02443 Stochastic Simulation

Simulation of Epidemics

June 25th, 2020

Troels Bagge Angelo	s153564
Christoffer Rode	s154323
Theis Ferré Hjortkjær	s163700
Leah Rahbek	s163718
Mette Bramsløw	s163721

DTU Compute
Department of Applied Mathematics and
Computer Science

Contents

1 Introduction

The topic of this project is modeling infectious diseases and their transmission within populations. It is inspired by the current COVID-19 pandemic, which has affected societies all over the world. Considering the different actions that has been taken to control this pandemic on a local and a global level, it has become evident that models are very helpful tools that can be used to issue guidelines and regulations for a population as well as for legislation for public health safety, but has also been a cause of discussion due to the various ideas of how to best build a model that resembles the real-life scenario of the specific epidemic.

1.1 SIR models

Compartmental models are used in epidemiology to describe and analyze epidemics unfolding in populations. The goal is to predict how infectious diseases spread through populations, and to predict the amount of infected people over time or at a certain point in time. How it affects the progress of the disease in different scenarios.[1]

The wording of "compartment" in the compartmental models refers to the idea that all individuals in a population will fall into a compartment such as healthy or infected.

The SIR model is one of the simplest compartmental models. The model consists of three compartments:

- **S**: The number of **susceptible** individuals in a population. These individuals are not infected but they can become infected.
- **I**: The number of **infected** individuals in a population. These individuals have the infectious disease so they are able to transmit it to susceptible individuals.
- **R**: The number of **recovered** individuals in a population. These individuals are immune to the disease, meaning that they cannot become infected again.

The SIR model will often have to be remodelled in order to take into account other aspects of the disease, e.g. if the disease is deadly to the infected individuals in the population. Thus, in addition to the three compartments in the SIR model, the SIRD model consists of an additional compartment:

- **D**: The number of **deceased** individuals in a population.

[1] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

Other possibilities for remodelling the SIR model includes for instance the incorporation of a compartment for the number of exposed individuals in a population such that the individual is *infected* but not yet *infectious* (the SEIR model). It is also an option that when infected individuals recover from the infectious disease, they do not become immune to it, so instead they are susceptible to the disease again (the SIS model).[2]

1.2 Project description

In this project, the dynamics of an epidemic will be investigated through stochastic simulation. The incentive of this project is to simulate infectious diseases under different conditions and settings, thus understanding some of the actions that can be taken to control an epidemic.

First, we will consider the differential equation model for an epidemic. We will solve the differential equations in order to demonstrate the development of an infectious disease. We will then simulate an epidemic by stochastic modelling by implementing a population with simulated infectious events. Lastly, we will test our stochastic simulation model under different conditions, e.g. simulating different events that either promotes or prevents the spread of disease such as travelling and quarantine.

2 Differential equation model

The SIR models are often made deterministic by using a system of differential equations with parameter values. This is in contrast to stochastic modelling, where the models are based on random movement, distributions for exposure- and recovery time, and so on.

The deterministic approach using a differential equation model is often useful when working with infinite populations, or with very large population numbers.

The compartmental model can be described using a system of differential equations together with the above mentioned terms for susceptible, infected, recovered, and deceased individuals, respectively, i.e. if working with a SIRD model. Assuming a population number of $N = S + I + R + D$, it is trivial that:

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

[2] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

where β is the infection rate, γ is the recovery rate, and μ is the death rate.

As the dynamics of an epidemic often is much faster than the dynamics of birth and dead in a population, it is fair to assume that the population number is constant, e.g. $N = 1000$ (disregarding e.g. quarantine and travelling between populations, etc). In figure 1, it is seen how an infectious disease may spread through a population at certain parameter values for β, γ, and μ.

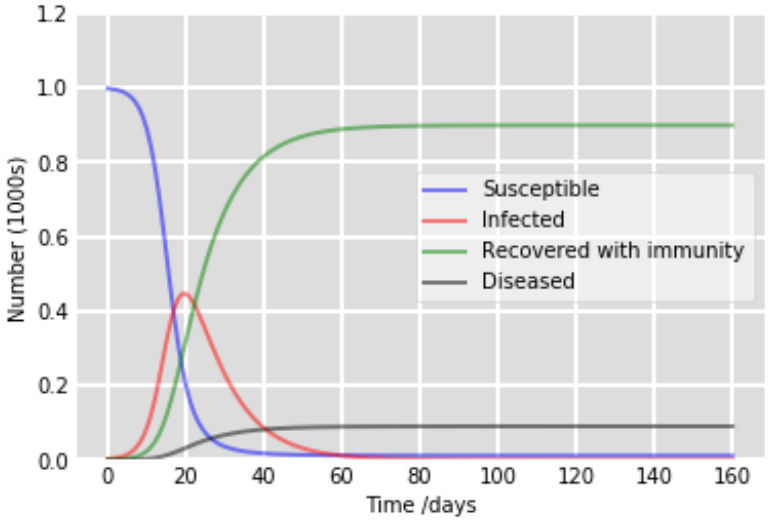


Figure 1: Plot of solution to a system of differential equations as stated by the SIRD model with parameter values of $\beta = 0.5$, $\gamma = 0.1$, and $\mu = 0.01$. The population number is $N = 1000$, and the number of infected individuals at initialization is $I_0 = 2$.

While the stated model is a simplification of the extremely complex dynamics behind an epidemic, it is still a powerful tool to provide an overview of the spread of an infectious disease. Depending on the infectious disease itself (e.g. its mechanisms) together with other factors such as population density and treatment availability, the parameter values can be changed to approach a real-life situation. Altering the parameter values in the system of differential equations, it is evident that:

- β: A high infection rate means a fast spread of the disease through the population, and more individuals infected by the disease at the peak of the epidemic (see figure 2). Infection rate can be decreased by e.g. hygienic initiatives and social distancing.
- γ: A high recovery rate means a slower spread of the disease through the population, and less individuals infected by the disease at the peak of the epidemic (see figure 3). Recovery rate can be increased by e.g. treatment availability.
- μ: A high death rate means more individuals dying of the epidemic. Accordingly, it also means a slower spread of the disease through the population, and less individuals infected by the disease at the peak of the epidemic (see figure 4). Death rate can be decreased by e.g. treatment availability too.

Infection rate (β)

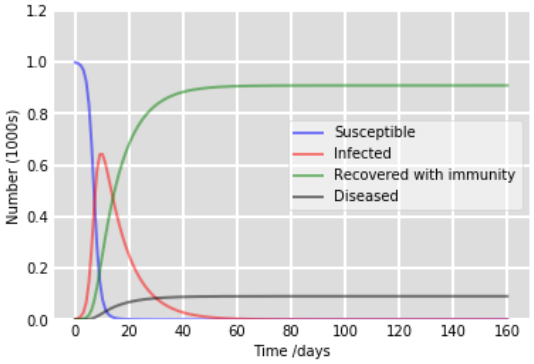


Figure 2: Plot of solution to a system of differential equations as stated by the SIRD model with an infection rate of $\beta = 1.0$. Other parameter values are as stated in figure 1.

Recovery rate (γ)

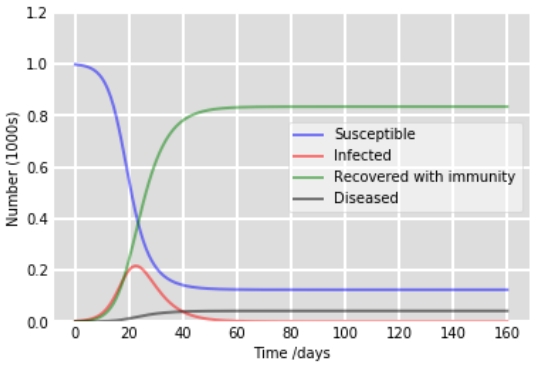


Figure 3: Plot of solution to a system of differential equations as stated by the SIRD model with a recovery rate of $\gamma = 0.2$. Other parameter values are as stated in figure 1.

