20599 - Simulation and Modelling
DSBA

Midterm Assignment

Niccolò Diana
3045616
Rebecca Solcia
3049890

March, 2020

Università Commerciale
Luigi Bocconi

1. Introduction

Last January marked one year of CoVid-19. This virus, causing acute respiratory symptoms, started spreading from the Central Chinese province of Wuhan and reached every corner of the world within a few months.

Despite the closure of the Italian borders to China on the 31st of January 2020, the Italian country has not been able to avoid the impending health disaster. While the first emergency hospital was being built in Wuhan in record time, Italy detected its patient zero: a 38-years old man living in Codogno, a small town in the Northern province of Lodi. From that day on, the government started implementing restriction measures that saw the 'zones system' as the very first attempt of containment. 'Red zones' were established in a total of eleven towns across the Northern Regions, and their inhabitants could not leave their place of residence nor meet in public or private places[1].
As of today, CoVid-19 still represents a problem that has been timidly contained, but has not yet been eradicated. With a total of three massive waves, 102 000 victims, and 3.2 millions of infected people, the virus is tracking record numbers. The Italian country, profoundly marked by the pandemic, has also been facing economic and social problems as a byproduct of the necessary containment measures.

Apart from being an assessment opportunity, this assignment constitutes a valuable chance to gain insight regarding the moment we are going through by applying the techniques we have been learning for the past two months.

2. Exercise 1: Calibration

To model the evolution of the CoVid-19 pandemic in the provinces of Lombardy, we have decided to implement a **SEIR model**.

This model has been widely used in the study of epidemic diseases and, despite being a simple extension of the SIR (Susceptible, Infected, Recovered), it proved to be helpful in predicting the spreading of infections. According to this compartmental model, the population is divided into four groups and each individual is either susceptible, exposed, infected or recovered (S,E,I,R) [2]. Susceptible individuals can potentially be infected by the virus, upon contact with an infected individual, at rate β. The E class, representing the main difference with the SIR model, comprises all the individuals that, after contracting the virus, incubate the CoVid-19 for a period L. In this case, the individuals are infected but not yet infectious. Infected individuals are, instead, people displaying the symptoms and exposed individuals access this class with probability one at the end of the incubation period, whose duration varies from person to person. Finally, once the disease has run its course and the infected individuals are no longer infectious, they develop antibodies and they move on to the Recovered class. The recovery process for infected individuals happens with probability γ.

When analysing the different stages of a CoVid-19 infection, and taking into account the characteristics of the dataset we were given, SEIR appears to be the best and simplest model to use. Specifically, the infection can be divided into three steps of increasing severity: an early stage, a pulmonary phase and an hyperinflammation phase. [3] The infection begins when a person in-
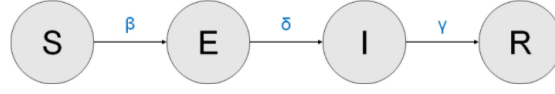
Fig. 1. SEIR model diagram

hales droplets or aerosols. In just 10% of the patients the infection can happen when the individual touches a contaminated object and then her hands comes into contact with the face. The peak of infectiousness for an individual is between the two days before the emergence of symptoms (if any) and the seven days following the appearance, on average. The median length of the incubation period is about 6 days. [4]

The variables we used to define our SEIR model are:

- N: the total population, which is kept constant at each time t and is equal to the sum of S(t), E(t), I(t) and R(t).

- β: the probability of a susceptible individual to be infected by an infectious individual

- δ: the inverse of the average latent time, governing the lag between having undergone an infectious contact and showing symptoms. In the equations, it brings people from the E category to the I category.

- γ: the proportion of infected individuals who recover. Similarly to δ, it is the inverse of the average time one person needs to recover from the disease.

