the people who have been in contact with a person who has been tested positive for COVID-19 by the government. This means that we need to investigate how many people are expected to test positive, from that number on we can estimate the amount of contacts that need investigation.

This chapter contains two possible extensions of the SEIRD-model which can be used for the simulation of the contact tracing queue. The first has an extra compartment called 'Tested', and the second extracts information from the classical SEIRD-model.

6.1. Compartment model 1

Using our knowledge form the research we have done on compartment models, it is possible to adjust an existing compartment model that is relevant for us. The differential equations that describe the dynamics for a certain compartment is its flow in of the minus its flow out.

Since our former compartment models do not include a compartment of people who are actually tested, we do not know how many people have to be traced. This is why we decided to include this compartment, the new compartment model can be found in Figure 6.1.

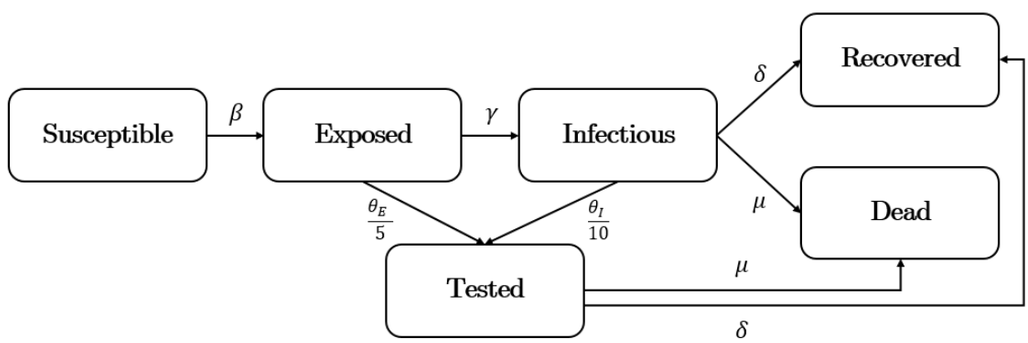


Figure 6.1: SEIRDT-model and its transition parameters

The θ_E and θ_I represent the fraction of their corresponding compartments who get themselves tested, which will automatically result in a positive test since we assume that everyone who is tested in compartment E and I will obtain a positive test result. We divide this number by 5 days (incubation time) and 10 days (infectious time) respectively since a person only tests positive once in the time span they are assigned to those compartments.

The differential equations that correspond to this compartment model are as follows

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta I^p S}{N} \\
\frac{dE}{dt} &= \frac{\beta I^p S}{N} - \gamma E - \frac{\theta_E}{5} E \\
\frac{dI}{dt} &= \gamma E - \delta I - \mu I - \frac{\theta_I}{10} I \\
\frac{dR}{dt} &= \delta I + \delta T \\
\frac{dD}{dt} &= \mu I + \mu T \\
\frac{dT}{dt} &= \frac{\theta_E}{5} E + \frac{\theta_I}{10} I - \delta T - \mu T
\end{aligned}$$

Unfortunately, this model does not work as hoped, because part of the individuals of compartments $Exposed$ and $Infectious$ are now put into one compartment. Due to this, the length of the stay in compartment $Tested$ is no longer correct since the parameter γ is no longer taken into account. We could adjust the compartment model in such a way that γ is also taken into account by splitting up the Compartment $Tested$ into two compartments, but instead we choose to create an external model explained in the next section.

6.2. Compartment model 2

Another way to simulate the queue is not to make another compartment model but to use one we have already implemented and extract information from it. We will do this using the SEIRD-model (Section 4.4).

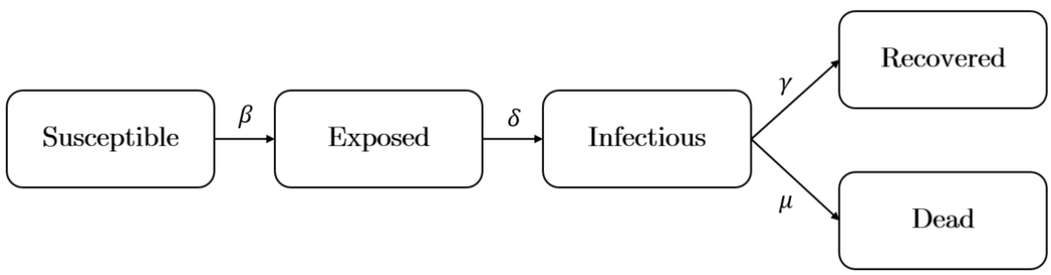


Figure 6.2: SEIRD-model and its transition parameters

From compartments $Exposed$ and $Infectious$ we can estimate how many contacts have to be investigated. We assume that an individual that is assigned to $Exposed$ will stay in that compartment for 5 days (incubation time is 5 days) and an individual that is assigned to $Infectious$ will stay in that compartment for 10 days (average infectious time).

We will make use of the parameters θ_E and θ_I (as described in Section 6.1 and their divisions by corresponding duration times to compute the number of people that test positive from both compartments E and I on time t. In Figure 6.3 you can see the visualisation for a model that retrieves the total amount of work that is needed at time t. κ represents the average number of contacts per individual and w_m and w_c the amount of time it costs to monitor or treat a new case respectively.

The left part of the overview in Figure 6.3 results in the amount of work it costs to investigate a contact of someone who tests positive on time t, this work is for the new cases who are contacts of people who tested positive on day t. The right part results in the amount of time it costs to monitor people, these individuals are contacts of people who tested positive in the last five days. The amount of people to be monitored can easily be retrieved from previous 'New contacts at time t' values calculated at time $t-1$ to $t-5$. When the work to investigate new cases and monitoring cases is added, we obtain the total work that is demanded to do proper research on contacts on day t.

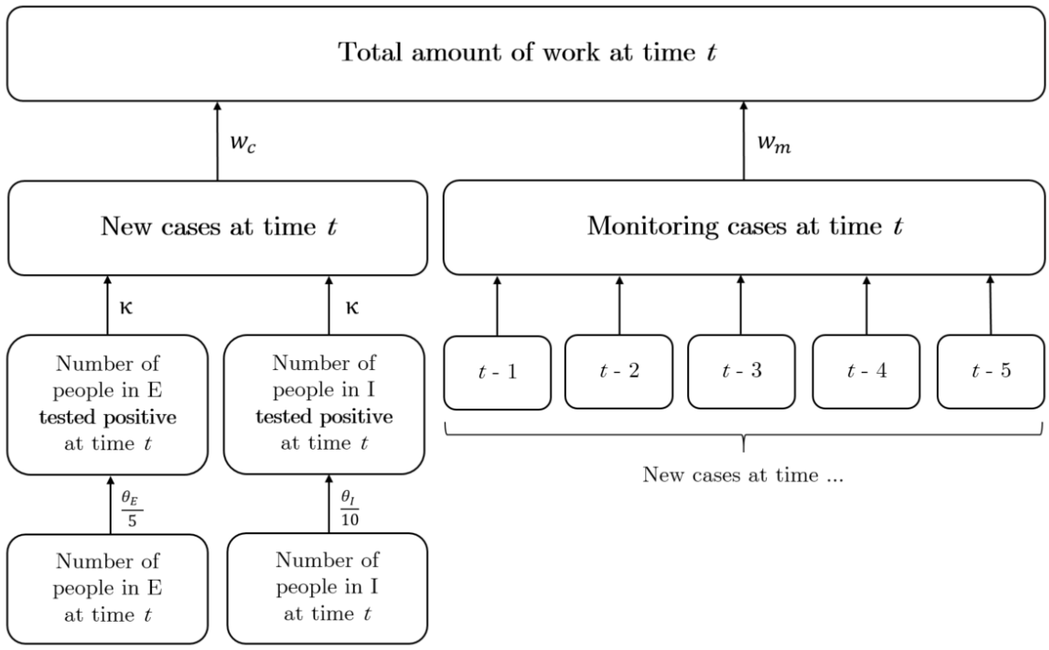


Figure 6.3: Visualisation of the total amount of work it costs to do contact tracing for day t

7

Queue simulation

The goal of the model made in the article this research is based on (Chapter 2), is different from the goal of the queue simulation in this chapter. The main difference is that the model in the article focuses on the queue of people who have been in contact with an infective **recently**, while this chapter focuses on the queue of amount of work it costs to investigate **all** individuals, which includes the people who are monitored. Another distinction is that the article has an analytically defined queue, which we will convert to a queue which can be approximated using numerical methods. The final change is that the compartment model on which the original queue is based on, is not used in this chapter due to lack of information on parameters, instead we use data retrieved from the SEIRD-model as described in Section 6.2.

In this chapter we will walk through the changes made to the original queue, which will result on a notation of the contact tracing which can be used for our research.

7.1. Queue

As mentioned above, the queue has a different goal in the original model, we will make the following changes in notation and definition to adapt the queue to our research. In Table 7.1 (Section 7.4) one can find the meaning of all the parameters used in the notation.

$\frac{dC}{dt} = \left[\text{flow}_{in}(t)\right] - \left[\text{flow}_{out}(t)\right]$	$\longrightarrow$	$Q(t) = \left[\text{flow}_{in}(t)\right] - \left[\text{flow}_{out}(t)\right]$
Change in new cases needed to be investigated at time t	$\longrightarrow$	Amount of work left to investigate new cases and monitoring cases at time t

7.2. Positive tested individuals

First of all, we want to define the amount of people that test positive on day t, from this number on we can estimate the flow in and flow out of the queue.

