

# **SIMULATION AND MODELING - ASSIGNMENT 1**

---

*Beatrice Guidotti*

On 30th January 2020 the World Health Organization announced Coronavirus outbreak to be Public Health Emergency of International Concern (PHEIC); 12 days later it was declared a pandemic. To date, the SARS-CoV-2 infected more than 121 Million people all around the globe, leading to 2.68 Million deaths. As for Italy, data display 3,26 Million confirmed cases and a total of 103,000 deaths as of 17 March 2021. Anyone contracting this infectious disease can experience mild to severe respiratory illness; the most common symptoms include fever, difficulty in breathing, body aches, loss of taste or smell. Coronavirus is particularly dangerous for those individuals with serious comorbidities as cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. Limiting and eventually ending this global pandemic requires implementation of multiple strategies, including social distancing and contact tracing. This work attempts at simulating the spread of Covid-19 by applying simple SIR and SEIR models, with and without the introduction of policies for restricting its dissemination.

## **Abstract**

The aim of this report is to describe different compartmental models able to simulate the spread of COVID-19 in thirteen provinces of the Italian Lombardy region. The first section focuses on a SEIR model, which considers four stages of infection: Susceptible (S), Exposed (E), Infected (I), and Recovered (R). For each of the thirteen provinces, different parameters have been calibrated on the basis of the available data. Calibration is essential since the tuned parameters allow to define the scope of the outbreak of Covid-19 as well as the rate at which new cases are expected to increase. In the following section, a simple deterministic SIR (Susceptible – Infected – Recovered) model accounting for location-dependent distribution of age has been implemented and calibrated for the province of Pavia, using an age-stratified contact matrix. The same model has been improved in section 3, so that to characterize the degree to which the spreading of the pandemic changes following the implementation of restrictive measures, such as the closure of schools and the introduction of compulsory quarantine for the entire population. The results obtained show that even simple SEIR and SIR models with appropriate underlying assumptions are able to provide a theoretical framework to uncover the evolution of SARS-CoV-2 over time.

## Section 1: SEIR model

In order to capture the spread of Covid-19, a classical SEIR model has been developed.

Individuals have been divided into four compartments: S (susceptible), E (exposed), I (infectious) and R (recovered). In particular:

- S indicates the portion of the population which is susceptible to infection;
- E indicates the portion of the population which is infected but asymptomatic;
- I indicates the portion of the population which is infected and has symptoms;
- R indicates the portion of the population which has recovered from the infection.

This model has been chosen because it allows to take into account the latent period of the infection. Indeed, the transmission process of COVID-19 occurs due to an initial contact between a susceptible and an infectious person, which leads to the virus inoculation. It follows a period of time in which newly infected individuals are asymptomatic. For simplicity, it has been assumed that during this incubation period individuals are also non-infectious. Hence, a person that comes into contact with an infectious individual becomes infected but still, she is not able to transmit the virus, thus she is included in the Exposed compartment. The addition of this latent period introduces a time delay into the system: the growth rate of the infection is slowed down due to individuals passing through the exposed group before they become effectively infectious, contributing to the transmission process.

The other main assumptions underlying the proposed model are that recovered patients become permanently immune to COVID-19, which can be justified by the fact that it has been proved that immunological memory to SARS-CoV-2 lasts for at least six months (Dan J.M. et al., 2021). This assumption is reasonable since the true data provided for the study cover a period of only 93 days. For simplicity and for the sake of this assignment, the model does not include virus-induced mortality, it considers a homogeneous mixing of the population and constant rates for the transition between compartments.