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta S \frac{I}{N} \\
\frac{dE}{dt} &= \beta S \frac{I}{N} - \delta E \\
\frac{dI}{dt} &= \delta E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}
\tag{1}
$$

While data about the actual population in each province (N) and the true data of new cases were available, the other parameters were not expressed and needed to be either estimated or accepted by assumption. More specifically, in the calibration phase we were required to model the beginning of a CoVid-19 wave affecting the twelve provinces of Lombardy. To tune the model, the very last observations in our dataset were discarded (in a number that varied depending on the province). The choice of taking into account just the first part of the dataset to train the models made particular sense because the specification of the SEIR we chose was describing a picture in which the growth of daily infected individuals was of exponential nature.

After a review of the literature, we initialized the simulation with just an infected individual and the rest of the population as susceptible, as this setting could best describe the initial phases of a

pandemic. In this way, we left β , the probability that an infected individual infects a susceptible individual, γ, the probability of an infected individual to recover, and δ, the probability of an exposed to become infected, as parameters to be estimated.

By using three different calibration routines (namely, Root Mean Squared Error minimization, Maximum Likelihood Estimator maximization and Metropolis Hastings algorithm) we have obtained different province-specific pairs of values of L and D. Sample outputs for Pavia for these routines (to be taken as an example of the outputs in the Notebook) can be found in Figure 9, Figure 10, and Figure 11 in the Appendix Section.

Despite this approach was particularly interesting, we performed this grid search just for the sake of curiosity. In fact, our data were not suited for this type of analysis. In the end, we chose to rely on a single value for both δ and γ and perform calibration leaving just the value of β to change across iterations.

In the best models, with $D = 14$ and $L = 6$, the parameters we obtained were:

Table 1. Parameters obtained through calibration

	LO	PV	SO	MN	MI	CO	BS	MB	LC	CR	BG	VA
R0	4.09	3.65	2.82	3.37	4.42	3.15	4.56	3.54	3.45	4.15	4.68	3.24
Beta	0.29	0.26	0.20	0.24	0.32	0.23	0.33	0.26	0.24	0.30	0.33	0.23
RMSE	9.56	5.95	2.4	4.57	15.29	3.23	19.56	5.14	5.28	7.38	22.89	3.49

Once all the needed parameters have been calibrated, we have run a final SEIR model on the whole period of 92 days for each province to graphically compare the true data with the simulated ones. The results confirmed our intuition as, in each plot, the final part of the two series diverged substantially: the simulated curve followed an exponential trend, while the true data decreased until reaching a level of zero. Here is an example:
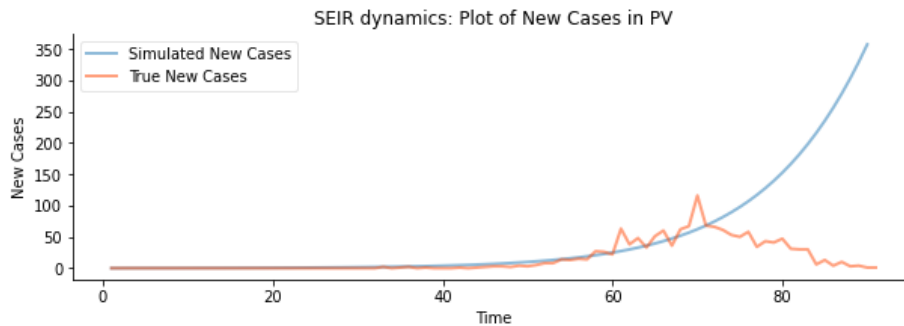
Fig. 2. SEIR simulation against true data for Pavia

This result is not particularly surprising. The SEIR model we have described did not include parameters that could regulate the introduction of restrictive measures that have been instead implemented in reality to reduce the number of cases.

3. Exercise 2: Adding age structure to the model

To carry out this exercise, which implied the use of a SIR model with age stratification, we decided to analyse the province of **Monza e Brianza**.

We first downloaded data about the population distribution from ISTAT. We divided such population into the desired 16 age groups and ended up with the grouping shown in Table 4 of the Appendix Section.