The amount of positive tested individuals on day t is defined as follows.

$$T(t) = \frac{\theta_E}{5} E(t) + \frac{\theta_I}{10} I(t)$$

Where θ_E is the fraction of $Exposed$ individuals who test positive in one day, and θ_I is the fraction of $Infective$ individuals who test positive in one day. We assume that individuals only test positive once in their stay in a compartment, we resolve this by dividing the θ.'s by the amount of days an individual is assigned to that compartment.

7.3. Flow in

At first we will focus on the flow_{in} of the queue. In the previous definition this existed of the amount of contacts of the individuals who test positive on day t, from now on this exists of the contacts of the people who test positive on day t plus the amount of people that need monitoring at time t multiplied by the corresponding amount of time it costs to contact trace them (w_c or w_m).

$$\text{flow}_{in}(t) = w_c T(t) + w_m \sum_{i=1}^{5} T(t-i)$$

7.4. Flow out

The flow out of the model is changed in a few ways. First of all, we will look at all cases that are worked on instead of only the new ones that come in. This means we are also interested in the cases that need to be monitored at time t. Secondly, the exponential expression will be converted to a linear function. And at last we will change the unit from number of people to total amount of work applied on contact tracing.

The original flow out is as follows.

$$\text{flow}_{out} = \frac{T(t)}{w_m \sum_{i=1}^{5} T(t-i) + w_c T(t)} \left(w_a M \Big(1 - \exp[-\alpha(w_m \sum_{i=1}^{5} T(t-i) + w_c T(t))] \Big) \right) \tag{7.1}$$

$$= \frac{1}{w_c} \frac{\text{Work to investigate new cases}}{\text{Total amount of work}} \left(w_a M \Big(1 - \exp[-\alpha(\text{Total amount of work})] \Big) \right) \tag{7.2}$$

The fraction in Equation 7.1 can be left out completely for our flow out notation since we are interested in the work that is put into all cases, not only the new ones. Furthermore, we lose the exponential expression since this expression is a smooth approximation of the linear relationship between work demand and work applied as illustrated in Figure 2.2. This results in the following notation of the flow out.

$$\text{flow}_{out} = \text{work}_{applied}(\text{work}_{demand}, \text{work}_{available}) \tag{7.3}$$

Where

$$\text{work}_{applied}(\text{work}_{demand}, \text{work}_{available}) = \begin{cases} \text{work}_{demand} & \text{if work}_{demand} < \text{work}_{available} = w_a M \\ \text{work}_{available} & \text{if work}_{demand} \geq \text{work}_{available} = w_a M \end{cases}$$

Parameter	Unit	Explanation
M	-	Number of contact tracers
κ	-	Average number of contacts per person per day
θ_E	-	Fraction of exposed individuals who test positive
θ_I	-	Fraction of infectious individuals who test positive
w_c	Days	Amount of time to investigate a new contact
w_m	Days	Amount of time to monitor existing cases
w_a	Days	Amount of time a contact tracer works

Table 7.1: Parameters simulation queue

7.5. Results queue simulation

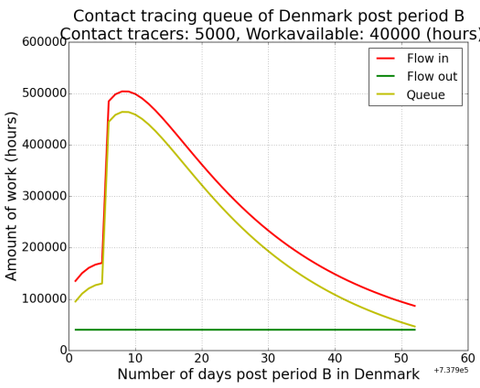
Using the parameter values in Table 5.2, we produce SEIRD-models to estimate the amount of $Infectives$ and $Exposed$ which we use to predict the amount of people whose contacts need contact tracing. We will need values for some extra parameters which can be found in Table 7.2. Since we do not have information on these values for now, we will make some educated guesses on their values. Substituting various values for the amount of contact tracers (M) gives us different situations of the queue.

Parameter	Value	Unit	Explanation
M	Variable	-	Number of contact tracers
κ	6	-	Average number of contacts per person per day
θ_E	1/6	-	Fraction of exposed individuals who test positive
θ_I	4/5	-	Fraction of infectious individuals who test positive
w_c	10/60	Days	Amount of time to investigate a new contact
w_m	4/60	Days	Amount of time to monitor existing cases
w_a	1/3	Days	Amount of time a contact tracer works

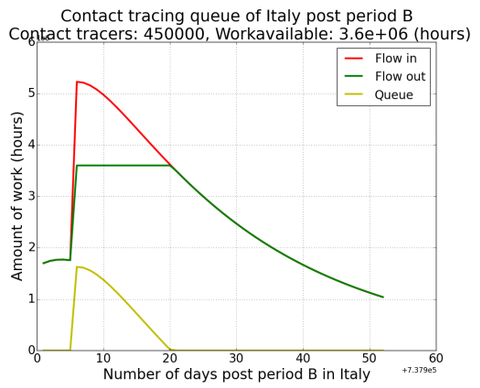
Table 7.2: Parameters simulation queue values

The model mentioned in this chapter is implemented in Python and we obtain the queues for contact tracing shown in Figure 7.1, including the SEIRD-model forecasts for corresponding country and period. We decided to plot the contact tracing queues for Denmark post period B and for Italy post period B and adjust the amount of contact tracers to show what differences it can make.

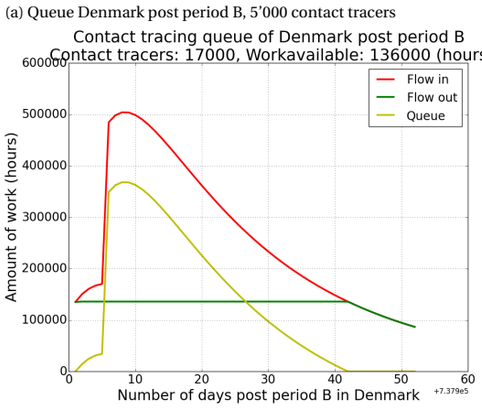
In Figure 7.1 you can find three situations. The first is plotted in the Figure 7.1a, the queue is long since there are not enough contact tracers. You can see this by the constant flow out, it is always equal to 4,000 since Equation 7.3 tells the flow out to be constant at $\text{work}_{available}$ when work_{demand} is exceeding $\text{work}_{available}$ at all times. The second situation is plotted in Figures 7.1b and 7.1c where flow out is eventually equal to the flow in which results in a queue of zero hours. The last situation is plotted in Figure 7.1d where you can see that there are a lot of contact tracers and in a short amount of time the queue is already equal to zero.



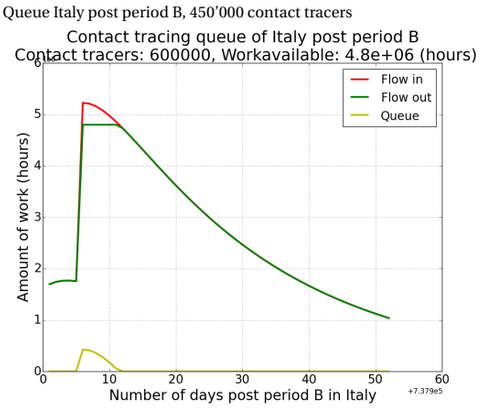
(a) Queue Denmark post period B, 5'000 contact tracers



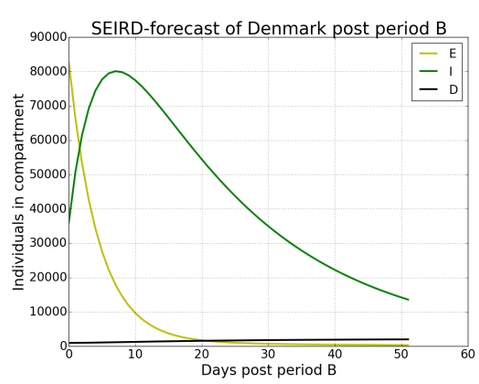
(b) Queue Italy post period B, 450'000 contact tracers



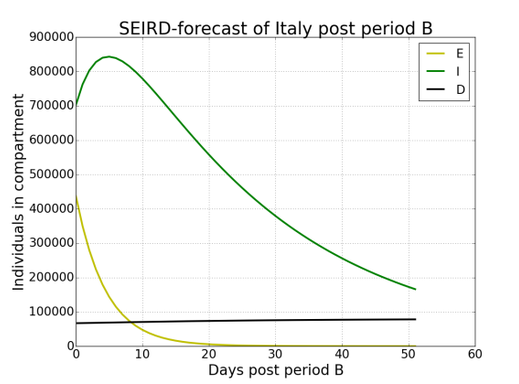
(c) Queue Denmark post period B, 17'000 contact tracers



(d) Queue Italy post period B, 600'000 contact tracers



(e) SEIRD-model Denmark post period B



(f) SEIRD-model Italy post period B

Figure 7.1: Contact tracing queue's of Denmark and Italy created with the corresponding SEIRD-models. Adjusting the number of contact tracers results in a different queue.

8

Amplify queue

The queue simulated in Chapter 7 is created to estimate how many contact tracers are necessary to investigate new cases and monitor them. If enough contact tracers are at work, one could say the situation is 'under control' since enough contacts of infective/exposed individuals are in contact with the government. The question is: What happens when there are not enough contact tracers and the people who have been in contact with COVID-19 positive individuals are not being monitored by contact tracers? What happens if the government is not able to get an overview of the status of all these potential infected individuals?