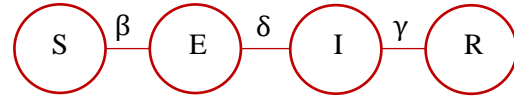


Figure 1: SEIR model dynamics

The proposed SEIR model is governed by the following system of differential equations:

$$\frac{\partial S_t}{\partial t} = bN - \beta \frac{I_t}{N} S_t - mS$$

$$\frac{\partial E_t}{\partial t} = \beta \frac{I_t}{N} S_t - \delta E_t - mE$$

$$\frac{\partial I_t}{\partial t} = \delta E_t - \gamma I_t - mI$$

$$\frac{\partial R_t}{\partial t} = \gamma I_t - mR$$

Where:

- $\beta$  is the *transmission probability*, namely the probability that a contact causes infection;
- $\delta$  is the *rate at which infected become infectious*, calculated as the inverse of the average duration of the latent period;
- $\gamma$  is the *recovery rate*, defining how fast an infectious individual recovers and calculated as the inverse of the average duration of the infectious period;
- $b$  is the *natural birth rate* in Lombardy, set equal to 0.0072<sup>1</sup>;
- $m$  is the *natural mortality rate* in Lombardy, set equal to 0.0099<sup>1</sup>.

With reference to the above-described model, unobserved quantities characterizing the dynamics of the epidemic have been calibrated, by fitting the model to real COVID-19 data from the thirteen provinces of Lombardy. In particular, different  $\beta$ ,  $\delta$  and  $\gamma$  have been calculated for each province by reducing the Rooted Mean Squared Error between the predicted and the true data, with each province associated with a different number of training days. The RMSE approach has been chosen mainly because it is the simplest one in terms of interpretability. In addition, Metropolis-Hastings algorithm and Maximum Likelihood Estimation have been tested on an arbitrarily selected province (Pavia) for the sake of proving that they all lead to similar results (See Appendix 1). The selection of the ranges for the calibration of the different parameters have been

based on domain knowledge: the basic reproductive number,  $R_0$ , has been calibrated to be between 1 and 3, since the current estimate for Covid-19  $R_0$  is around 2.5<sup>2</sup>; the average latent and the average infectious periods have been selected to vary between 3 and 14 days since, according to the data provided by WHO<sup>3</sup>, the incubation stage for COVID-19 is on average 5-6 days, but can last up to 14 days; moreover, people remain infectious for up to 10 days after the beginning of symptoms in moderate cases, and up to 20 days in severe cases. By conditioning on the observed number of infected individuals, it is possible to obtain a posterior distribution of the number of daily infected.

Although there is some level of error between the estimated and the actual numbers of new daily cases, the predicted values reasonably resemble the real data on the interval deployed for calibration. Nevertheless, if we integrate the SEIR equations with the optimal parameters over a longer period of time, e.g. 100 days, the estimated pattern will show the exponential growth lasting much longer with respect to the increase displayed by the real curve. Indeed, one should not expect the rest of the simulated curve to resemble the pattern observed in the real data because the parameters have been calculated by minimizing the RMSE on data characterized by an increasing pattern, therefore the predicted pattern will present an exponential growth even after the threshold point (i.e. the limit day used for calibration), as shown in Figure 2.

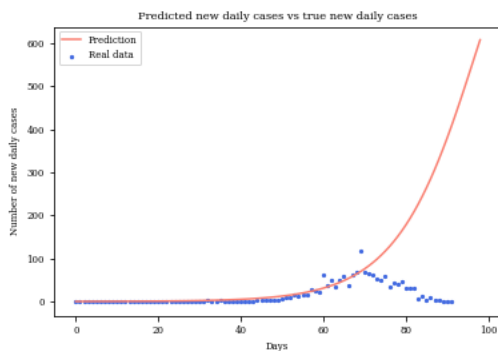


Figure 2: New daily cases: true data over predicted curve

## Section 2: Adding age structure to the model

This section describes the development of a simple SIR model accounting for age structure. In particular, the population has been divided into sixteen 5-age groups, where the first class represents individuals between 0-4 years and the last class comprises people aged 75+. The interactions between individuals are defined by a contact matrix  $C$ , where the entry  $c_{ij}$  indicates the average number of contacts between an individual belonging to class  $i$  and a person within age group  $j$ .

