

Simulation & Modeling of Covid-19

Name student:
Steffen BROCKMANN
Louis LACOMBE

Date final version: March 28, 2021

I Introduction

The first cases of the respiratory disease COVID-19, caused by the novel coronavirus called SARS-CoV-2, in Italy were confirmed on January 31st, 2020. In the following three weeks, clusters of cases were detected in the northern parts of Italy, including the region of Lombardy. By March 8th, 2020, the whole of Lombardy was placed into lockdown and on the following day all of Italy, after which all non-essential businesses were closed and movement of people was severely restricted. This report simulates the spread of COVID-19 in the first weeks in the region of Lombardy. The first exercise fits a calibrated compartmental epidemiological model, the SEIR model, to each region's real case numbers. The second exercise expands the model by a contact-matrix based on age-groups' interactions and the third exercise explores stochasticity related to the spread of the disease and restrictive measures are simulated to slow down the spread of the virus.

II Data

Daily data points of individuals newly-infected with COVID-19 in each of the 12 Provinces of the Lombardy region over a period of 92 days were given. Exemplary, the data of the province of Milano is visualized below (Milano will be used throughout this report as main example). Visualizations and data regarding all other provinces can be found in the Appendix (subsection I).

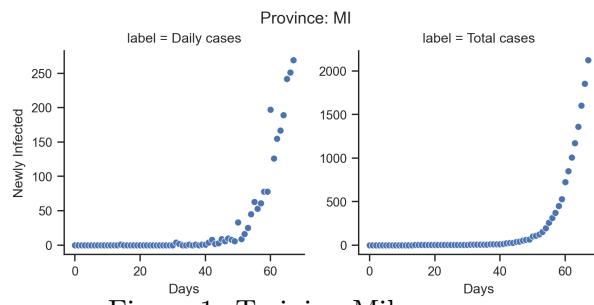


Figure 1: Training Milano cases

III Exercise 1

According to the WHO, COVID-19 has an average time that it takes from the initial exposure to the first symptoms of 5 - 6 days, although this so-called incubation period can range from 1 - 14 days (*Coronavirus disease (COVID-19)*, 2021). Due to this finite infectious period (meaning once infected people do not stay infected forever, like with some sexually transmitted diseases) and lasting immunity after an infection (according to the British Medical Journal reinfections seem to be extremely rare and only happen after long time intervals (Stokel-Walker, 2021), thus they are excluded from this model which only looks at a time period of three months), *SI* and *SIRS* models can be ruled out for this simulation. Therefore, the most suitable compartmental model for COVID-19 seems to be the *SEIR* model (equation 1) which expands the basic *SIR* model by an incubation period during which COVID-19 patients have been infected but are not yet infectious themselves. According to the European Centre for Disease Prevention and Control (ECDC) (*Timeline of ECDC's response to COVID-19*, 2021), this time frame is determined by the beginning of infectiousness which is usually two days before the first symptoms. The *SEIR* model has been shown to be a well-suited compartmental model to simulate COVID-19 in the paper by Prem et al. in 2020. Consequently, a *SEIR* model is adopted to model the spread of the disease that includes the following parameters: the probability of disease transmission in contact between a susceptible and an infectious subject β (from which the force of infection is computed), the incubation rate alpha (rate of latent individuals becoming infectious), the recovery rate γ (average duration of the infection), as well as the birth rate b and the mortality rate m .

$$\begin{aligned}
\frac{dS}{dt} &= bN - \frac{\beta IS}{N} - mS \\
\frac{dE}{dt} &= \frac{\beta IS}{N} - (\alpha + m)E \\
\frac{dI}{dt} &= \alpha E - (\gamma + m)I \\
\frac{dR}{dt} &= \gamma I - mR
\end{aligned} \tag{1}$$

$$\begin{aligned}
\frac{dS}{dt} &= bN_i + \sum_j \beta_{ij} I_j S_i - mS_i \\
\frac{dI}{dt} &= \sum_j \beta_{ij} I_j S_i - mI_i \\
\frac{dR}{dt} &= \gamma I_i - mR_i
\end{aligned} \tag{2}$$

While the birth rate b is assumed to be similar to the mortality rate m (in a three month time window this seems reasonable) and the gamma is determined by the average duration of the disease D (for COVID-19, D is assumed to be 6 days), β and α have to be calibrated to best fit the real data (see Figure 2 as an example for the cases in the first 68 days of Milan, all other provinces can be found in subsection II in the Appendix). As the calibration method, the Mean Squared Error (MSE) was used. The MSE is the average squared difference between the estimated values of daily cases and the actual value found in the data set. This was iterated over a range of $R0s$ (to calculate the beta) and alphas and the model with the lowest MSE was chosen. In the case of Milano, the resulting model had a MSE of 15.44, the $R0$ with the minimum MSE was 3.83, with a beta of 0.6383, and the alpha was 0.09. To validate the results a log-likelihood calculation was applied as well that yielded similar values for all parameters (see Appendix subsection II for all results).