The SIR model we worked with throughout the whole Exercise 2 is the one that follows.

$$
\begin{aligned}
\frac{dS_i}{dt} &= -q \cdot \sum_{j=1}^{16} c_{ij} \cdot S_i \frac{I_j}{N_j} \\
\frac{dI_i}{dt} &= q \cdot \sum_{j=1}^{16} c_{ij} \cdot S_i \frac{I_j}{N_j} - \gamma I_i \\
\frac{dR_i}{dt} &= \gamma I_i
\end{aligned}
\tag{2}
$$

Specifically, Susceptible individuals of age-class i become infected upon contact with an infected from class j, and such contact happens with a probability dictated by the element c_{ij} of the contact matrix. As usual, the transition from Infected to Recovered for age-class i depends on the parameter γ (which is the same across every group $i = 1, ..., 16$) and the number of infected individuals belonging to i.

In this exercise, the most delicate part was the calibration of the model. In fact, this specific age-stratified setting implied that the parameter R0 was defined as

$$
R0 = \frac{q}{\gamma} \cdot \rho(C)
\tag{3}
$$

with $\rho(C)$ as the spectral radius of matrix C. We calculated the maximum of the eigenvalues of C by using the Gelfand Formula and then checked for convergence. The results of the calibration of the model, assuming $\gamma = \frac{1}{14}$, are shown in Table 2. It is also important to remember that we started the pandemic from a single individual belonging to the 75+ age group.

Table 2. Parameters obtained through calibration

	SIR model
R0	2.59
Q	0.00887
RMSE	4.03

As we can see, the $R0$ of this setting differs from the one we have obtained while calibrating the SEIR in Exercise 1. Despite the two $R0s$ differ from one another by one unit, it is important to note that for the SEIR model, the latency period changes the expected duration of an epidemic: as the latency period decreases, the duration of the epidemic also decreases and converges to the

SIR model. Since we were calibrating the two models on the exact same time-span (that moreover coincided with just the first part of the true pandemic, where growth was exponential), the only way for the SEIR simulation (with a fixed latency period) to resemble the SIR simulation was to have a greater number of secondary cases.

After having calibrated the model, we ran the SIR over a longer period of time, to allow for a visualization of stabilized dynamics. While the aggregate dynamics are shown in Figure 3, Figure 12 in the Appendix illustrates the evolution of the pandemic for six different age sub-classes.
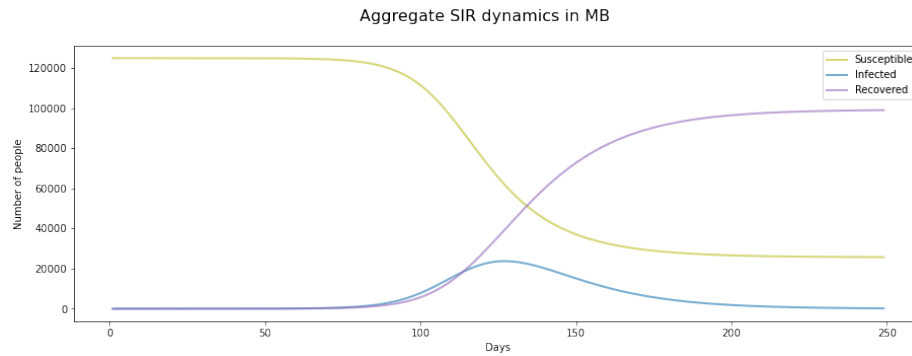
Fig. 3. Aggregat SIR simulation for Monza and Brianza