Death rate (μ)

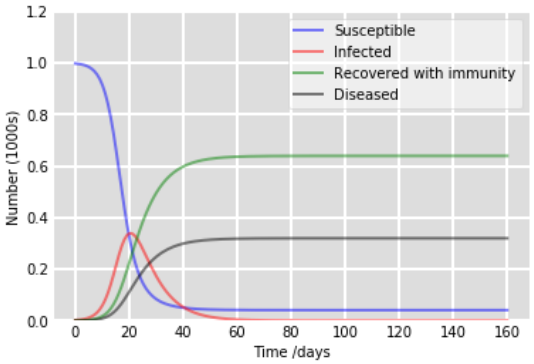


Figure 4: Plot of solution to a system of differential equations as stated by the SIRD model with a death rate of $\mu = 0.05$. Other parameter values are as stated in figure 1.

3 Stochastic simulation model

In this section, the method for simulating a pandemic through stochastic modelling is described. Stochastic modelling in this context is defined as using appropriate assumptions, random distributions and programming to simulate populations, movements and borders of a given system (e.g. communities, cities or countries).

The variables of the simulation program generated in this project will be tested for their influence on the model (see section 4), and later the simulation program will be used to explore a list of objectives (see section 5).

3.1 Simulating a population

One way of simulating a population is by defining people as coordinates in a coordinate system. A country could then be defined as a confined space in the system, such as a square in the 1st quadrant with borders from (0 , 0) to (x , y). At the beginning of a simulation, people are spawned uniformly at random within the boundaries of the system, see figure 5.

The movement of people is simulated as a random walk with the mean 0 and the variance as a parameter the user specifies. A more complex simulation would have a continuous time space allowing the individuals to move around continuously.

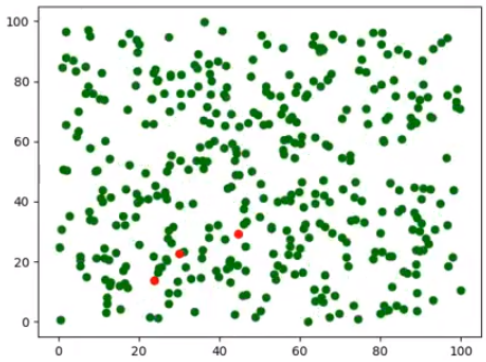


Figure 5: Movement within a population.

The result of this simulation is a population of people moving around within the coordinates from (0 , 0) to (x , y) as a function of time (arbitrary units).

3.2 Simulating infection in a population

To simulate infection, people are defined as ”susceptible”, ”infected”, ”recovered” or ”deceased”, as stated by the SIRD model. An infection event can take place if an infected

person pass within a certain distance to another susceptible person. In our model, the probability of infection is determined by a Bernoulli distribution, and probability increases as the Euclidean distance between the two persons decreases. At each time unit, the distance between all pairs of individuals is thus calculated, and if the distance is within a radius, a stochastic event is carried out to determine if the infection is successful.

Once people are infected, they may either become ”immune” or ”deceased”, depending on a variable ”age”. The ”age” variable is a random age given to each person in the system, with death being a probability given by age intervals. That is, there are ten intervals with each 10 life-years (0-9 years, 10-19 years, 20-29 years etc), where each age interval has a certain probability of death as each time-step passes. The system is designed so that disease progresses with a mean infection time, but infection time for each individual is randomly chosen, centered around the mean infection time, which is poison distributed. As disease progresses over time-steps through the population, if infection happens, there is a counter that decreases until an individual is declared as ”immune”. If at a random time step, the probability for death is hit, the individual is declared ”deceased” and the death is added to a deceased counter. After simulation over a given number of time-steps, the epidemic progress for a given number of individuals is plotted, revealing total death-rate for the population, death-rate by age and reproductive number.

3.3 Additional simulation implementations

Further implementations have been made to the simulation, which are described below.

3.3.1 Travelling

One of the important variables when dealing with pandemics is dealing with individuals travelling to other communities or countries. To explore the effects of this, multiple systems can be simulated simultaneously. Systems can have different parameters such as available space, population number, movement rates and so on. In this simulation, individuals travelling to other systems will have their original coordinate in their home systems removed and relocated in another systems for a given time. The likelihood of people travelling to another system is another stochastic variable that can be adjusted.

3.3.2 Hospitalization

Two other variables define hospitalization of patients; hospital beds and hospital factor. When hospital beds are available, death rates are reduced by the hospital factor. The higher the hospital capacity, the better the ability to handle highly infected populations. Adding these hospital parameters to our simulations are important in order to understand the effects of the ”flatten-the-curve” health strategy.

3.3.3 Quarantine

Lastly, another important parameter is quarantine. In our simulation, infected individuals only have a chance to go in quarantine, based on a Bernoulli distribution. This removes their coordinate for a fixed time period.

3.4 Computational requirements

The simulation of a pandemic requires a lot of computational efforts. Simulating movement of people as a random walk with distance as a metric for infection, is a indirect way of simulating the spread of a infectious disease throughout a population. A less computational demanding simulation could more closely resemble the SIRD model by using stochastic processes to determine the change of people in the different compartments using appropriate distributions.

3.4.1 Time Complexity

Assume only a single population is being simulated with a population size P. Using the notation from the SIRD model, we denote S as the susceptible, I as infected, R as immune and D as deceased. Using asymptotic notation, most of our the steps in the simulation is upper bounded by $O(P)$, as a constant amount of work is performed for each person in the population. The most time consuming step in our simulation however is when infected people infect others. This steps requires that the distance from each infected person to all people in the population is calculated. Therefore, this requires $O(I \cdot S \cdot d)$ time, where d is given by the dimension in the space which is 2. As d is constant this is removed in asymptotic notation and we end with a time complexity of $O(I \cdot S)$. This complexity corresponds well with our observations that initially the computational steps are fast, but as more people get infected our program severely slows down.

4 Applying the stochastic simulation model

In this section, the resulting model of the stochastic simulation is tested to show how different parameters can affect how an epidemic develops in a population. The results of these tests are described and analyzed here.

4.1 Reference simulation of fixed variables with plots

The benchmark simulation is done for an epidemic progressing over time steps $t = 400$. The default setting is:

- *Population size*, $N = 200$.

- *Shape*, the dimensions of the country people are allowed to move in. A default setting is a square of coordinates (0, 0 ; 100, 100).
- *Time*, $\mu = 400$. The amount of steps taken per simulation.
- *Movement* $= 5$. Movement defines the maximum distance people are allowed to move per time unit.
- *Infection distance* $= 2$. The radius of infection.
- *Recovery rate* i.e. the mean probability of recovering given by a Poisson distribution, $\gamma = 100$.
- *Death rate* i.e. the probability of dying per time step μ. Death rates are correlated with age intervals as shown in the following table:

Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90+
Death rate	0.00	0.0001	0.0005	0.0005	0.0008	0.001	0.001	0.004	0.004	0.006

The plots below shows a sample of plots of the stochastic simulation (see figure 6). The stochastic models have outputs that differ slightly amongst themselves from iteration to iteration. Overall, the plots of the epidemic progression over time are very similar to the discrete modelling described previously, but death rates vary as well as the age groups with most deceased when the simulation runs.