To simulate a situation in which more individuals will be assigned to the compartment $Exposed$ (and less to the compartment $Susceptible$) when the queue is too long, we adjust the SEIRD-model in the following way.

$$\frac{dS}{dt} = -\frac{\beta I^p S a}{N} \tag{8.1}$$

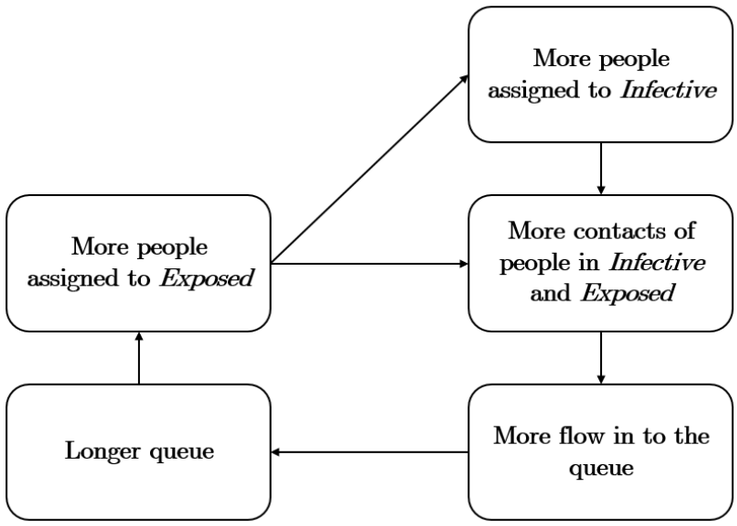
$$\frac{dE}{dt} = \frac{\beta I^p S a}{N} - \delta E a \tag{8.2}$$

$$\frac{dI}{dt} = \delta E a - \gamma I - \mu I \tag{8.3}$$

$$\frac{dR}{dt} = \gamma I \tag{8.4}$$

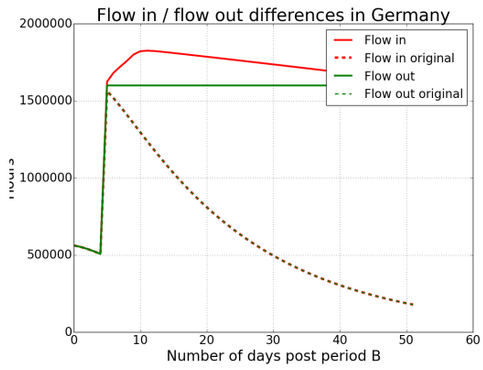
$$\frac{dD}{dt} = \mu I \tag{8.5}$$

The a is the amplification factor which is set equal to a number between 1.1 and 5 if the queue of day $t-1$ exceeds the queue$_{max}$ (set by the government), if the queue$_{max}$ is not exceeded, the value of a is set equal to one as in the normal SEIRD-model. This makes sure that more individuals are assigned to the compartment $Exposed$ when the government has not enough control over citizens that might have been in contact with COVID-19 infected individuals. This implementations allows the queue to be dependent on itself, which can easily result in an unstable situation, visually represented below.

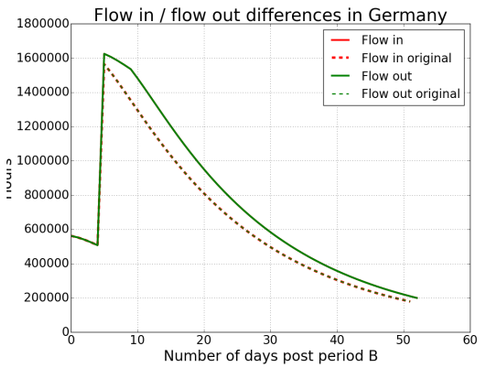


In Figure 8.1 one can find a plot in which the difference between the the original and the new model becomes visible for the flow in / flow out, the queue and the SEIRD-model. The plots are a simulation of

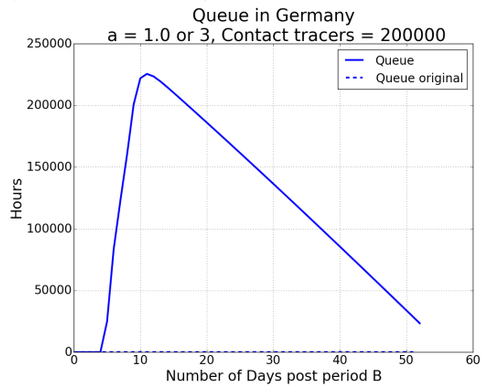
the situation in which the amplification factor is set equal to 3 when the queue$_{max}$ = 1,000 is exceeded. All plots are estimations of the amplified queue in Germany post period B, the left plots simulate the amplified queue when there are 200,000 contact tracers available and the right plots when there are 210,000 contact tracers available.



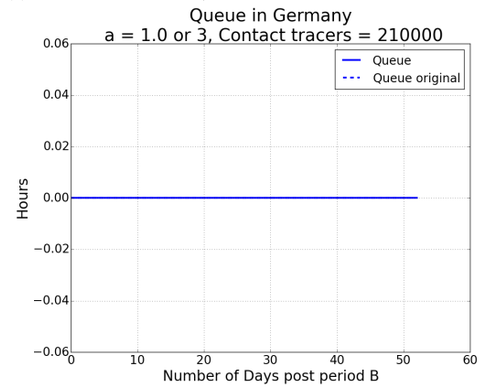
(a) Flow in and flow out 200,000 contact tracers



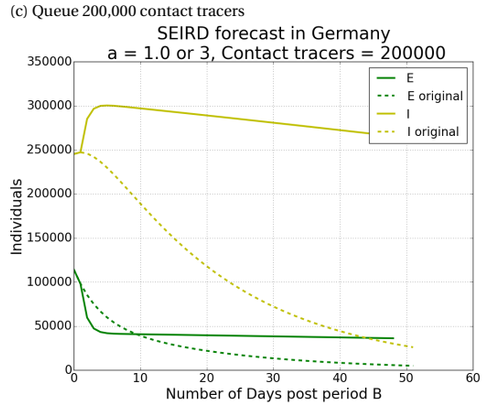
(b) Flow in and flow out 210,000 contact tracers



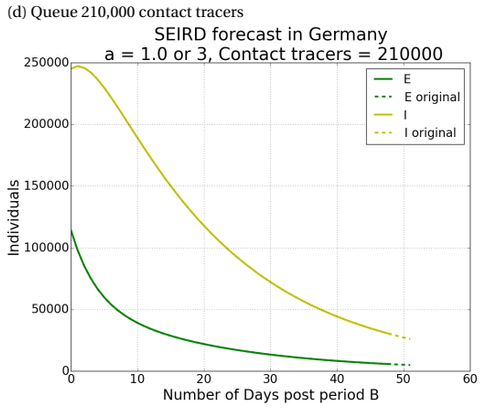
(c) Queue 200,000 contact tracers



(d) Queue 210,000 contact tracers



(e) SEIRD-model 200,000 contact tracers



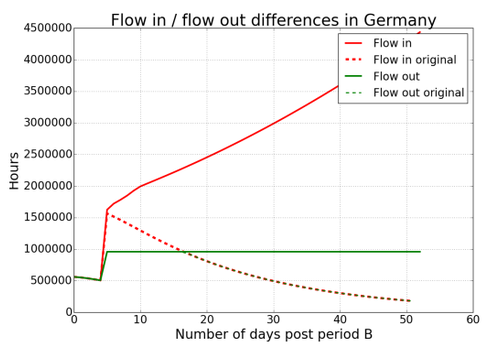
(f) SEIRD-model 210,000 contact tracers

Figure 8.1: Plots of the SEIRD-models including an amplification factor of 1.0 or 3 as defined in Equations 8.1 and 8.2. The amount of contact tracers is variable to show the difference between the outcome of the models.

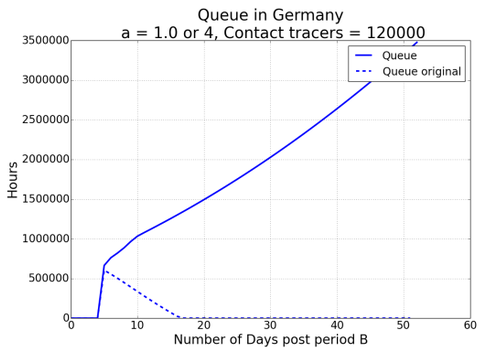
Both situations plotted in Figure 8.1 result in a stable solution. The only difference is that there is more flow in when there are 200,000 contact tracers available which results in a longer queue. Furthermore, there exists no queue if there are 210,000 contact tracers since the flow in is equal to the flow out. This means that the amount of people assigned to $Exposed$ is not amplified.

To simulate an unstable situation, we set the amplification factor equal to 4 and the contact tracers available is 120,000, the results can be found in Figure 8.2.

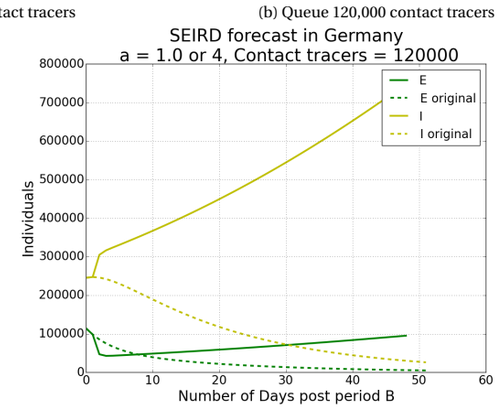
This model can be used to find the amount of contact tracers that are needed for a stable situation, but only if there is proof that when the government has no control over its citizens that have been in contact with COVID-19 infected individuals, there are more people exposed to the virus.