To answer the final question of Exercise 2, asking which are the groups that are infected the most, we calculated the fraction of infected people within each age group. As it is possible to see from Figure 5, about 10% of the individuals aged 15-29 got infected during the peak of the pandemic (around day 110). Despite this 14-years aggregation is quite wide, it is not surprising that adolescents and young adults have always the greatest percentage of in-class infected. By the same token, all the other classes (except for the ones comprising the oldest people) peaked to 8% during the most critical moment of the pandemic, with the 0-15 class rising a little bit earlier than the adult groups. Even if we started the pandemic from the 75+ class, the 60-74 and the 75+ class always displayed a lower fraction of infected within the respective classes. Once again, the contact matrix can explain this trend: older people tend to have less contacts, on average, than other age groups, both in within- and between-class terms.
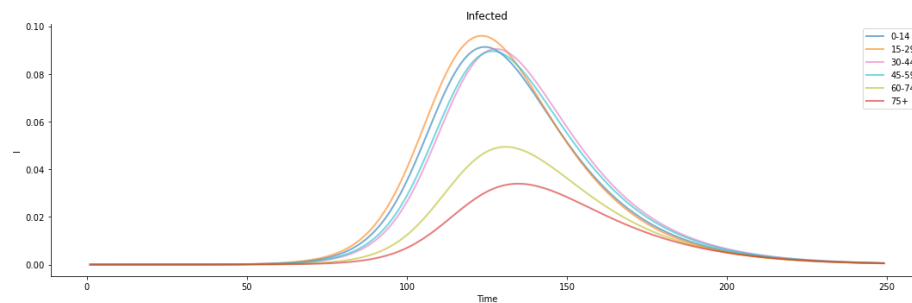
Fig. 4. Proportion of infected of the SIR simulation for Monza and Brianza

4.   Exercise 3: Improve the model

To extend the age-stratified SIR model we presented in the previous exercise, we decided to implement two restrictions, resembling the ones taken by the Italian government. Specifically, we allowed the model to accommodate for the adoption of measures forcing individuals to wear masks and to quarantine. On the other hand, no demographic characteristics (such as natural death and birth rates) were added to the model since the analyzed time-span was very small.

The overall approach to this exercise was simple, yet powerful: we built two different models to then be able to make a comparison. For the first model, we drew the restrictions from the data, chopping it in three parts and calibrating the model in each separately, allowing us to capture different phases of the process' evolution. The second model was instead built upon the use of stochastic and deterministic measures that allowed us to analyze its behavior.

4.1   Calibration with restrictions

With this first approach we considered three distinct time spans in the domain of the data in order to calibrate our model piecewise. By looking at the data and its pattern, we have selected $t_m = 50$ and $t_q = 72$ as the days in which the laws of wearing masks and quarantine come into effect, respectively. Since we have calibrated a $d = 14$, we expected to see the effects of quarantine at $t_{qe} = t_q + d = 86$, assuming the average period as a threshold after which all the effects were visible. In this case, the restriction coefficients were ψ and δ an were calibrated (using the RMSE minimizaion technique) considering them as a unique element. The measure-updated set of differential equations then became as follows.

$$\frac{dS_i}{dt} = -(q - \phi \mathbb{1}_{t \geq 50} - \delta \mathbb{1}_{t \geq 86}) \cdot \sum_{j=1}^{16} c_{ij} \cdot S_i \frac{I_j}{N_j}$$

$$\frac{dI_i}{dt} = (q - \phi \mathbb{1}_{t \geq 50} - \delta \mathbb{1}_{t \geq 86}) \cdot \sum_{j=1}^{16} c_{ij} \cdot S_i \frac{I_j}{N_j} - \gamma I_i \qquad (4)$$

$$\frac{dR_i}{dt} = \gamma I_i$$

Once the restriction parameters were calibrated, we ran a simulation with our new age-stratified SIR to be able to assess the graphic performance of our new model.