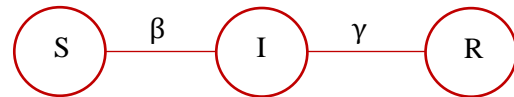


Figure 3: SIR model dynamics

The age-stratified SIR model is characterized by the following system of equations:

$$\frac{\partial S_i}{\partial t} = - \sum_j \beta_{ij} \frac{I_j}{N_j} S_i$$

$$\frac{\partial I_i}{\partial t} = \sum_j \beta_{ij} \frac{I_j}{N_j} S_i - \gamma I_i$$

$$\frac{\partial R_i}{\partial t} = \gamma I_i$$

Where:

- $\beta$  is the *transmission matrix*, also called WAIFW matrix, in which the element  $\beta_{ij}$  denotes the transmission from an individual in class  $j$  to a person belonging to group  $i$ .  $\beta$  is defined as the product of  $q$ , the *intrinsic probability of transmission*, assumed to be the same across all the age classes, and  $C$ , the *contact matrix*;
- $\gamma$  is the *recovery rate*.

For the development of this model, it has been assumed that the size of the population is constant, i.e.  $\sum_j S_j + I_j + R_j = N$  and that the recovered individuals are immunized, as assumed for the SEIR model described in Section 1. For simplicity, the model does not include mortality due to Covid-19 and assumes that there is no latent period in the infection process. The number of individuals in each age group has been calculated by multiplying the initial number of susceptibles for the proportion of people per age group. The latter has been

defined according to ISTAT demographic data (*Popolazione residente al 1° gennaio 2021*<sup>4</sup>) simply by summing over age groups and dividing by the total number of individuals in the population. The first infected individual has been arbitrarily assigned to age group number 11, i.e. age group 55-59 years. This implies that at the outset of the epidemic, nearly everyone is susceptible, therefore we expect an optimal  $R_0 > 0$ , since data show that the epidemic spreads out. The model has been calibrated using the RMSE approach. In particular, since

$$\beta = q * C$$

and

$$R_0 = \frac{q}{\gamma} \rho(C),$$

the optimal basic reproductive number (in terms of lowest RMSE) has been selected, leading to the optimal value of the transmission matrix. The value of  $R_0$  calibrated in the SIR is lower than the optimal  $R_0$  for the SEIR model calculated in Section 1. This because the previous model includes a latency period, which slows down the rate at which the peak of new infectious cases is reached. On the contrary, within a SIR model, as soon as a susceptible person comes to contact with an infected, she immediately transitions to the I compartment. Thus, the SEIR model shows a slower growth rate after infection since the individual has to pass through the E class before being able to transmit the virus to other people. Furthermore, it can be noticed that the dynamics of the three compartments are slightly different for each age group. Indeed, the numbers of Susceptible, Infected and Recovered over time follow the standard SIR dynamics we would expect; still, each group is characterized by a different rate at which individuals transition from S to I, due to the peculiar values of the contact matrix, leading to distinct slopes of the Susceptible and Infected curves (See Appendix 2).

In addition, by looking at the predicted rate of new daily infected per each age group, it is possible to assess that 15-19 years and 10-14 years groups are the most infected ones, irrespectively of the age of the first infected individual. (Figure 4)

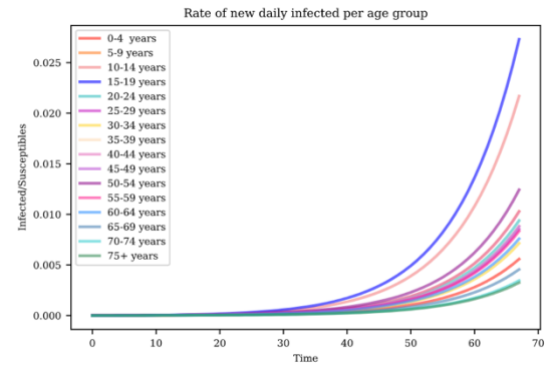


Figure 2: Rate of new daily cases per age group

This result is in line with expectations, since the values of the contact matrix are higher for these age ranges, which suggests that members of these groups have the same role of super-spreaders: their higher numbers of contact imply a higher probability of contracting the disease, as well as a higher chance in affecting other individuals, given that  $q$  and  $R_0$  are constant across the whole population.