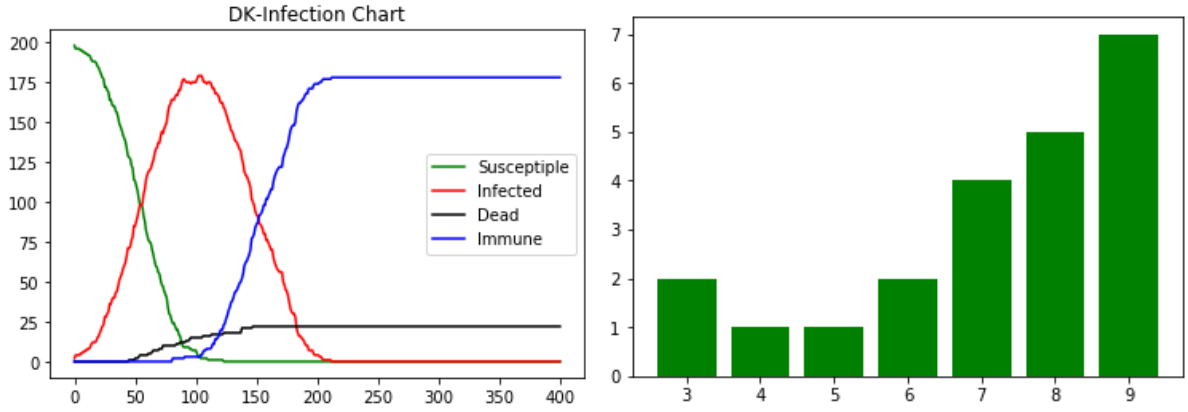


Figure 6: Reference simulation of an epidemic progressing with same starting values, ten replicates. **Left:** Example of one simulation iteration. **Right:** Histogram of dead individuals corresponding to the simulation iteration, sorted by age group.

To estimate how large the window is of expected dead people between each simulation, the same simulation was run 10 times and the amount of dead was recorded. On average, the amount of dead people is $\mu = 27.7$ with a confidence interval of $CI = (17.18, 38.22)$

using $\alpha = 0.05$, corresponding to $\mu = 27.7 \pm 10.52$.
Non-parametric bootstrap was then used to increase precision of the confidence interval. From the initial mean list, $N = 100000$ replications was done, where for each replication, a random number from the mean list was drawn and added to a new list of means. Ideally the simulation would have been done 10000 times instead of bootstrapping, but that would require too much computational power. A new mean, standard deviation and confidence interval was calculated from the new list of non-parametric bootstrapped means: $\mu = 27.69$, $s = 1.70$, $CI = (24.36, 31.02)$ using $\alpha = 0.05$. This means that $\mu = 27.69 \pm 3.33$.

4.2 Effects of the implemented variables

To understand the effect of using different variables, a seed(42) was set in the random generator. This makes sure that the same random number will be generated and the models therefore only are varied on basis of their variables.

4.2.1 Population size

The population size is a constant and is by default set to 200. The effect of adjusting the population size is shown in figure 7. The higher the population size, the higher the population density, and thus the higher the infection rate. This can refer to the difference between living in a city (where population density is high) and living on the countryside (where population density is lower). This is evident in figure 7. When population size is set to 100, the infection rate is gradually increasing over a longer time span. A small part of the population completely avoids infection. When the population is set to 500, all are infected within 50 time units.

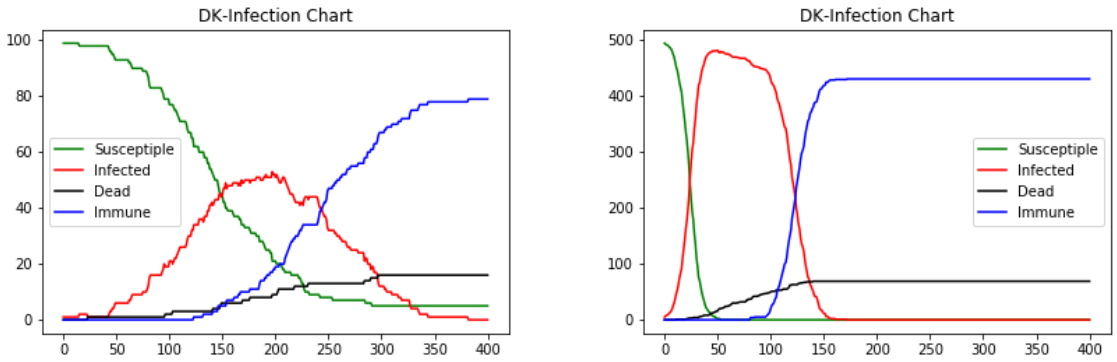


Figure 7: Charts for population sizes. **Left:** A population size of 100. **Right:** The population size of 500.

4.2.2 Movement

The variable ”movement” is based on a normal distribution. For each time-unit, every person will move from their position with a normal distribution with mean 0 and a variance that the user specifies. If the person is moving towards the border of the confined space, they will not be allowed to cross the border. This is done by rejection sampling. As seen in figure 8, the movement rate does effect the infection rates strongly when the movement is lowered. This is comparable to asking people stay at home and only interact very little with the environment around them. It is also seen that if there already is a high movement rate, the effect of increasing it even more is limited. This is a consequence of the random walk as the expected movement in both directions is 0. One interesting topic that could have been investigated is the consequences of only setting the elderly stay at home.

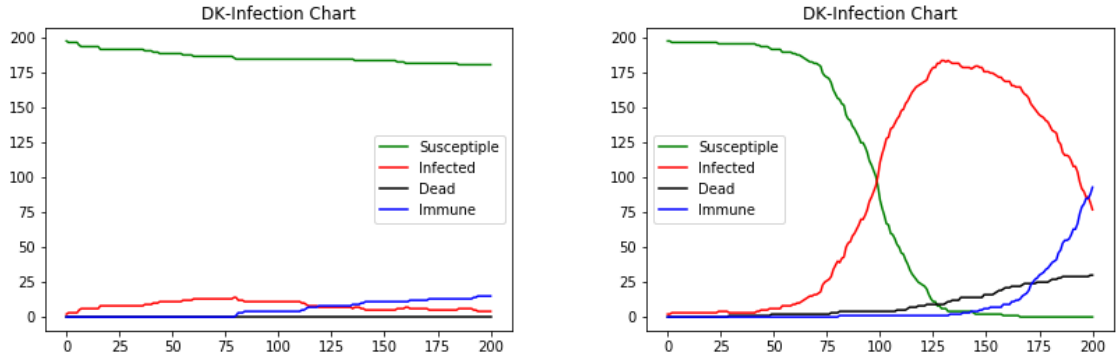


Figure 8: Charts for movement rates **Left:** The movement rate has a standard deviation of 1. **Right:** The movement rate has a standard deviation of 20. It seen that a low movement rate lowers rate of infection quite dramatic, where the increased movement rate does not effect infection chart very much except pushing the exponential growth a bit

4.2.3 Infection distance

Our model for describing the transmission of infection is based on distance. If a person is within a given distance of a person carrying the infection, the person has a probability of getting infected. By calculating the Euclidean distance from an infected person to every healthy person in the system, it is checked whether a healthy person is within infection distance. The closer in proximity the person is to an infected person, the more likely the person is to be infected. This proximity is linear in our model, where the probability of infection is 1 if the distance is 0.
The default infection distance is set to 2 distance units. This parameter can in reality be translated into how the disease spreads between people. If the disease spread through body fluids it should be set down. Alternatively, if it spreads as aerosols it should be increased. It can also give an idea of the effects of using masks and other protective gear

which influence the distance that aerosols moves when released from an infected person. When looking at figure 9 it is obvious how lowering infection distance affects infection rate. When the infection distance is set to 3 units, infection rate dramatically increases. This indicates that there is a very high sensitivity for the infection rate and that distance is an important metric for pandemics.