As it is possible to see from Figure 5, the introduction of masks slowed down the contagion process, while quarantine inverted the exponential tendency of the pandemic. The joint action of masks and quarantine blocked the spread of the disease in the sense that the peak of the curve for the new daily infected was reached at just 120 (out of about 130,000 individuals inhabiting the province of Monza e Brianza). Despite this result might seem surprising at first sight, it resemble quite well the curve for the new daily infected we were given at the beginning of this project. By the same token, the plot for the Susceptibles is characterized by straight lines: the majority of the population could not even reach the Infected compartment.
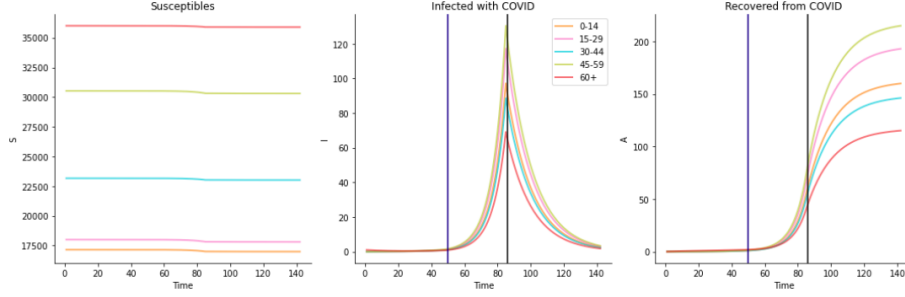
Fig. 5. SIR Model with masks (blue) and quarantine (black), calibrated on data

4.2 Stochastic Framework

In this second framework, we allowed for stochasticity in the spread of the infection. We sampled the number of individuals transitioning from a group to the other from a Binomial distribution, which we parametrized by the vector λ (that needed to be computed) and the vector of susceptible population per age group.

At this stage, we did not perform calibration as the nature of the two measures differed from the ones we presented before. The quarantine consisted in a lockdown happening overnight, implying that from t_i to t_{i+1} the contact matrix C dropped immediately, therefore turning the lockdown into a deterministic measure (parametrized by the vector ψ). In this case, ψ_i represented a "contact reduction" coefficient varying for each age group. Intuitively, the elements of the ψ vector were lower on the ends of the spectrum (for kids and elders), and higher for the age-groups in the middle, who naturally experienced more contacts because of their higher levels of social interactions (such as groceries and work in primary sectors exempted from quarantine, to name a few).
On the other hand, masks are a physical good of which the country has experienced several shortages during the initial stages of the pandemic, and they also carried a highly aleatory subjective compliance. For these reasons they will be implemented as a stochastic measure acting on C indirectly, through reducing q. We called this last parameter, regulating the compliance to the introduction of masks, ω.

Therefore, rewriting our system of differential equations defined piecewise we got the following.

$$\frac{dS_i}{dt} = \begin{cases} -\lambda_i \cdot \sum_{j=1}^{16} c_{ij} \cdot \frac{I_j}{N_j} & t < 50 \\ -\lambda_i \cdot \sum_{j=1}^{16} c_{ij} \cdot \psi_j \cdot \frac{I_j}{N_j} & t \geq 50 \\ -\lambda_i \cdot \omega \cdot \sum_{j=1}^{16} c_{ij} \cdot \psi_j \cdot \frac{I_j}{N_j} & t \geq 86 \end{cases}$$

$$\frac{dI_i}{dt} = \begin{cases} \lambda_i \cdot \sum_{j=1}^{16} c_{ij} \cdot \frac{I_j}{N_j} - \pi_i & t < 50 \\ \lambda_i \cdot \sum_{j=1}^{16} c_{ij} \cdot \psi_j \cdot \frac{I_j}{N_j} - \pi_i & t \geq 50 \\ \lambda_i \cdot \omega \cdot \sum_{j=1}^{16} c_{ij} \cdot \psi_j \cdot \frac{I_j}{N_j} - \pi_i & t \geq 86 \end{cases} \tag{5}$$

$$\frac{dR_i}{dt} = \pi_i$$

Where the parameters λ and π are defined as follows

$$\lambda \sim Binomial(S_{t-1}, q, 1)$$
$$\pi \sim Binomial(I_{t-1}, \gamma, 1) \tag{6}$$

And the parameters ω and ψ are retrieved through logical assumptions and the researchers' experience.

$$\omega \sim Uniform(0.5, 0.7) \tag{7}$$

Table 3. Values of the vector ψ

| ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_8	ψ_9	ψ_{10}	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{15}	ψ_{16}
0.1	0.1	0.15	0.2	0.25	0.3	0.35	0.35	0.35	0.35	0.3	0.3	0.2	0.1	0.05	0.05

These plots in Figures 6, 7 and 8 resemble the average of 50 simulations on the population dynamics overtime with no restrictions, only masks, and both masks and quarantine, respectively. Numbers in the middle and right-most graphs are averaged not only through simulations, but also among age groups, to favor a clearer presentation of the graphs.