(a) Flow in and flow out 120,000 contact tracers



(b) Queue 120,000 contact tracers



(c) SEIRD model 120,000 contact tracers

Figure 8.2: Plots of the SEIRD-models including an amplification factor of 1.0 or 4 as defined in Equations 8.1 and 8.2. Setting the amplification factor at 4 results in unstable situation.

9

Extending the SEIRD-model with vaccinations

From January on, it is possible to be vaccinated for COVID-19. Since this will decrease the amount of individuals assigned to the compartment $Susceptible$ drastically, it is important to take a look at the difference it can make for the amount of infected individuals in a country. This can be done using a compartment model in which an amount individuals get assigned to a compartment, $Vaccinated$, each day. The model can be found in Figure 9.1 and will be referred to as the 'SVEIRD-model'. [3]

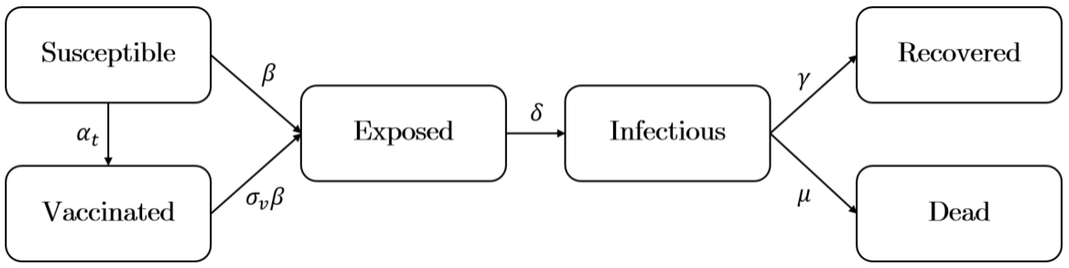


Figure 9.1: SVEIRD-model

The corresponding differential equations are as follows.

$$\frac{dS}{dt} = -\frac{\beta I^p S}{N} - \alpha_t S \tag{9.1}$$

$$\frac{dV}{dt} = \alpha_t S - \frac{\beta \sigma_v I^p V}{N} \tag{9.2}$$

$$\frac{dE}{dt} = \frac{\beta I^p S}{N} - \delta E + \frac{\beta \sigma_v I^p V}{N} \tag{9.3}$$

$$\frac{dI}{dt} = \delta E - \gamma I - \mu I \tag{9.4}$$

$$\frac{dR}{dt} = \gamma I \tag{9.5}$$

$$\frac{dD}{dt} = \mu I \tag{9.6}$$

We define α_t as the vaccination rate and σ_v as the vaccine efficacy rate which both will be elaborated on in Section 9.2.

The last terms in Equations 9.2 and 9.3 are in the form for the following reason. When a vaccine has an efficacy rate of 66%, it does not imply that 34% of the individuals shot with the vaccine are still susceptible. It means that if someone who is vaccinated is in contact with a person infected by COVID-19, the chances are 34% this vaccinated individual is infected by COVID-19. This means that we still have to take the transmission rate (β), the chances of people being in contact with each other (multiplying V by I) and the social distancing (p-value) into account.

This chapter contains the results of this new model, and eventually we will sketch different scenarios by adjusting the value for α_t (vaccination rate) to see what difference it can make for the future. We want to emphasise that this chapter is not testing accuracy anymore but investigates possible different scenarios if more time and money is invested in vaccination.

9.1. Data time frames

On the beginning of this bachelor thesis, there was data available until 29-04-2021. Since there is more data available now, we have imported this data on $Infectives$, $Recovered$ and $Deaths$ for Denmark, Austria, Italy, Poland and Germany. We are going to combine the data on vaccinated people with the newly imported data to estimate parameters for the SVEIRD-model.

To estimate the amount of individuals assigned to $Vaccinated$, we will only look at the data for individuals that are fully vaccinated (some vaccinations require two shots 14 days apart) because there is no reliable information on the efficacy of these vaccinations when someone has had only one shot. One can find the amount of individuals being fully vaccinated in Denmark each day in Figure 9.2, this figure shows that from mid March on, a steady amount of individuals are being vaccinated and from May on this number increased significantly.

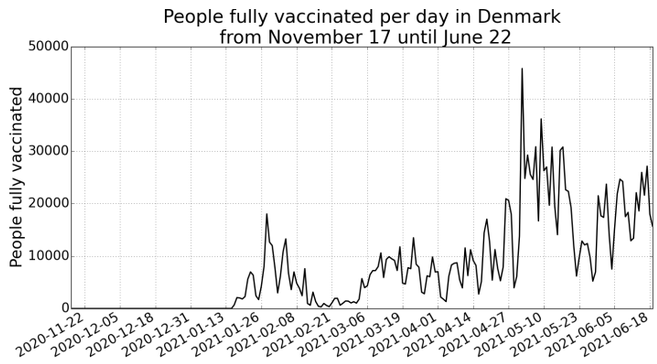


Figure 9.2: Fully vaccinated people difference each day in Denmark

In Figure 9.3 one can find an overview of the time frames with whom the SIR-, SEIRD- and SVEIRD-model parameters are calculated. The μ and p-values are the same as for the SEIRD-model we used in previous chapters.

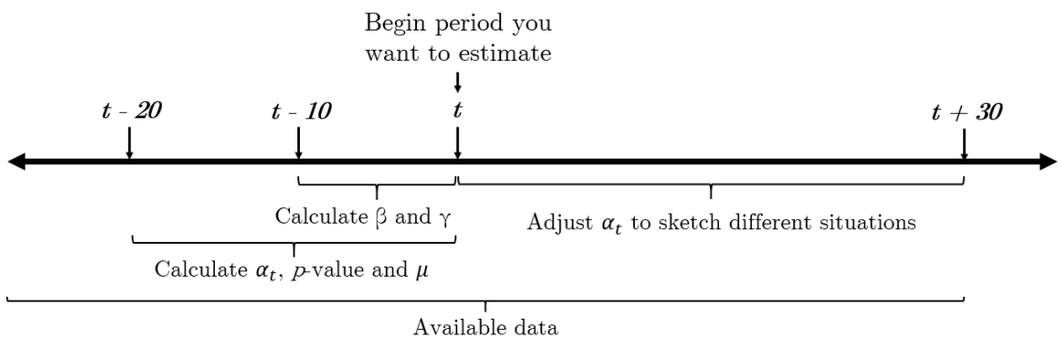


Figure 9.3: Time frames used to calculate parameters

9.2. Parameters α_t and σ_v

9.2.1. Vaccination rate α_t

The parameter α_t is the vaccination rate at day t, which is the percentage of susceptible individuals vaccinated that day. We assume that only individuals in compartment $Susceptible$ get vaccinated.

To decide what value α_t takes, we divide the average of vaccinations over 20 days before the period we are investigating by the amount of individuals that are assigned to the compartment $Susceptible$ on the previous day. $V_{obs}(t)$ is defined as the observed data on vaccinated individuals.

$$\begin{aligned}\alpha_t &= \frac{\sum_{t=1}^{20} V_{obs}(t-i)}{20} \cdot \frac{1}{S(t-1)} \\ &= \frac{\text{Mean 20 days pre period}}{S(t-1)}\end{aligned}$$

Since the compartment $Susceptible$ decreases each day due to vaccination, α_t increases each day.

The average vaccinations per day for the period 02-05-2021 until 22-05-2021 per country can be found in Table 9.1. From now on, we will refer to this period as 'Period D'. To give a better idea of the amount of people vaccinated each day with respect to the population size, we added a column that contains the percentage of inhabitants vaccinated each day.

Country	Period	Average vaccinations per day	Average / population size
Denmark	D	23445	0.40%
Austria	D	20344	0.22%
Germany	D	254100	0.30%
Italy	D	191541	0.32%
Poland	D	127330	0.34%

Table 9.1: Average vaccinations in the period from 02-05-2021 until 22-05-2021 per country.

9.2.2. Vaccination inefficacy σ_v

The parameter σ_v is the vaccination inefficacy rate. Research has been done to find the efficacy rate of certain vaccines, these values will be used in the model and can be found in Table 9.2.

Vaccine	Efficacy	σ_v	Source
Pfizer	95%	0.05	WHO*
Johnson & Johnson	66%	0.34	WHO**
Moderna	94%	0.06	WHO***
Astrazeneca	63%	0.37	WHO****

Table 9.2: Inefficacy rates for COVID-19 vaccines.
*https://www.who.int/news-room/feature-stories/detail/the-j-j-covid-19-vaccine-what-you-need-to-know?gclid=CjOKCQjw_dWGBhDAARIsAMcYuJyOSptG_NIAYChDKf7vyqgNV3NEuKqH6GFvUg96YLm1tYqphZUweVYaAvLUEALw_wcB, consulted 29-06-2021
** https://www.who.int/news-room/feature-stories/detail/the-j-j-covid-19-vaccine-what-you-need-to-know?gclid=CjOKCQjw_dWGBhDAARIsAMcYuJyOSptG_NIAYChDKf7vyqgNV3NEuKqH6GFvUg96YLm1tYqphZUweVYaAvLUEALw_wcB, consulted 29-06-2021
*** https://www.who.int/news-room/feature-stories/detail/the-moderna-covid-19-mrna-1273-vaccine-what-you-need-to-know?gclid=CjwKCAjwieuGBhAsEiwA1Ly_nc9ymLAmAQ-66XmLkP65KWM5yt9nBZblQQ5-grCOXVWtDGStPSkKeRoCUJ8QAvD_BwE, consulted 29-06-2021
**** https://www.who.int/news-room/feature-stories/detail/the-oxford-astrazeneca-covid-19-vaccine-what-you-need-to-know?gclid=CjwKCAjwieuGBhAsEiwA1Ly_nW2flSHzkRC3x5ly03nVxjc4pxrcKxyNRJtd8xJIrp4xK_aoSUm5dRoC9QEQAvD_BwE, consulted 29-06-2021

9.3. Results for effect of different vaccines

It is obvious that there will be less infections when more people are vaccinated. But how much does it differ? In Figure 9.4 one can find the difference between the SVEIRD- and the SEIRD-models of Austria for Pfizer and Johnson. We choose to plot the infectives and exposed for Johnson and Pfizer since Moderna and Astrazeneca have efficacy rates which are like those of Pfizer and Johnson.