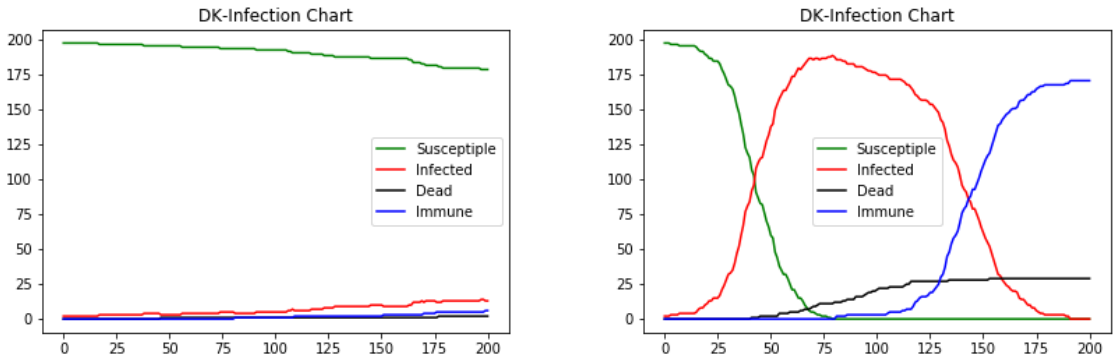


Figure 9: Charts for infection distances **Left:** The distance of infection set to 1. **Right:** The distance of infection set to 3. It seen that a low movement rate lowers rate of infection quite dramatic, where increase gives an earlier peak and sharper rise infection

4.2.4 Recovery rate

Infected people have two destinies; either they die or they survive through a recovery time and become immune. The recovery time is determined by a Poisson process, which provides a time for when the infected person will recover. The default setting for recovery time is λ=100. As shown on figure 10, when λ is lowered, a smaller part of the population dies and greater part becomes immune and there are even some who doesn't become infected.

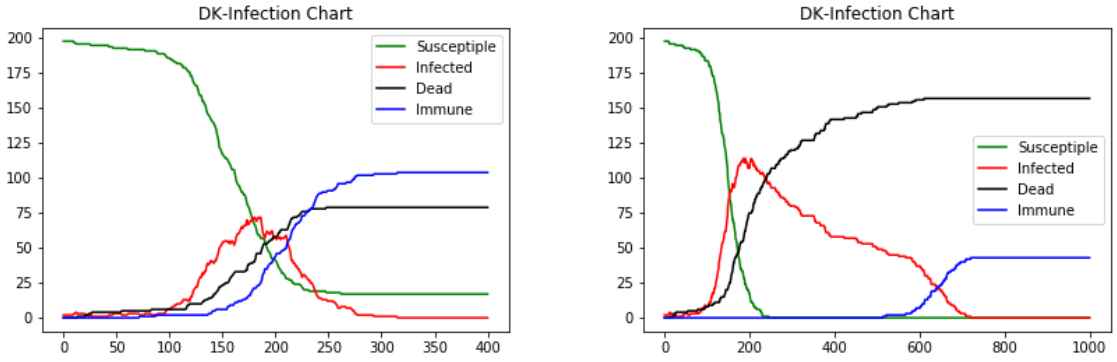


Figure 10: Effect of altering the recovery rate. The death rate has been increased ten times here compared to default. **Left:** The recovery rate is decreased to $\lambda = 50$ time units. **Right:** The recovery rate is increased to $\lambda = 500$ time units.

When λ is increased to 500, people die at a very high rate and very few people manage to recover and become immune. As a side-effect, everyone happens to get infected. Therefore recovery rate seems to be a very strong factor towards the outcome of the disease.

4.2.5 Death rate

Death rate is a variable that determines the probability for any given infected individual to die. For every step in time, the death rate decides the probability of whether the individual remains infected or the individual dies from the infection. The death rate is implemented as a variable that is dependent on the variable 'age', with low probability of dying for young individuals and significantly higher probability of dying for older individuals. The mean default probability per time-step for death rate across age groups is $\mu = 0.00179$.

In figure 11, the death rate is increased by a factor 10 and a factor 100, respectively. Changing the death rate affects not only the number of dead people but also the dynamics of other groups of the population. As the death rate increases, there is a slower spread of the disease through the population, and less individuals are infected by the disease of the peak of the epidemic. If the death rate is too high, i.e. if the infectious disease is too deadly, it will not spread through the population. This is due to individuals are almost instantly moved from the 'infected' compartment to the 'dead' compartment, where they are removed from the population and become unable to transmit the infection. This phenomenon that also can be seen in real disease cases.

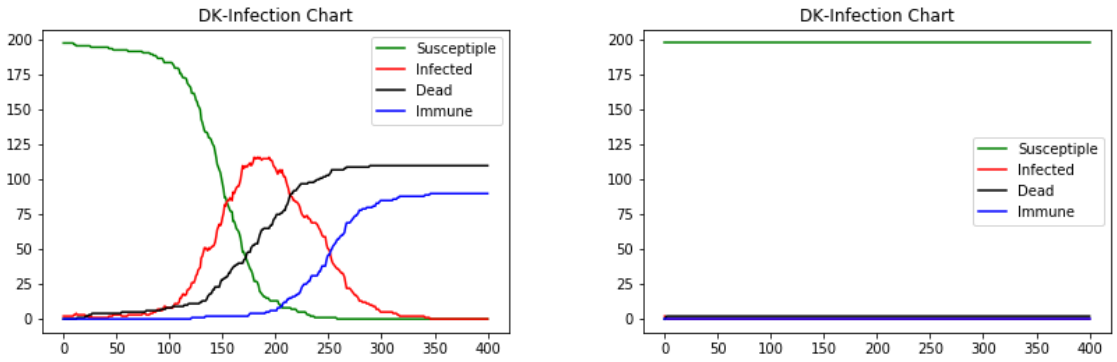


Figure 11: Changing the death rate. **Left:** The death rate is increased by a factor 10. **Right:** The death rate is increased by a factor 100.

4.2.6 Quarantine

A quarantine model is also added to the program. In the model, infected people draw from a Bernoulli distribution defined by a probability parameter p. If a success is drawn, the infected individual is taken out of the system for a certain time constant. In figure

12, the chance of being quarantined is changed. When chance is lowered (here shown as 1%), the effects of quarantine becomes insignificant. An obvious change is seen when the value is increased to 10%. Here, not even half the population become infected. This tool is also used in the current pandemic, where testing and quarantine has shown to be a strong tool.