### Section 3: Improve the model

Here, the age-stratified SIR model described earlier is extended through the addition of two restrictive measures that have been implemented in the Lombardy Region in order to reduce the spread of the virus: the closure of schools and the compulsory quarantine.

The former mirrors the importance of sheltering by different age groups: in the previous section it has been shown that individuals between 10 and 19 years of age are the ones presenting the higher risk of contracting the infection. This suggests that demography must be considered for the introduction of population-level specific interventions. The closure of schools limits the contact between individuals aged 0 to 24, which in turn decreases the probability of transmission from and to people within this age range. Compulsory quarantine will be introduced afterwards, further reducing the contact matrix for the whole population. The assumptions underlying these restrictions are that all individuals are compliers, and that everyone is subject to quarantine.

Furthermore, it has been assumed that the closure of schools has a heterogeneous effect on the different clusters: individuals between 0 and 24 years are the most affected one, since school

<sup>4</sup> <http://dati.istat.it/Index.aspx>

represents the main moment of gathering for youths. The reduction in the number of contacts is even higher if two individuals belong to the same age group. As for people aging 25+, the schools closure reduces the average number of contacts, but this decrement is lower than the one for younger people since the former are assumed to have other major moments of gatherings throughout their day (See Appendix 3).

The imposition of the previously-mentioned restrictions has been implemented by applying two different approaches: a deterministic and a stochastic one. Both of these algorithms involve two diverse scalings of the contact matrix at two different points in time and lead to similar results. Since the model under analysis is an SIR, it does not consider the latency period that Covid-19 entails. Therefore, it is assumed that the effect of the restrictions will be seen from the exact day of their introduction.

From the available data, it is possible to assess that new cases started increasing from day 50 up to the peak reached at day 69. It is reasonable to close the schools when the cases start showing an increase, for instance at day 55. Since quarantine is a much harsher restriction than the former, it will be introduced at a point in which it is clear that the number of new daily infected is growing exponentially: in particular, day 69 has been chosen to be the starting time for the compulsory quarantine.

As for the deterministic approach, three time ranges have been defined and a SIR with different parameters is run for each time frame. First, the SIR model has been run from time 0 to 54; then, the composition of the S, I, R compartments characterizing that last step has been used at time 55 as the input for an SIR model considering the scaled contact matrix representing school closure. The latter has been used to calculate the numbers of Susceptible, Infected and Recovered individuals from time 55 to 68. Finally, in a similar fashion, the same model has been run from time 69 to 92, using as

initial population the composition of the S, I, R groups at time 55 and applying a further scaling of the contact matrix representing compulsory quarantine. The three models are then combined together in order to create a cumulative plot showing the overall effect of the implemented restrictions (see Appendix 4.1).

In the second approach, instead, an SIR model with stochasticity in the infection transmission and recovery processes has been developed. One hundred simulations have been run in order to assess how the stochasticity component affects the output. The same reductions in the average number of contacts that have been used for the deterministic approach are applied here, at the same points in time.

Both approaches lead to similar results, as shown in Appendix 4: the marginal impact of regulations in reducing the spread of infection is different by age groups and it mainly depends on the values of the contact matrix of each age group. This is evidence of the importance of introducing age-specific restrictions: limiting people with a high average number of contacts can significantly decrease infections. Indeed, at time 55, the closure of schools mainly affects the infection process of people aged between 0 and 24, leading to a reduction of the growth rate of infection for these individuals. At time 69, all the age classes experience a sharp decrement in the number of new infected, due to a further and harsher scaling of the contact matrix, representing the effect of quarantine.