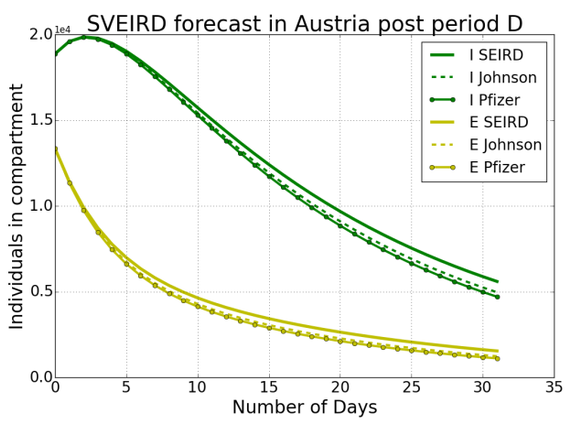


Figure 9.4: SVEIRD- and SEIRD-models post period D plotted in one figure for Denmark. The vaccines plotted are Pfizer and Johnson.

Since it is hard to see the exact difference from the plots, and because the other countries are not plotted, one can find the difference on day 30 between the amount of $Exposed$ and $Infected$ individuals from the SEIRD- and SVEIRD-model in Table 9.3 for both vaccines for all countries. Next to the exact amount one can find the percentage of the amount of $Exposed$ and $Infective$ decrease with the corresponding vaccine after 30 days.

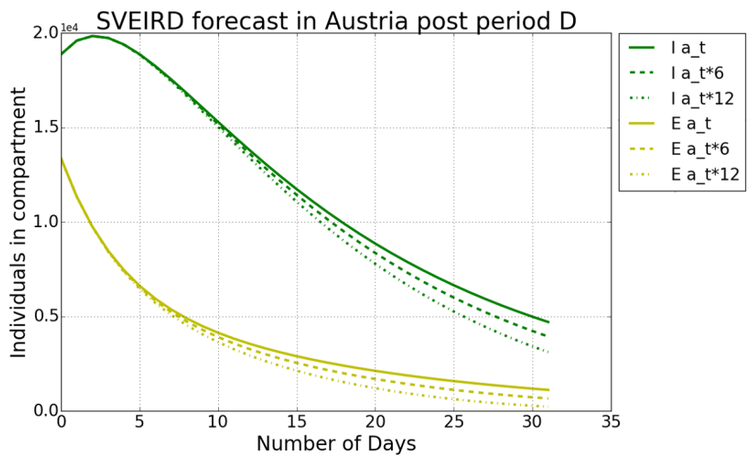
Country	E difference Pfizer	I difference Pfizer	E difference Johnson	I difference Johnson
Denmark	40.25 – % 27.26	118.90 – % 2.41	28.05 – %19.00	82.73 – %1.67
Austria	432.35 – % 27.95	892.58 – % 15.94	308.656 – % 19.95	630.70 – %11.26
Germany	1865.21 – % 11.41	3577.81 – % 4.33	1305.63 – %7.99	2495.95 – %3.02
Italy	23.38 – % 6.182	62.35 – % 0.03	16.24 – %4.29	43.31 – %0.02
Poland	21.86 – % 7.15	64.808 – % 0.03	15.19 – %4.97	45.0 – %0.02

Table 9.3: Differences in amount of individuals assigned to $Exposed$ and $Infective$ when different vaccines are used

We find that the difference between the SEIRD- and SVEIRD-model is not very big. This could be due to the fact that we have only taken fully vaccinated individuals into account when calculating the vaccination rate α_t. Since there is evidence that you are also (partly) protected if you have only had one shot, this rate might be higher then we think. It could also be possible that countries with a higher. What can also cause an unrealistic vaccination rate, is that we use the average vaccinations of period D to calculate with, while in reality more people are being vaccinated when we look at more recent vaccination numbers. (Appendix D)

9.4. Results for different vaccination rates

In Section 9.3 we concluded that the vaccination rate may be unrealistic. In this section we will adjust the vaccination rates by multiplying the original values for α_t by 6 and 12 and plotting the results in the same figure for Denmark to show what the possible forecasts are for higher vaccination rates. In this section we will assume that everyone is going to be vaccinated with Pfizer ($\sigma_v = 0.05$).



(a) SVEIRD-model of Denmark post period D

Figure 9.5: SVEIRD-model for different vaccination rates.

It makes sense that a higher vaccination rate results in less people being assigned to $Infective$ and $Exposed$, which is also visible in Figure 9.5.

9.5. Applications

9.5.1. Different α_t

Until now, we have based the vaccination rate on the amount of people being vaccinated in the past. Another option is to make an estimate of the vaccination rate by looking look into the future. The vaccination rate can increase if there are more vaccination stations, vaccines and vaccination workers. For example, if authorities know that there will become more vaccines available, α_t can be set higher at $t+30$. If a time series α_t is made which depends on all these factors, one can make an accurate forecast with the SVEIRD-model.

9.5.2. Different σ_v

Another application could be to give weights to the sorts of vaccinations being shot in a country. What if 50% of the inhabitants (who actually get the vaccination) receive Pfizer, 25% Moderna and 25% Astrazeneca? One can set the inefficacy rate as follows and create forecasts with it.

$$\sigma_{overall} = 0.5 \cdot \sigma_{Pfizer} + 0.25 \cdot \sigma_{Moderna} + 0.25 \cdot \sigma_{Astrazeneca} \tag{9.7}$$

9.6. What's next?

We have created a compartment model which takes vaccinated individuals into account and the results are what we expected. Of course, when more individuals get vaccinated, there will be less infections. And when individuals get vaccinated with a vaccine which has a higher efficacy rate, there will also be less infections. We now want to combine these findings with the queue we have created earlier to investigate whether we can make recommendations for the government.

10

Queue if vaccinations are available

The model for the queue in this chapter works exactly the same as described in Chapters 7 and 8, but for the queue in this chapter we use the amount of people assigned to $Exposed$ and $Infective$ from the SVEIRD-model instead of the SEIRD-model.

First we are going to show the results for the standard queue, after that the results for the amplified queue. In Figure 10.1 one can find the contact tracing queues in Denmark for different vaccination rates. It is obvious that when more individuals get vaccinated, less individuals are infected and exposed which results in less flow in.

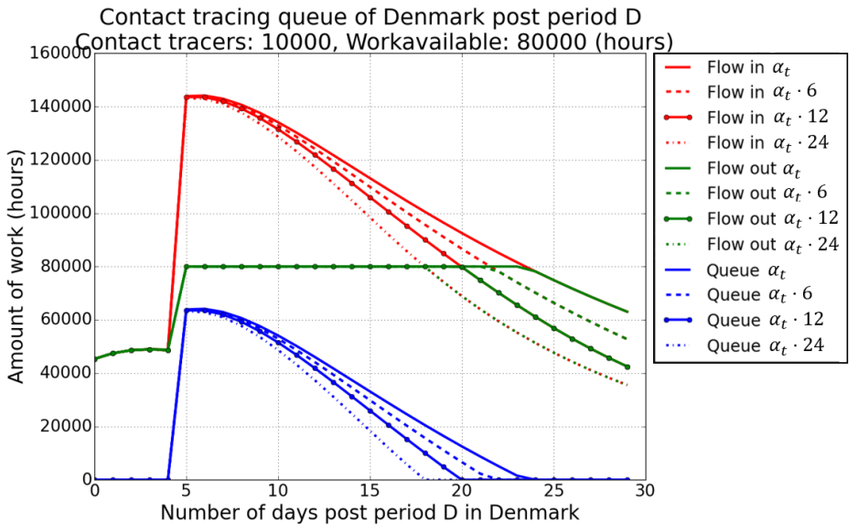


Figure 10.1: Contact tracing queue with different vaccination rates.

To conclude this chapter, one can find the SVEIRD-model forecasts for Denmark post period D if the amount of exposed individuals are dependent on the length of the contact tracing queue. It make sense that when more individuals get vaccinated, less individuals get exposed and infected which implies a shorter queue (as became clear from Figure 10.1). This also implies a less amplified queue which eventually can result in much less people assigned to $Exposed$ and $Infective$.

To illustrate the difference in contact tracing queue for a vaccination rate equal to α_t and $\alpha_t \cdot 12$, one can find their plots in Figure 10.3.