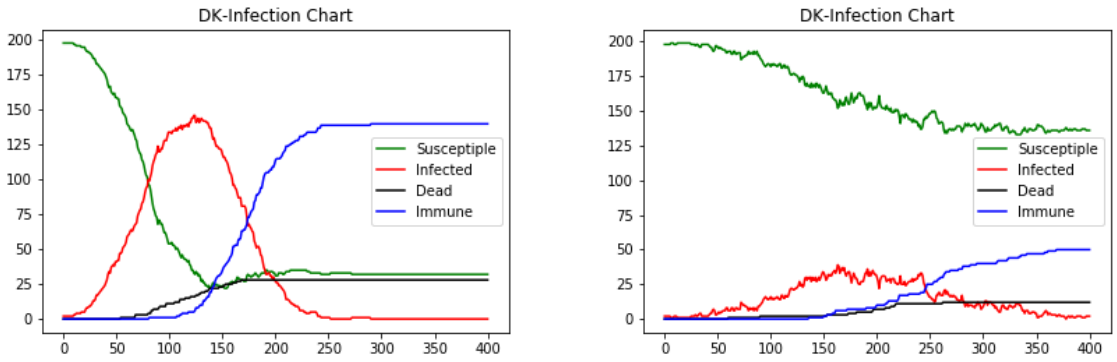


Figure 12: Models for changing quarantine probability **Left:** Here there is a quarantine probability at 0.01 **Right:** Here there is a quarantine probability of 0.1

4.2.7 Hospitalization

Another effect that has a strong influence on a disease is hospital capacity. When there is not enough space on hospitals people do not receive any treatment which increases the likelihood of dying. The model handles this by changing the death rates whenever a certain threshold of infected has been reached. One thing the model does not regulate now is movement of hospitalized individuals. This means they still move freely around. This could be possibly be improved later given more time. Figure 13 gives some different scenarios of hospitalization. It is seen that in all scenarios everyone is infected, but the amount of deceased varies quite a lot. When there is an increase in the amount of people who can be hospitalized and therefore don't decease, there is also a higher peak on the amount of infected. This is probably due to the fact that the hospitalized people are still moving around in the model, infecting more people and since they are not dying the probability of getting infected increases. When the mortality rate is not decreased to zero, the model begins to climb back to the look of the model without any hospitalization.

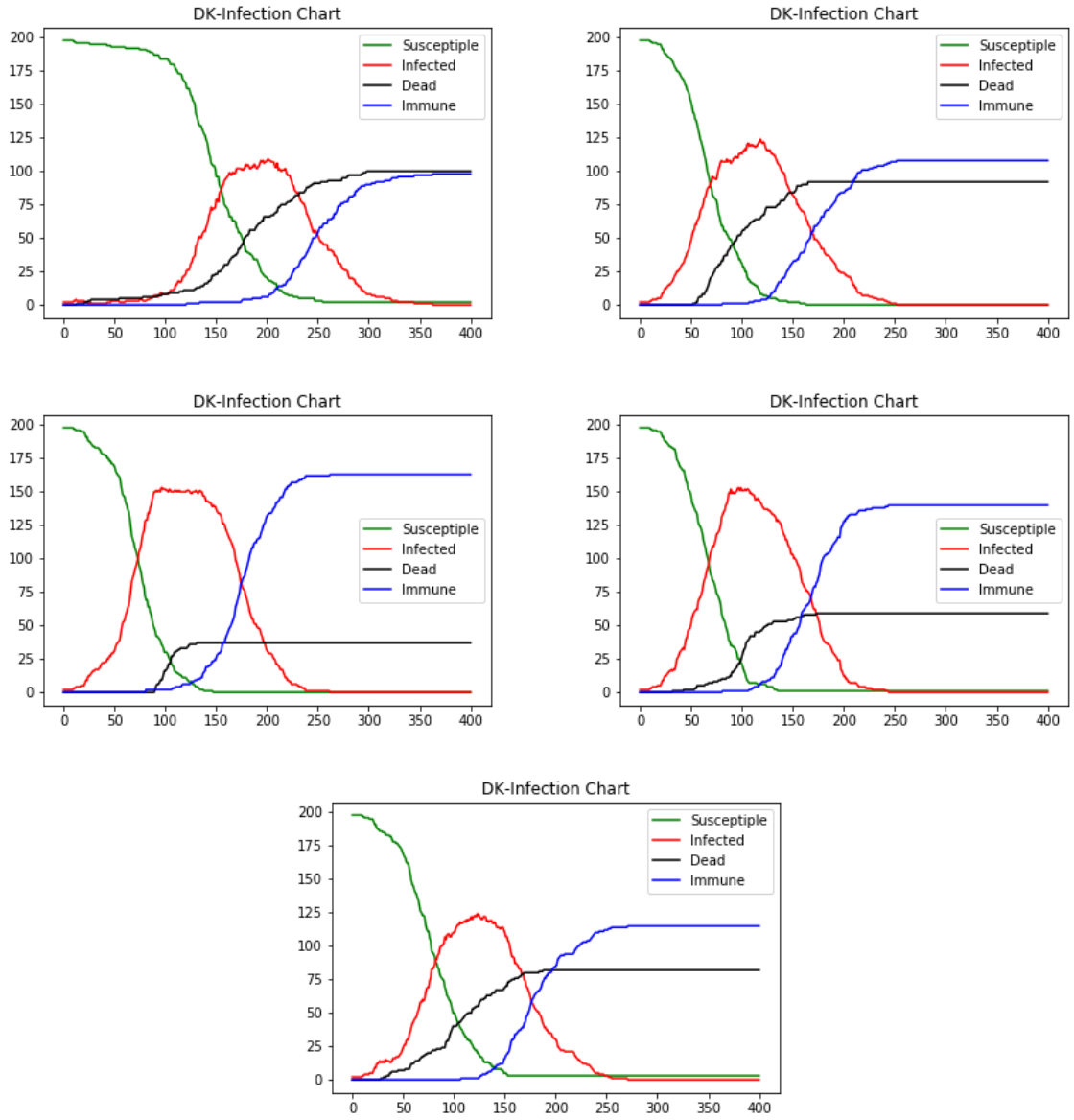


Figure 13: Models for changing the amount of hospital beds available. The death rate is increased by 10x from default. The mortality is decreased with 100% unless otherwise is stated. **Left top:** The model with out any beds. **Right top:** The model with 50 hospital beds. **Left middle:** The model with 150 hospital beds. **Right middle:** The model with 150 hospital beds, but with a decrease in mortality of 80%. **Bottom:** The model with 150 hospital beds, but with a decrease on mortality of 50%

5 Case studies for the stochastic simulation model

In this section, different cases are studied using the stochastic simulation model.

5.1 Killing 95% of the population

In order for the infectious disease to wipe out more than 95% of the population, two of the fundamental variables can be altered: the recovery rate and the death rate. As concluded in a previous section, dramatically increasing the death rate comes at the cost of killing infectious people before they can pass on the infection. Also, a too low recovery rate would mean that the infection not is spreading effectively through the population, as infected individuals are recovering too quickly. Thus, the objective is to alter the recovery rate and the death rate to become sufficiently high and low values, respectively.

In figure 14, a simulation with 96.5% deceased individuals have been attained by a recovery rate of $\gamma = 100000$ and a death rate of $\mu = 0.03$. Note that the death rate is the same in all age groups i.e. the probability of dying of the disease is independent of age. This is done to simplify the simulation, as it is simply regarded as a proof-of-concept simulation.

Performing 10 subsequent simulations, it is evident that the death rate is stable around 95%. For a population size of $N = 200$, following numbers of deceased people were obtained: $N_{\text{deceased}} = (193, 185, 187, 190, 195, 187, 191, 193, 189, 190)$. These are equivalent to a mean value of 190 deaths with a standard deviation of 2.97 (i.e. $95.0 \pm 1.5\%$).

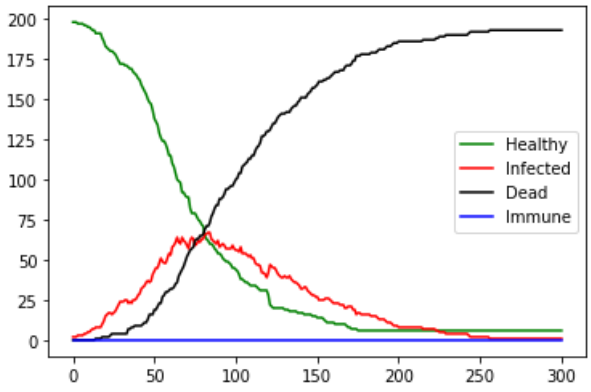


Figure 14: Killing 95% of the population. Default settings for the model have been used, but recovery rate and death rate has been altered to $\gamma = 100000$ and $\mu = 0.03$, respectively.