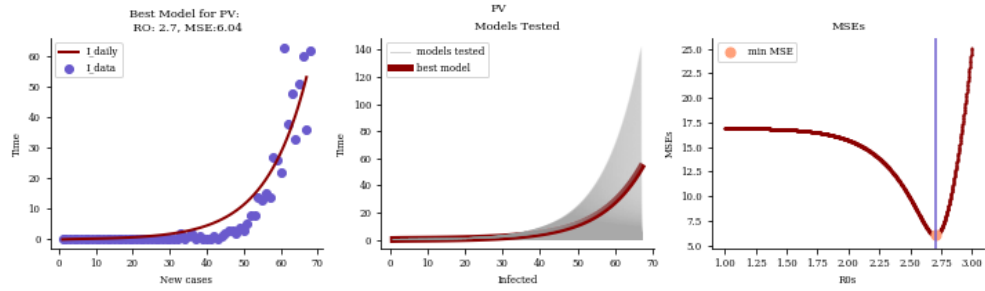
### **Final considerations**

Mathematical modelling tools as SEIR and SIR proved to provide an effective way of predicting the spreading and pattern of an infectious disease as Covid-19. Notwithstanding, simple SEIR and SIR models are able to provide a theoretical framework to uncover the factors underlying the spread of Coronavirus and its evolution over time.

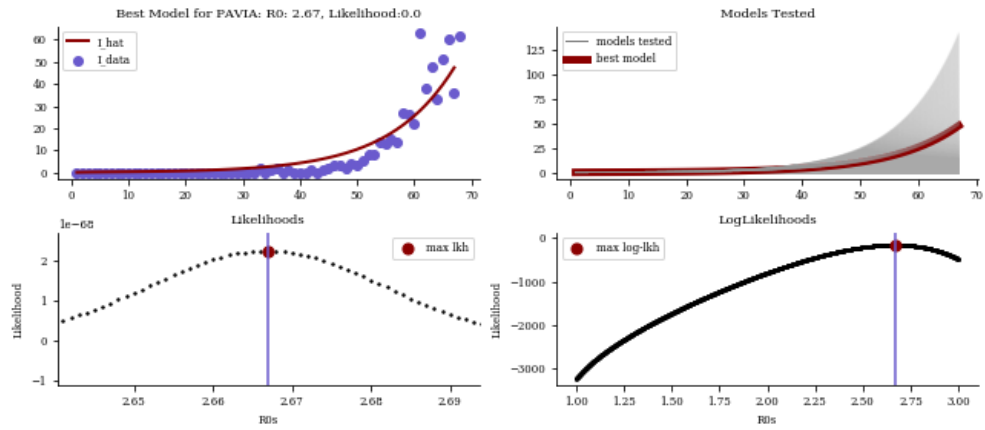
# Appendix

## Appendix 1

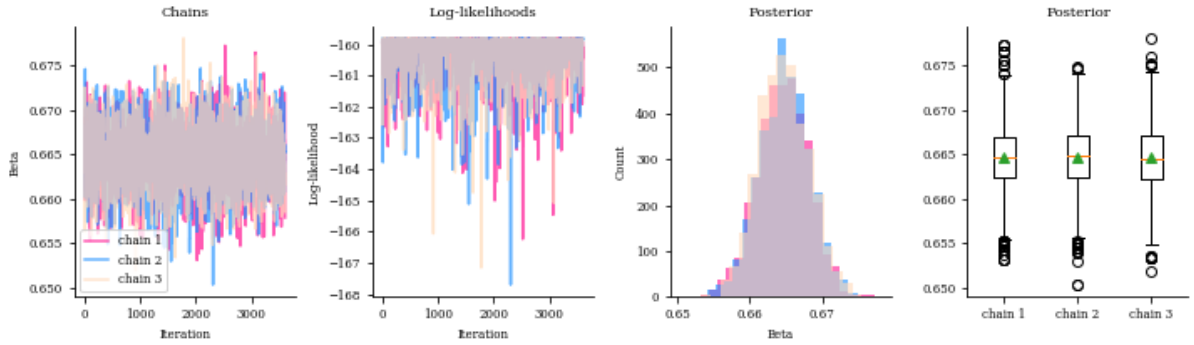
Results from the different calibration techniques, applied to the city of Pavia.