With no restrictions, the average number of infections over age groups reached a relatively high peak of 1000 while, when introducing masks, the spread of the infection is slowed down considerably and spread over more than twice the time horizon.
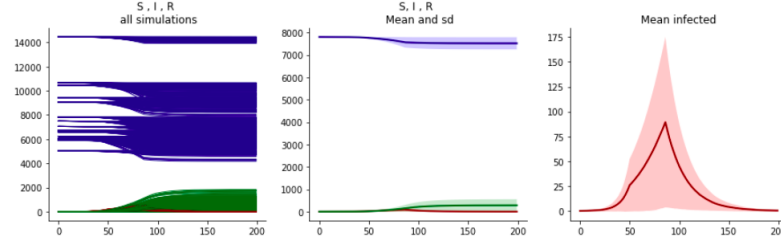
Finally, in Figure 7, we can see that quarantine is a significant restriction that, if perfectly complied with, slows down the spread considerably.
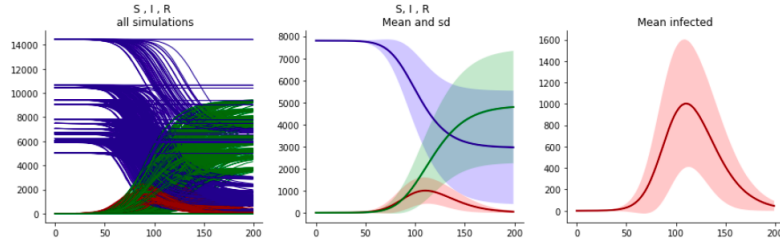
Fig. 6. Stochastic SIR with no restrictions
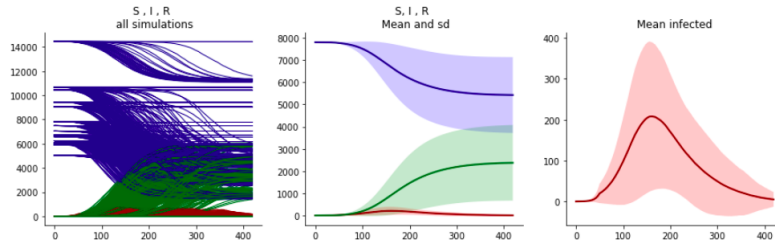
**Fig. 7.** Stochastic SIR with stochastic masks at t_m

Fig. 8. Stochastic SIR with stochastic masks at t_m and deterministic quarantine at t_{qe}

References

[1] IlSole24Ore (2020) Cose che noi umani, https://lab24.ilsole24ore.com/storia-coronavirus/. [Online; accessed 14-March-2020].

[2] Keeling MJ, Rohani P (2008) *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press), . https://doi.org/10.2307/j.ctvcm4gk0,. URL http://www.jstor.org/stable/j.ctvcm4gk0

[3] Siddiqi (2020) Covid-19 illness in native and immunosuppressed states: A clinical-therapeutic staging proposal. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation* 39(5):405–407. https://doi.org/10.1016/j.healun.2020.03.012