5.2 Effects of forced quarantine/lockdown

Amidst the COVID-19 situation, societies around the world enforced lockdown strategies to suppress infection rates. Lockdowns force people to behave responsible by staying home and avoid interactions with other people. Lockdown can reduce infection rate

and allow the health care system to keep up with infected people. One of the issues with lockdown strategies is that once the period of lockdown is over, the same old usual behavior continues and infections are resumed.

The effects of this can be simulated in our model in a simple manner. When a fraction of the whole society reach a threshold of infection, movement of people is frozen for a given time period.

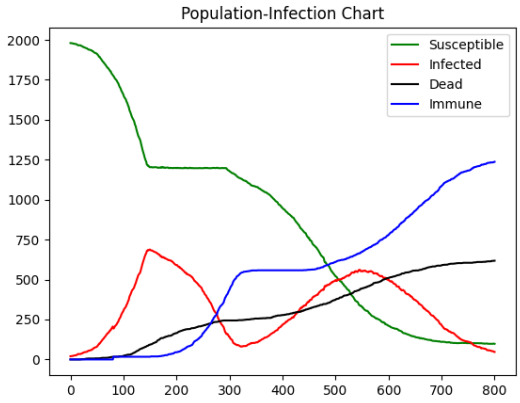


Figure 15: SIRD plot of a lockdown.

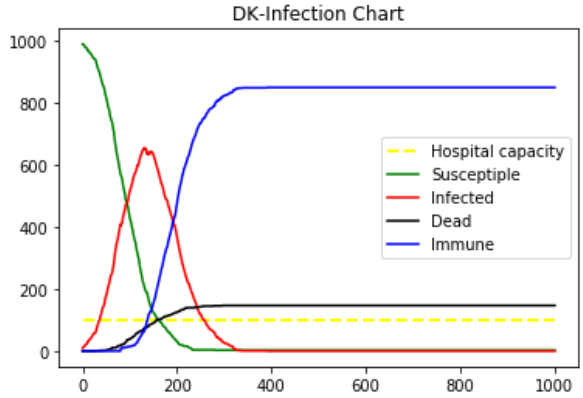
As shown in figure 15, a lockdown occurs at around time = 150 where 35% of the population is infected. The number of infected people reduce significantly during lockdown, because people are only able to infect those they are frozen within vicinity. Then, once lockdown period is over, infection resumes. This creates a cyclical behavior shown in figure 15.

In reality, not all individuals obey the law enforcement of staying home, and thus, there will still be some people moving around, including people with important work to take care of such as health care workers. Therefore another model could investigate the effects of lockdown where the population had a chance to disobey the lockdown.
In addition to figure 15, a .gif file (pandemic_lockdown.gif) is a attached in the appendix that shows how the lockdown is simulated in our program. This animation show how the lockdown actually affects the population and provides insights which can not be seen from figure 15 alone.

5.3 Flatten-the-curve effects

We will now study the effects of how the flattening-the-curve health strategy can reduce deaths in a society. In a real society, it is preferred to keep the infection rate low such that people are infected gradually. This allows the health care system to keep up with infected people. To reduce infection rate in a society, people can reduce interactions with each other.

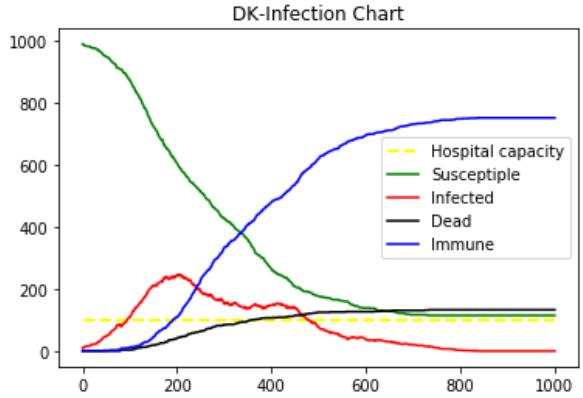
To simulate flatten-the-curve effects in our model, hospitalization was implemented. A limited amount of hospital beds is specified and the hospital factor was set to 0.5, meaning that death rates are reduced by 50% if infected people can get hospitalized. We can now adjust the movement variable; if movement is reduced, people interaction is also reduced, which reduces infection rate.



Death counts

149	159	133	164
132	128	159	138
158	133	141	140
140	149	153	139
146	152	163	141

Figure 16: Left: SIRD plot with movement=2. **Right:** Death count, 20 iterations. μ: 145.85.



Death counts

133	112	113	121
96	126	124	120
111	130	130	131
130	127	108	95
126	152	109	102

Figure 17: Left: SIRD plot with movement=1. **Right:** Death count, 20 iterations. μ: 119.8.

Under these conditions, the simulation shows that lowering movement result in slower and more gradual increase of infection. This allow more people to receive hospital treatment. Approximately 27 more lives are saved from reducing movement by half, i.e. the death rate is lowered by 2.7%. This is not a huge difference, and could be explained by the fact that the hospital capacity is still exceeded in many cases even though movement is set to 1.

Still, this simulation agrees with the effects of the flatten-the-curve health strategy. This model is a reasonable simulation of a real health care system, but it does not describe the full extent of it. In reality, health care is given to infected people that are diagnosed as critically vulnerable. A more authentic model could have used the variable ”age” to determine whether people need health care treatment or not.

5.4 Roskilde Festival 2020

In the wake of the corona-crisis in the summer of 2020, a major concern for many people, especially young people, was the cancellation of Roskilde festival 2020. But what would a scenario have looked like if Roskilde Festival 2020 had taken place?

Assuming obtaining max capacity of 130.000 people at the festival, the festival area is 2.500.000 m^2. This is equal to a mean population density of 0.052 (person per m^2). Any (sub)populations modelling Roskilde Festival should therefore have a population density close to this number. This is a number used for scale-down of population size, to obtain same population density and the real Roskilde scenario, thus making the model comparable to Roskilde.

In this simulation, it is assumed that there are three populations: the general living area, where population densities are normal but the movement higher (0.05 person per m^2), the concert area i.e orange stage, where movement is lower but population densities are higher (0.1 person per m^2), and finally, the silent and clean area, where the population density is normal and movement is lower (0.05 person per m^2). Another correction from earlier models is that the people participating will be between the age of 15-50. Therefore the generating data is biased towards including only people from the age of 15-50.

The recovery rate for COVID-19 is well above 100 hours (around 4 days), but to look at the epidemic in this closed system, we are assuming a mean recovery time of 4 days. Otherwise, everyone would get infected but the total probability of death occurring is higher if the mean infection time is longer than this. Infection distance is presumed to be standard, in a radius within 2 meters.

In summary, the assumptions are:

- Population size, for each population $n = 1000$ people for concert area, $n = 2000$ for each of the living areas
- Recovery rate, i.e. the mean probability of recovering given by a Poisson process, $\gamma = 100$ hours
- Time-steps = it takes around 200 hours for all of Roskilde Festival, so each-time step is 1 hour.
- Death rate, i.e. the probability of dying per time step, μ = [0: 0, 1: 0.0001, 2: 0.0005, 3: 0.0005, 4: 0.0008, 5: 0.001, 6: 0.001, 7: 0.004, 8: 0.004, 9: 0.006]
- Movement rate = 1 (concert area), 10 (living area, also accounting for transport time), 2 (silent and clean) (unit: meters per time step)
- Movement distribution = normal distribution

- Area for populations = 100x100 meters (concert area), 100x1000 meters (living area, silent and clean)
- Hospital capacity = 10 beds per population (as this is to show only how the epedemic would develop in Roskilde)
- No quarantine options
- Travelling between populations happens at random, with an average travel time of 5 hours

The resulting simulation can be seen below in figure 18.