Appendix 1.1: RMSE



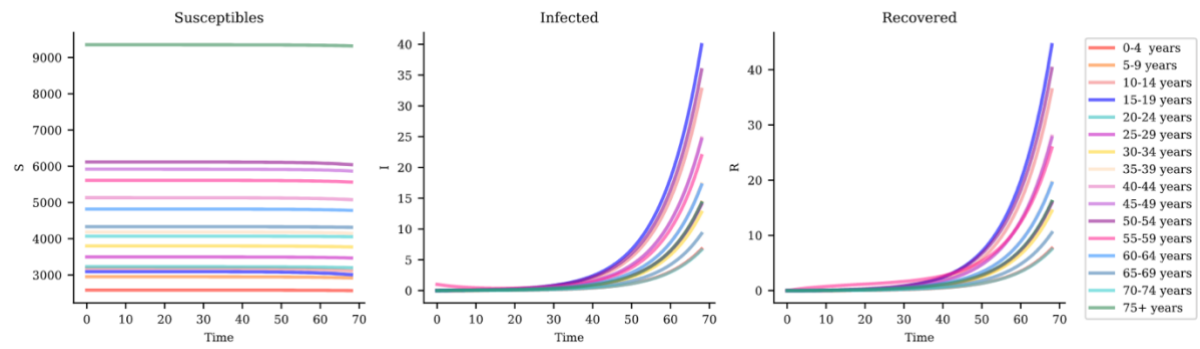
Appendix 1.2: MSE



Appendix 1.3: Metropolis-Hastings



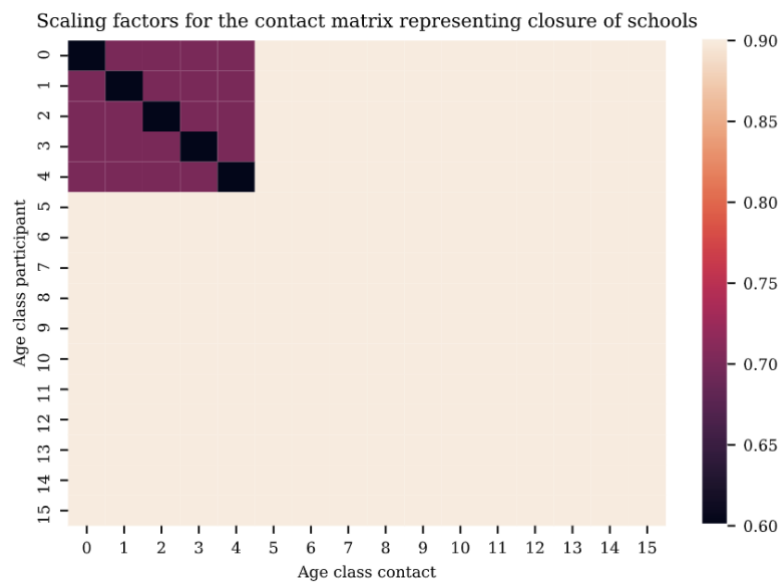
## Appendix 2



Appendix 2: SIR dynamics for different age groups

## Appendix 3

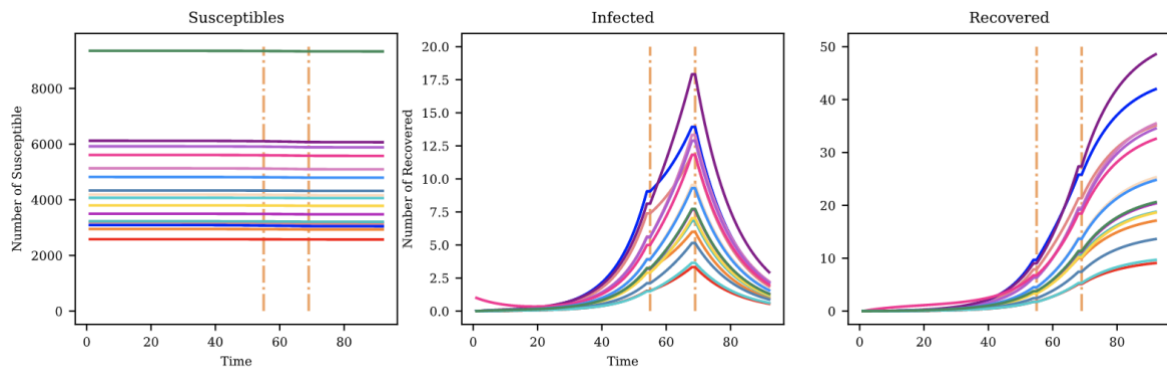
Graphical representation of the scaling of the contact matrix following the closure of schools.