[4] He (2020) Temporal dynamics in viral shedding and transmissibility of covid-19. *Nat Med* 26:672–675. https://doi.org/10.1038/s41591-020-0869-5
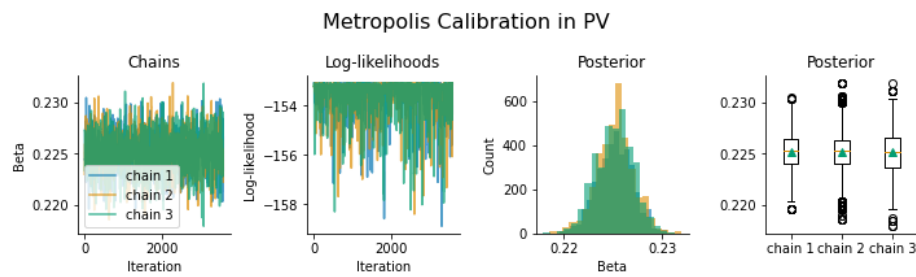
Appendix A: Supplemental Materials
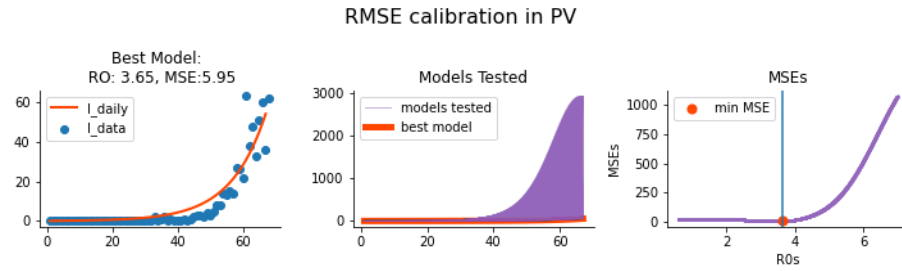
**Fig. 9.** RMSE calibration output for Pavia
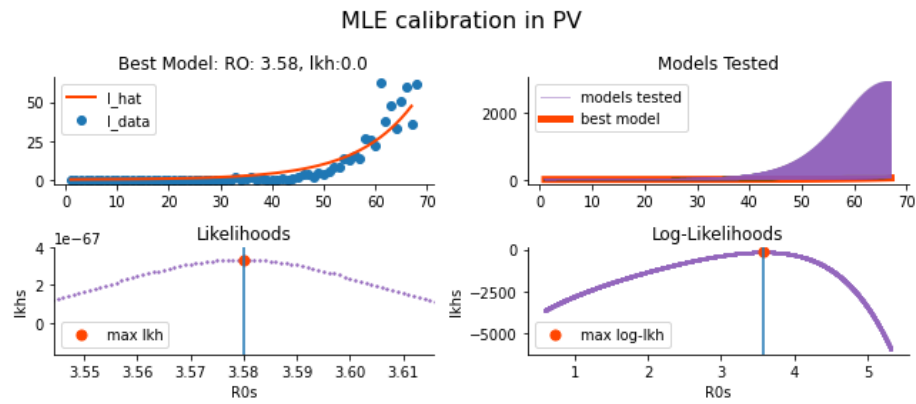
Fig. 10. Maximum Likelihood Estimator calibration output for Pavia

Fig. 11. Metropolis Hastings calibration output for Pavia

Table 4. Age groups in Monza e Brianza

Age group	Number
0-5	34974
5-10	41045
10-15	43510
15-20	41784
20-25	41444
25-30	42226
30-35	46022
35-40	52186
40-45	63243
45-50	72800
50-55	74230
55-60	65732
60-65	54492
65-70	49124
70-75	46686
75+	100695
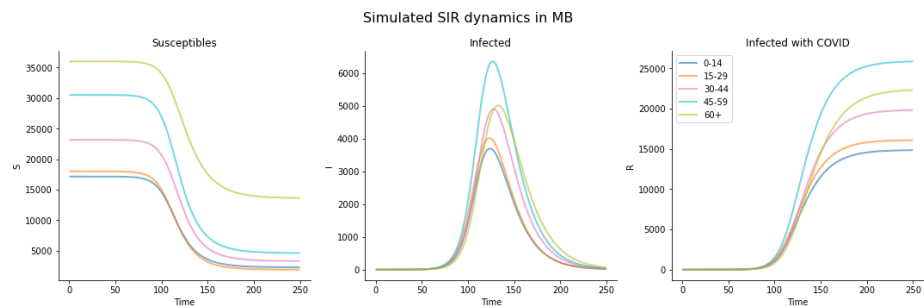
Fig. 12. Age-stratified SIR simulation for Monza and Brianza