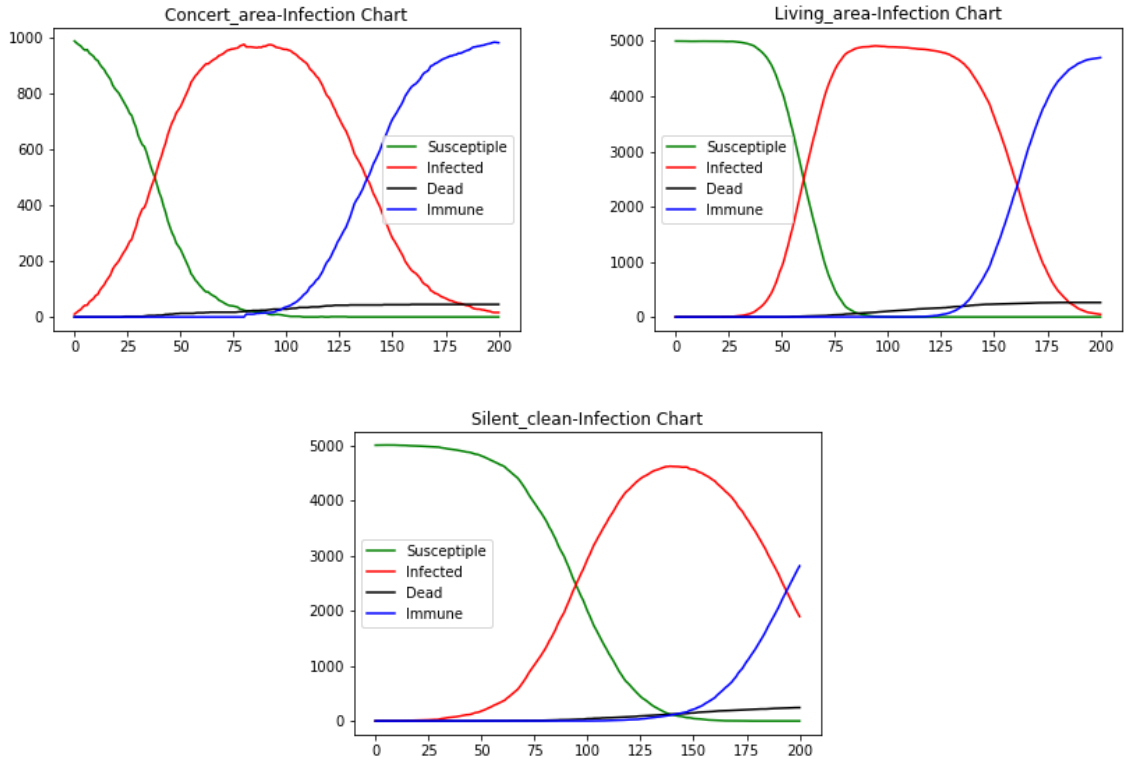


Figure 18: Roskilde 2020 Simulation. **Left:** Concert area with population density 0.10 person per m^2 **Right:** Living area with population density 0.05 person per m^2, **Bottom, center:** Silent and clean living area with population density 0.05 person per m^2

Figure 18 shows how easily COVID-19 could have spread in a Roskilde festival scenario. The higher population density, the higher the movement, the quicker the infection spreads. It can be seen by the bottom center plot that the progression of the epidemic is much slower in the silent and clean portions of the population than in both the concert area (left plot) and the living area (right plot), even with travelling between populations, because of the doubled density in the concert area and the increased movement of the living area compared to the silent and clean area. The curve is much more flat in the

living areas than in the concert area, because of the lowered population density. It can also be observed that in all populations, there is a low death rate because of the age distribution in our population (the populations only contain people age 15-50).

5.5 International pandemic

In order to model different infection rates realistically, it is also possible to include different countries. The infection charts in figure 19 were generated using parameters from the table below. The parameters does not resemble any real world situation, but it should give an overview of how interactions between countries of different sizes affect the whole world. It should be noted that the following parameters were identical across all countries: *Mean Recovery Rate=80*, *Travel Time=5* and *Death Rate* using the same numbers as the benchmark simulation.

	Population	Area	Movement	Hospital Factor	Hospital Beds	Quarantine Rate	Quarantine Duration
Belgium	400	62500	2	0.1	100	0	2
China	1500	160000	4	0.1	100	0.05	10
Denmark	100	10000	2	0.1	100	.002	5
Spain	500	40000	2	0.1	200	0.001	2
Estonia	20	10000	2	0.05	Inf	0	2
France	600	10000	2	0.1	100	0.02	5
Germany	200	10000	1	0.1	100	0.02	2
Italy	350	10000	2	0.01	60	0	2
Sweden	200	10000	10	0.2	100	0.01	2

In figure 19, some interesting interactions can be observed. Belgium is defined a large country and thus, only few infections happen. Furthermore, we see that the total number of people in Estonia increases over time. This is due to the fact that people travel to other countries, and in some cases they will stay within the country. The curve of China is quite flat. This is a result of the big area, along with a high quarantine rate for the infected people. When we aggregate all the numbers in figure 20, it is noted that a single peak of infected people is observed and afterwards the curve slowly flattens. For this specific simulation a quarter of the world population dies, which would be the result from a deadly pandemic that highly affects the world.

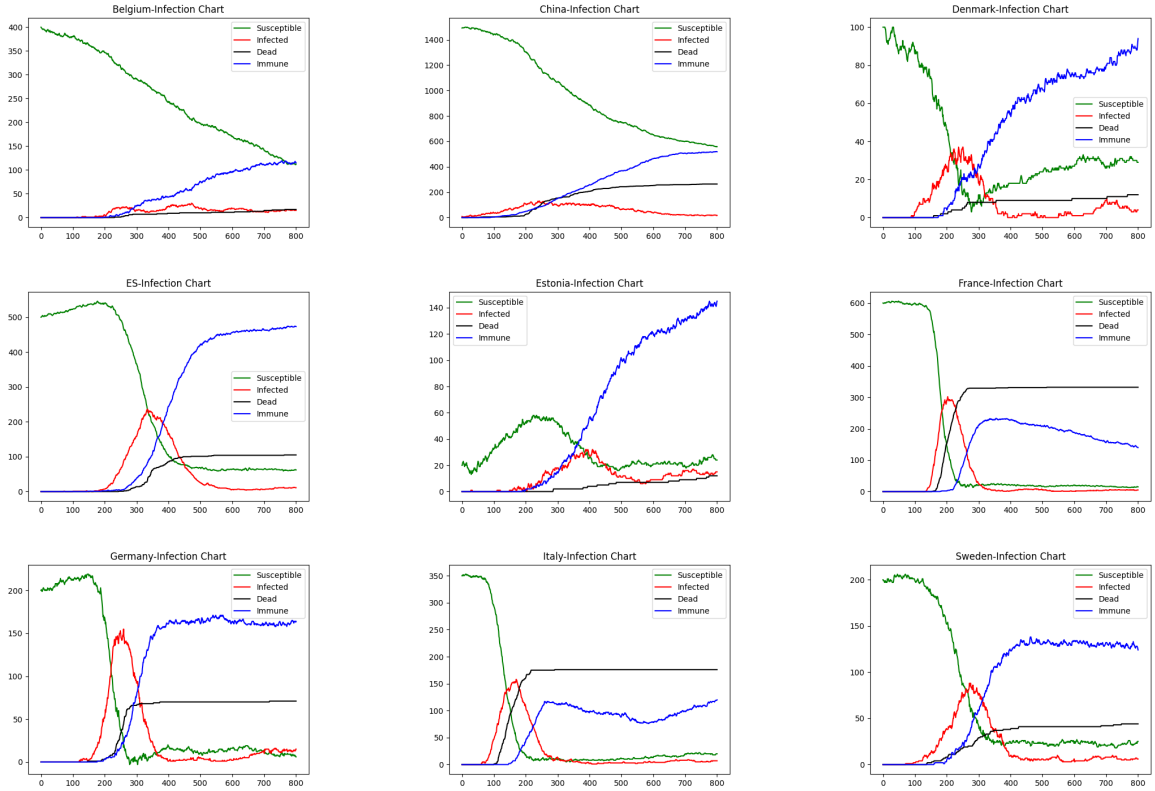


Figure 19: Infection charts for individual countries. **From top left to bottom right:** Belgium, China, Denmark, Spain, Estonia, France, Germany, Italy, and Sweden.