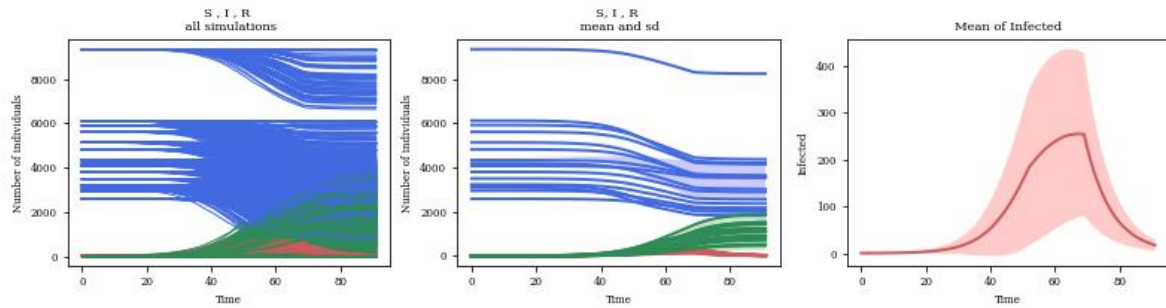
Appendix 3: Heat map of the scaling of the contact matrix

## Appendix 4

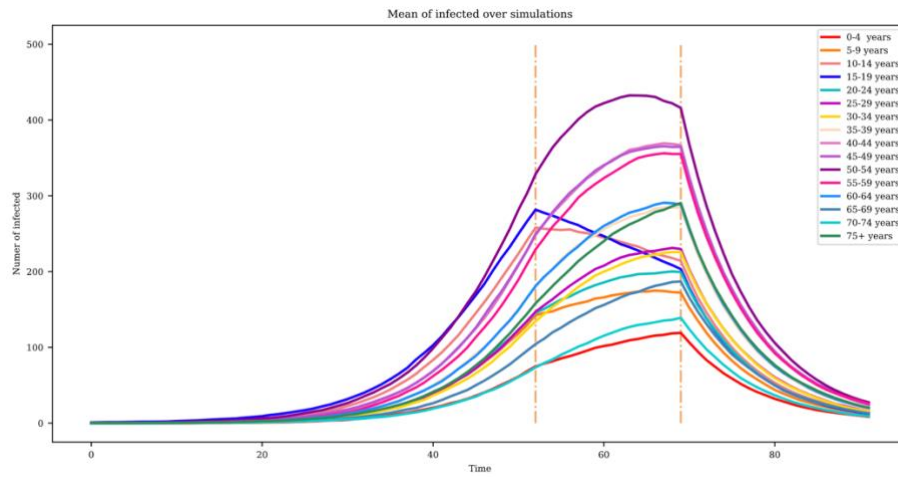
Comparison between deterministic and stochastic approaches for the introduction of restrictive policies.



Appendix 4.1: SIR results with deterministic approach



Appendix 4.2: SIR results with deterministic approach - 100 simulations



Appendix 4.3: Average over simulations of infected



## References

Dan J.M. et al., 2021; *Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection*. Science.

He, S., Peng, Y. & Sun, K., 2020; *SEIR modeling of the COVID-19 and its dynamics*. <https://doi.org/10.1007/s11071-020-05743-y>

Wani U.A., Bakshi A., Wani M.A., 2020; *Dynamics of COVID-19: Modelling and Analysis*. Journal of Infectious Diseases Epidemiology. doi.org/10.23937/2474-3658/1510128

World Health Organization, 2021; *Coronavirus disease 2019 (COVID-19) Situation Report – 73*