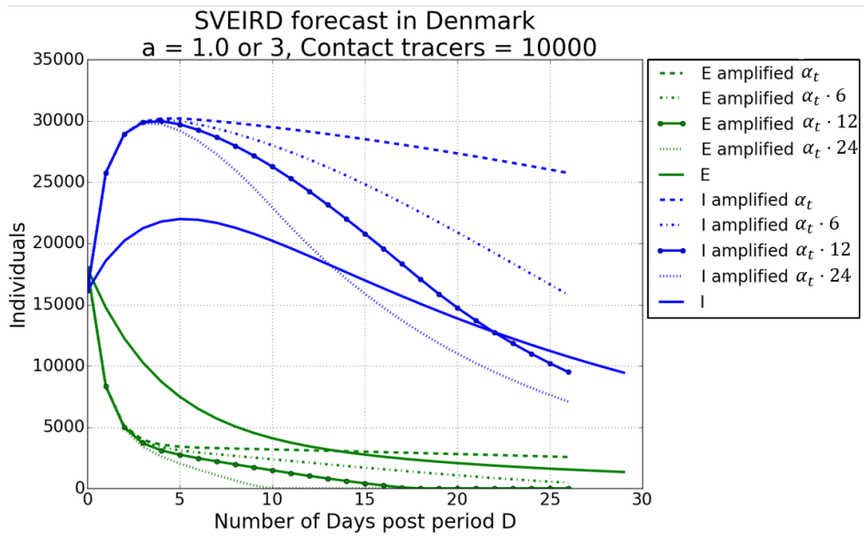
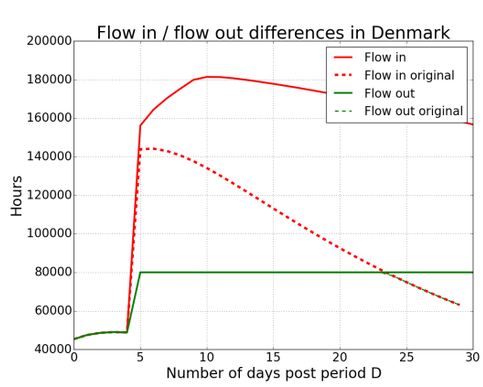
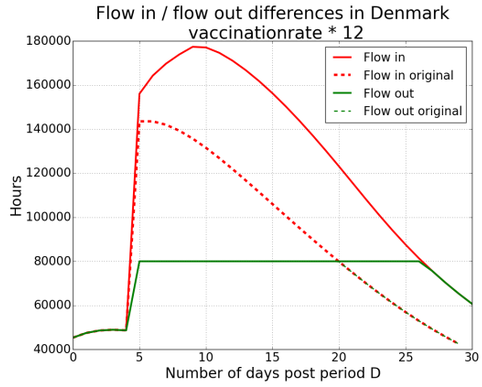


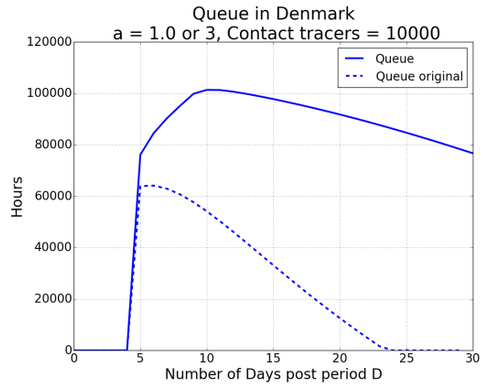
Figure 10.2: Amplified contact tracing queue with different vaccination rates.



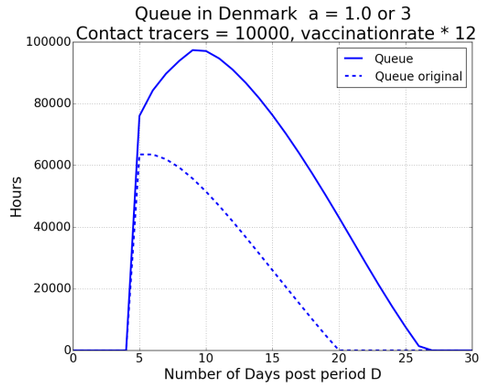
(a) Flow in and flow out of the amplified queue, vaccination rate α_t



(b) Flow in and flow out of the amplified queue, vaccination rate $\alpha_t \cdot 12$



(c) Amplified queue, vaccination rate α_t



(d) Amplified queue, vaccination rate $\alpha_t \cdot 12$

Figure 10.3: Flow in, flow out and actual queue (described in Chapter 8) for two different vaccination rates.

11

Conclusion and recommendations

Conclusion

By the first part of this report we can conclude is that it is possible to adjust compartment models in such a way it is applicable to a specific virus. The characteristics of COVID-19 can be implemented by adding compartments $Exposed$, $Deceased$ and $Vaccinated$ to the classical compartment models. These are important since the incubation time, mortality rate and vaccination rate can make a significant difference in forecasting the dynamics of COVID-19. The behaviour of the inhabitants can be implemented by turning the bi linear term βIS in a non bi linear term by raising I to the power p. This power makes sure that there is no homogeneous spread of people, this simulates social distancing and quarantine.

Secondly, one can simulate a contact tracing queue using the compartment model mentioned above. To keep the pandemic under control, we need contact tracers to investigate infected individuals and their contacts. The amount of contact tracers needed can be estimated using the compartment models. One can use the estimation of the amount of $Exposed$ and $Infectives$ to create a forecast contact tracing queue. We then found that it is possible to simulate the amplification of the queue if there are not enough (possibly) infected individuals investigated.

Finally, it is possible to simulate the dynamics of the virus if vaccinations become available with the SVEIRD-model. One can use the model to investigate what percentage of the inhabitants has to be vaccinated with a specific vaccination. This can be useful when vaccinations have a different efficacy rate.

Recommendations

The behaviour of the inhabitants is changing a lot throughout the entire pandemic, this indicates that the p-value in the SVEIRD-model is not representative for the different lock downs the countries have been through. This means that it might be valuable to recalculate these frequently and not take the mean of them, as we have done in this research.

When simulating the queue, the values of the most parameters were educated guesses because their values were not available. When these models would be used in real life, one has to investigate the values of these parameters to get a realistic model. For example, when the queue is created, we took guesses for the values for θ_E and θ_I, while these values are very important for the estimation of the queue. Also, when the queue is amplified it is important that the parameters are realistic. In Chapter 8 we set the amplification factor at 3 or 4 to show a clear difference for the situations, but it could be that in reality these factors have a much lower or higher value. What is also important for the simulation of the queue, that authorities decide what percentage of the individuals they want to have contact traced, since it might not be realistic to have a full coverage as goal. This decision can then be implemented in the queue simulation, this way we can still have a proper investigation done, while not everyone who has been in contact with infected individuals has to be contact traced. This can be very helpful, especially in times of an infection peak and there are not contact tracers available.

All models in this report are used to make a forecast of the virus for thirty or sometimes fifty days. We decided to do that in this report to emphasise our findings. In reality, we know that COVID-19 is very unpredictable and can most of the time not be estimated so far in the future. We would recommend to forecast only 15 days in advance, especially in unstable times.

Bibliography

[1] Erin C.S. Acquesta, Walt Beyeler, Pat Finley, Katherine Klise, Monear Makvandi, and Emma Stanislawski. Modeling resource demands and constraints for covid-19 intervention strategies. siam news, 2020.

[2] Dilip K. Bagal, Arati Rath, Abhishek Barua, and Dulu Patnaik. Estimating the parameters of susceptible-infected-recovered model of covid-19 cases in india during lockdown periods. Chaos, Solitons & Fractals, 140:110154, 2020.

[3] Rabih Ghostine, Mohamad Gharamti, Sally Hassrouny, and Ibrahim Hoteit. An extended seir model with vaccination for forecasting the covid-19 pandemic in saudi arabia using an ensemble kalman filter. Mathematics, 9(6):636, 2021.

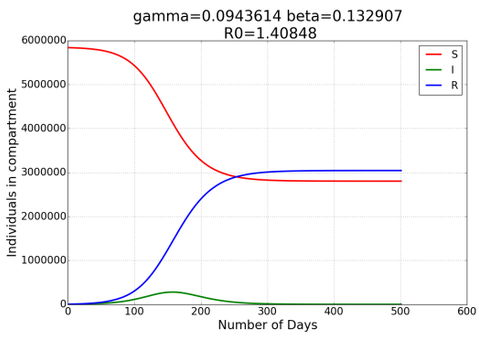
[4] Elena Loli Piccolomini and Fabiana Zama. Preliminary analysis of covid-19 spread in italy with an adaptive seird model. arXiv preprint arXiv:2003.09909, 2020.

[5] Manojit Roy and Mercedes Pascual. On representing network heterogeneities in the incidence rate of simple epidemic models. Ecological complexity, 3(1):80–90, 2006.