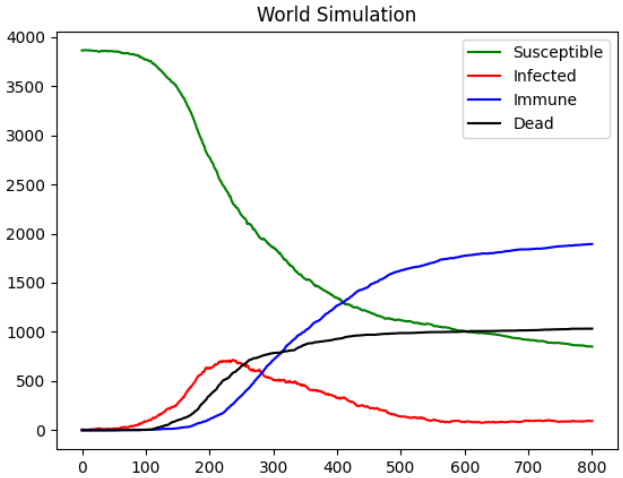


Figure 20: Aggregated statistics of world simulation.

6 Discussion

6.1 Discussions on the model

We have applied a seed in our model to ensure easier comparable outputs when changing for instance a rate, as seen throughout the report. We have mainly been doing one simulation with the set seed (42), which means that the models we have generated are based on a specific set of random numbers. If the model where to be tested in reality, it would be used without a seed, and used to generate multiple simulations as shown in the start of section 4.

The current model gives a simple idea of how an infection spreads during an interaction. It should be noted that with such a model, many assumptions needs to be considered, which make it increasingly complex. Currently the model uses normal distributed random walk in a squared space. This sort of random walk in a square can hardly be said to represent human movement. People tend to interact in groups and move along in patterns (think about the movement through a supermarket for example). People are also able to move on top and through each other. Patterns like these will hardly ever be completely random. This could be adjusted by making groups of people interact more often with each other, than with the rest and by adding certain movement patterns in the systems.

Another thing that could be adjusted in the model is the ability of the disease to spread. The current model is based on distance to other people. The closer you are to a person, the more likely you are to be infected. Studies in current COVID-19 pandemic have shown that the virus particles can stay on surfaces for days. This has shown to be an issue especially in slaughter houses and other low temperature work places [3]. This sort of transmission could be modelled, especially if people return to certain spots (like door handles etc.), where people by some probability will leave particles, which can be picked up within certain time-rate with a certain probability.

Since every person in the model has to calculate their Euclidean distance with any other person in the model, the computational power used when scaling the model is not linear, but quadratic. This means that generating big simulations, with millions of people as there is in a country, will take a huge amount of computation. This issue could possibly be optimized by lowering the amount of computation needed. One option could be to fraction a country into many smaller systems (communities) and allow some cross-travelling between the different areas. We could also look into reducing the time complexity of the distance calculation, by utilizing KD-Trees[4] for effective distance queries.

Our current model is not based on any specific time unit or distance. This makes the model quite flexible for different scenarios, but it also means that it can be hard to

[3] https://news.sky.com/story/coronavirus-hundreds-test-positive-for-covid-19-in-german-slaughterhouse-12009442

[4] https://en.wikipedia.org/wiki/K-d_tree

interpret the actual results and turn them into simulations that represent an actual real life event.

6.2 Discussion on the results

When it comes to the simulations, there is quite big effects on the spread of an infection based on the parameters. Population density showed to have an effect on the rate of infection, but not on whether everyone got infected. This indicates that this parameter will mainly be interesting adapt in connection with the hospital capacity parameter, showing what happens when a hospital is overloaded.

The movement factor is only having an effect up towards a certain rate, but low movement rate is a strong factor towards eliminating the infection. As with movement, another strong factor for eliminating the disease is limiting the infection distance - this factor does eliminate the disease when its lowered, but creates a steep infection rate when increased.

The death rate is a very interesting parameter, since too high a death rate will kill people before they infect further. Therefore, in order to create a truly deadly pandemic a disease needs to find a balance point for not being too lethal to kill people instantly before the disease spreads, but neither be too mild such that people become immune before they die. A factor that really plays into how lethal the disease is the recovery rate. The longer time people are sick, the higher chance they have of dying from the disease and spread it further. These factors were illustrated in section 5.1.

Our models with quarantine shows how strong a factor that can be for eliminating an epidemic. The testing capacity for the pathogen needs to be quite high and targeting infected people specifically in order for quarantine to be really effective, which can be hard to implement in reality, since people who are infected needs to be tested and a high test capacity can be expensive to implement.

Lastly, in our model the hospitalization affects the death rate. In general our models has used diseases with a high mortality. In these cases the hospital has a strong effect on survival of people, especially if they provide a cure. This is a very simple idea of hospitalization which, in real life, is not likely to be able to provide a cure that efficiently cures every person no matter age and other diseases.

6.3 Future perspectives

The following points where discussed in the group as possible options for expanding the capabilities of the model even further:

- Incubation period.

- Asymptomatic people.
- Leaving virus particles on surfaces.
- Central hubs.
- The introduction of a vaccine.
- Optimizing computational time of our program: utilizing KD-Trees

Given a situation where we had higher computing resources and more time, more complex systems like the ones mentioned above could be implemented and tested. Instead of a discrete time step, we would have used an continuous time step. Also, it would have been interesting to simulate larger populations confined within larger countries.

7 Conclusion

The aim of this project was to build a stochastic simulation model, which was programmed and tested in order to model infectious diseases and their spread in populations. The basis of the program is the so-called SIRD model, which is used to describe and analyze the dynamics in a population consisting of susceptible, infected, recovered and deceased individuals, respectively.

In this project, it has become evident that a stochastic simulation model can be used to simulate the anticipated development of a pandemic following a SIRD model. While a deterministic model such as a differential equation model has no randomness such that it will always produce the same outcome from a given starting condition, a stochastic model is able to simulate different outcomes with the same parameters, i.e. the exact same inputs will yield different outputs. This can help understand the sensitivity of the outcome, given the parameter space. Moreover, a stochastic simulation model offers a range of conditions that can be implemented with relative ease, e.g. quarantine, hospitalization and travels between populations.

Using stochastic simulation and modelling, it has been found that a combination of conditions like movement, infection range etc. has an impact on how a disease spreads through a population. The model is still very simple and has only been tested with limited variables and population sizes. Many more hypothesis can be tested with the model if more computational power and time is given.

If one were to understand how a single parameter changes the dynamics of the system as a whole, a comprehensive study could easily be carried out just for this. This fact displays that this report only provides an overview and is quite superficial in the means of studying pandemics with stochastic simulation. In layman terms, we have only scratched the surface of pandemic simulations.