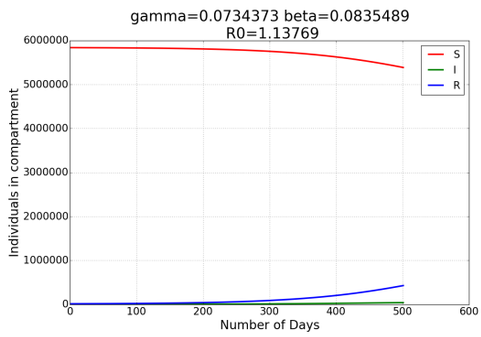
Appendices

A

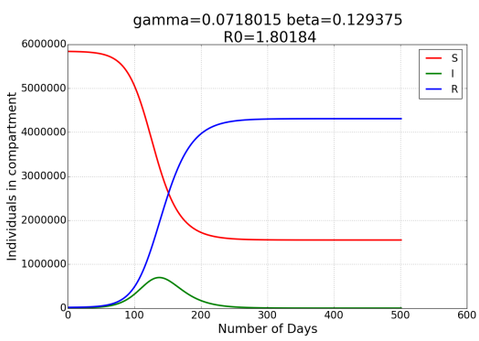
SIR-models for Denmark



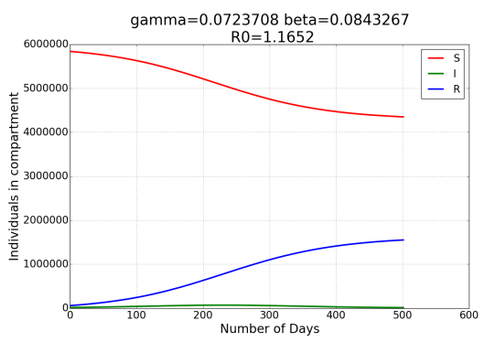
(a) SIR-model period 0



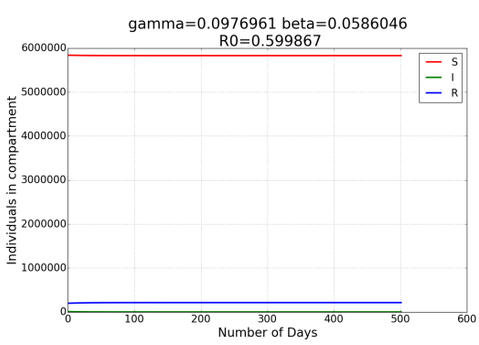
(b) SIR-model period 1



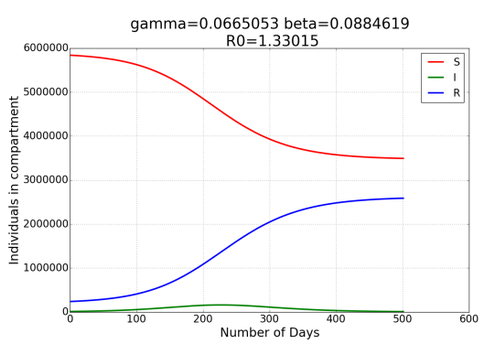
(c) SIR-model period 2



(d) SIR-model period 3



(e) SIR-model period 4



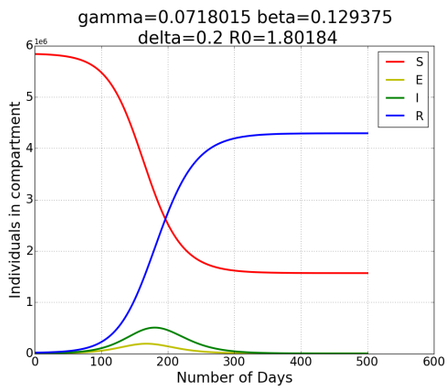
(f) SIR-model period 5

Figure A.1: SIR-models created using the data in the period mentioned in the sub caption.

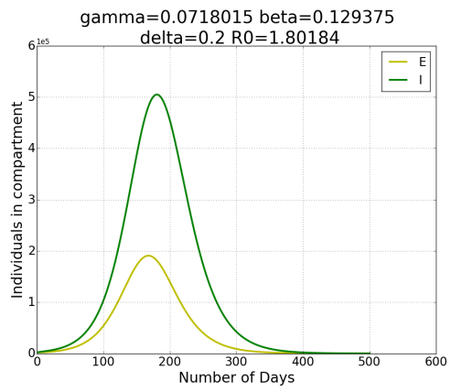
B

SEIR- and SEIRD-models for Denmark

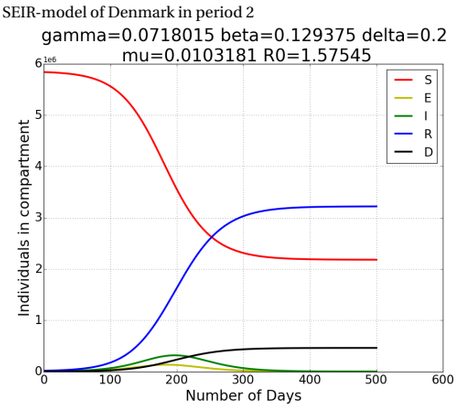
SEIR- and SEIRD-models for Denmark post period 2.



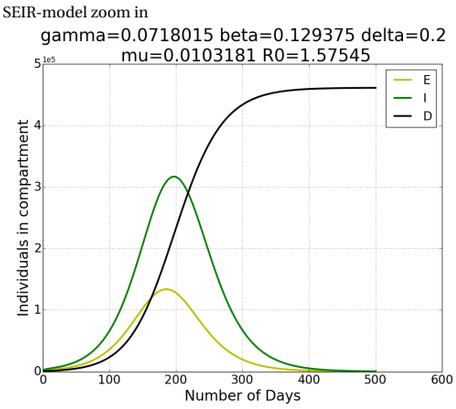
(a) SEIR-model of Denmark in period 2



(b) SEIR-model zoom in



(c) SEIRD-model of Denmark in period 2



(d) SEIRD-model zoom in

Figure B.1: SEIR- and SEIRD-models for Denmark in period 2, next to them their corresponding figures for E, I and D separately for a more detailed representation

C

Infected individuals

Number of people infected by COVID-19 in Austria, Germany, Italy and Poland. The periods chosen to use data from for the forecast using the SEIRD-model are highlighted and labelled with A, B and C.

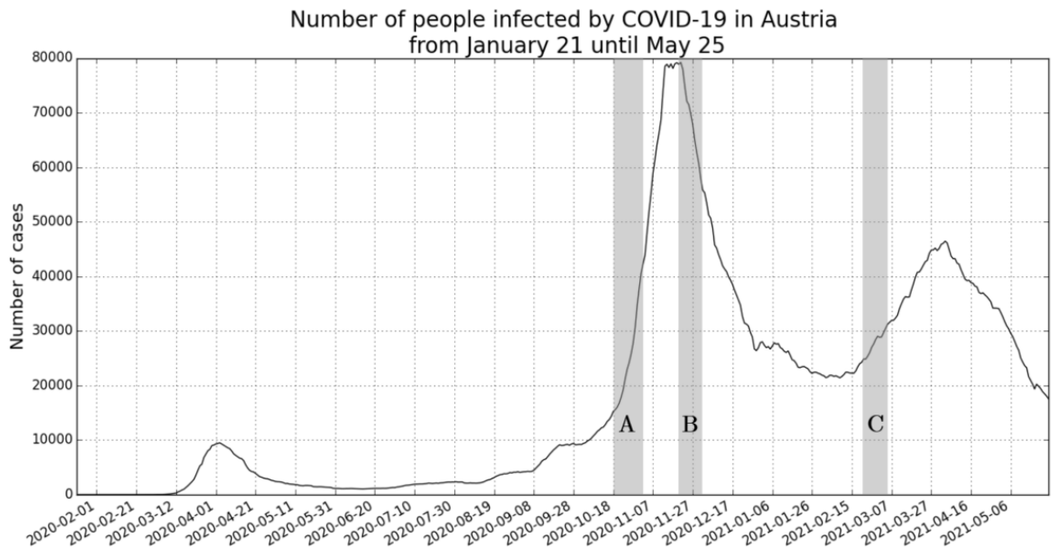


Figure C.1: Infections in Austria

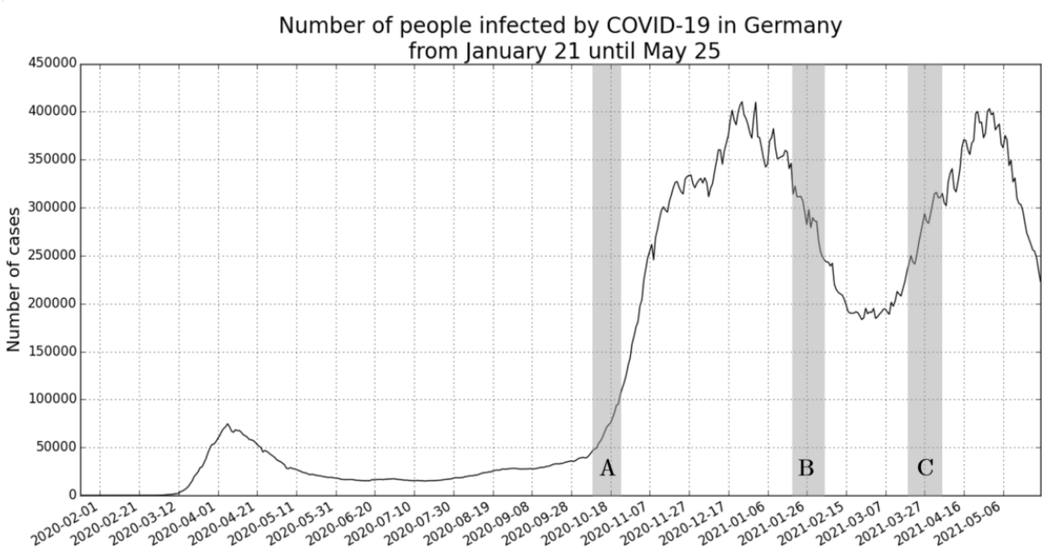


Figure C.2: Infections in Germany

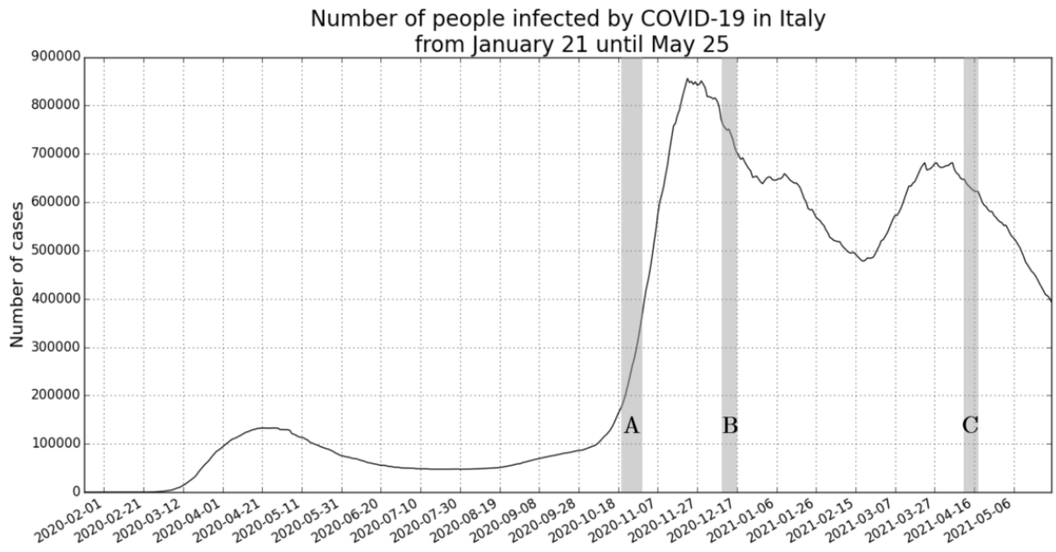


Figure C.3: Infections in Italy

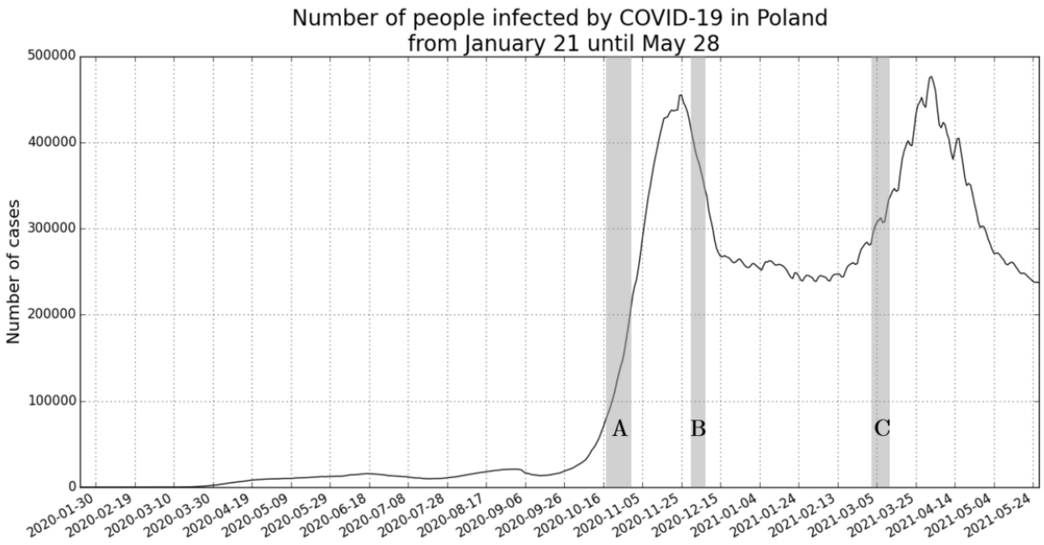


Figure C.4: Infections in Poland

D

Fully vaccinated individuals per country

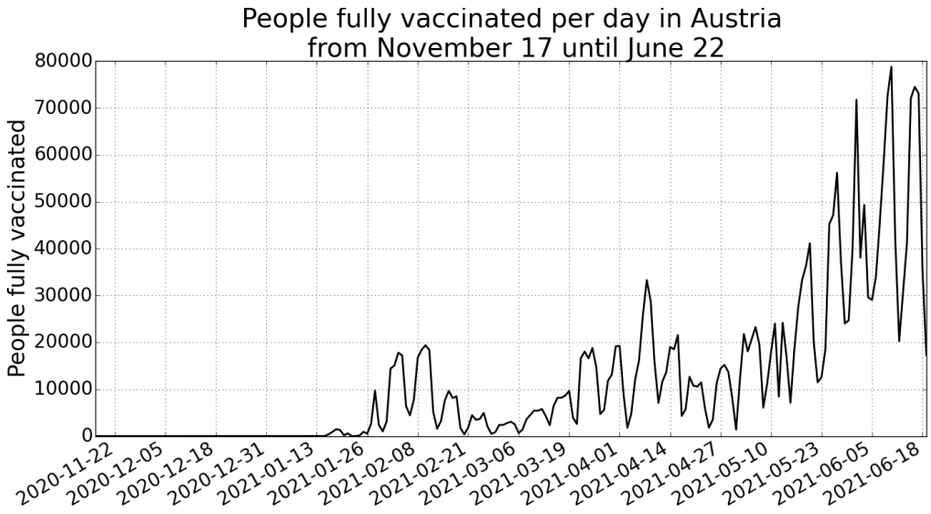


Figure D.1: Fully vaccinated individuals in Austria

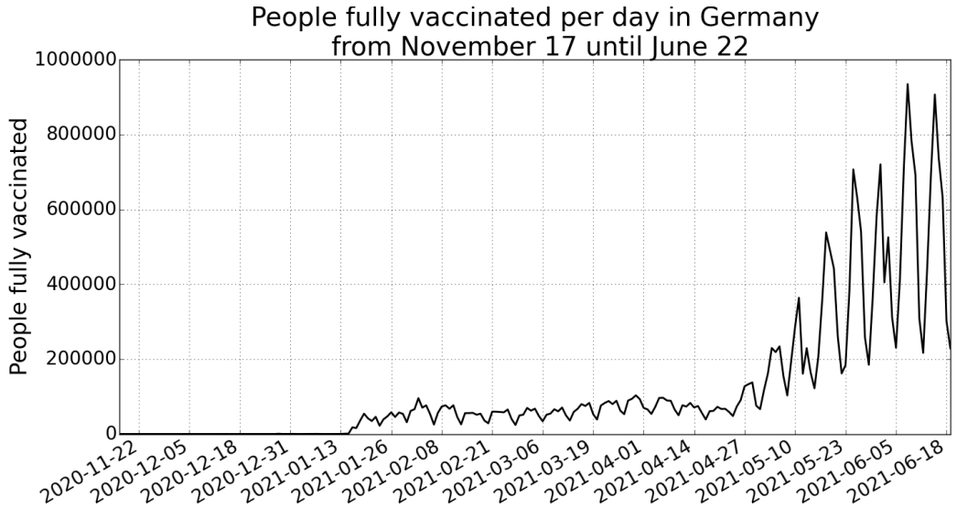


Figure D.2: Fully vaccinated individuals in Germany

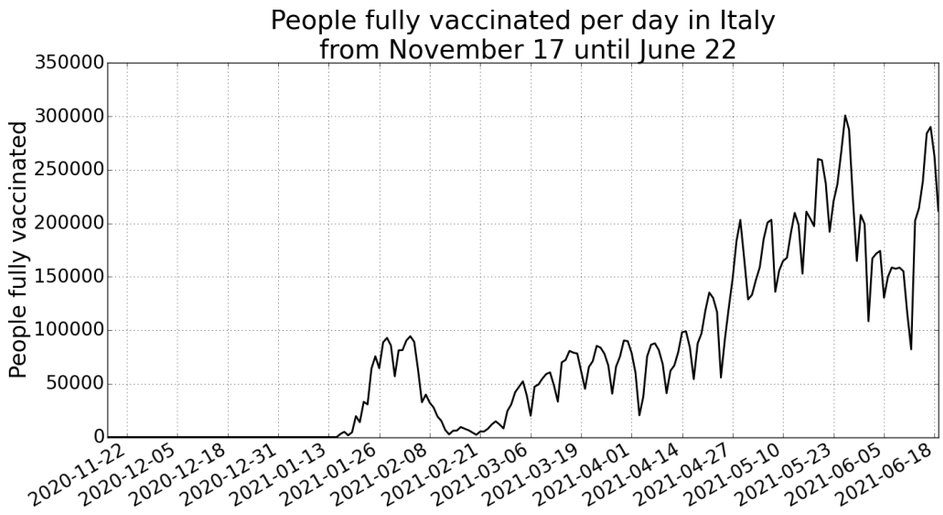


Figure D.3: Fully vaccinated individuals in Italy

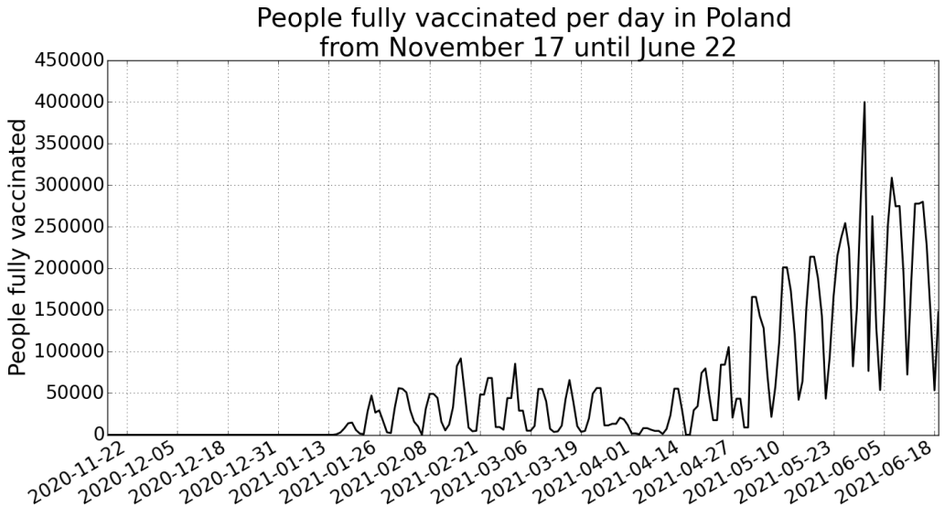


Figure D.4: Fully vaccinated individuals in Poland