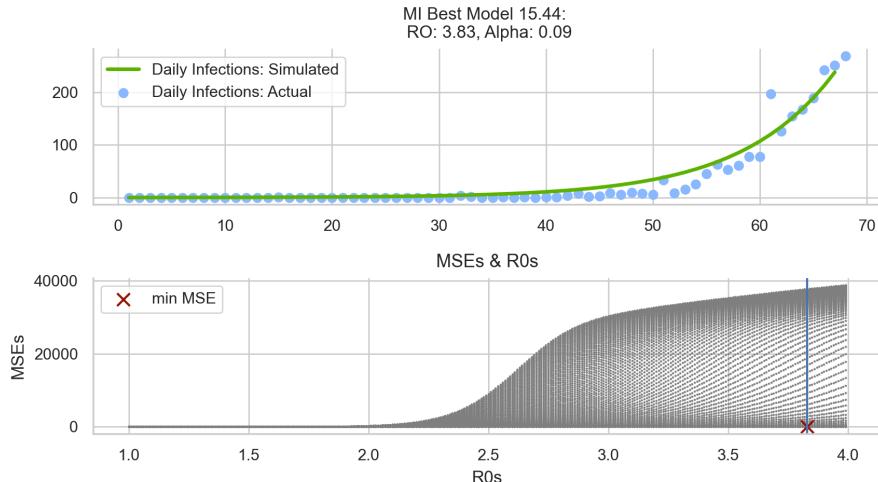


Figure 2: Milano infections

What is clearly visible in the plot is that the model and calibration fit the data very well for the specified days (e.g. 68 for Milano), but afterward the exponential growth in the real-life data suddenly stops, whereas the modeled cases grow further and further. This is most probably due to measures implemented by the local authorities including the first lockdown at the beginning of March which together with measures like mask-wearing, restrictions of gatherings, etc. dramatically reduced the number of newly infected people as we can observe in Figure 3 where the extended timespan is modeled.

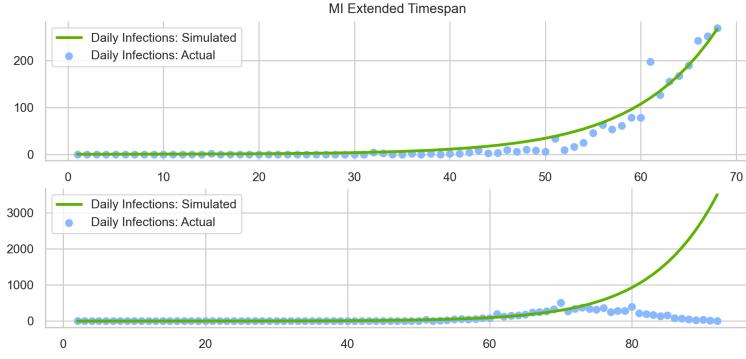


Figure 3: Milano infections

IV Exercise 2

For this simulation, a *SIR* model (equation 2) is used but extended by a demographic component. A contact-matrix divided by sixteen age classes (0-4, 5-9, ..., 75+) indicates the average number of contacts in between each of the age groups. To model the spread of COVID-19 in the province of Milano the precise age distribution of this province was used and then combined with the contact-matrix. The resulting matrix is visualized in Figure 4 where brighter colors indicate more contacts.

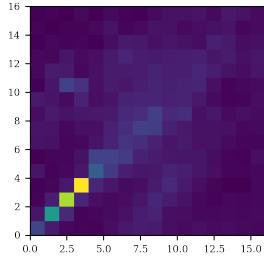


Figure 4: Matrix of average contact per age class

For the calibration of the age-stratified *SIR* model that takes the contact matrix into account, the same assumptions as in the previous exercise were used, but this time only beta needed to be optimized (again MSE was used). All susceptible, infected and recovered individuals were summed up for that task since the infection data that was provided was not stratified by age groups. Details on the calibration can be observed in the Appendix Figure 24. The resulting calibrated model was then used to simulate the spread of COVID-19 in each age group and can be seen in Figure 5.

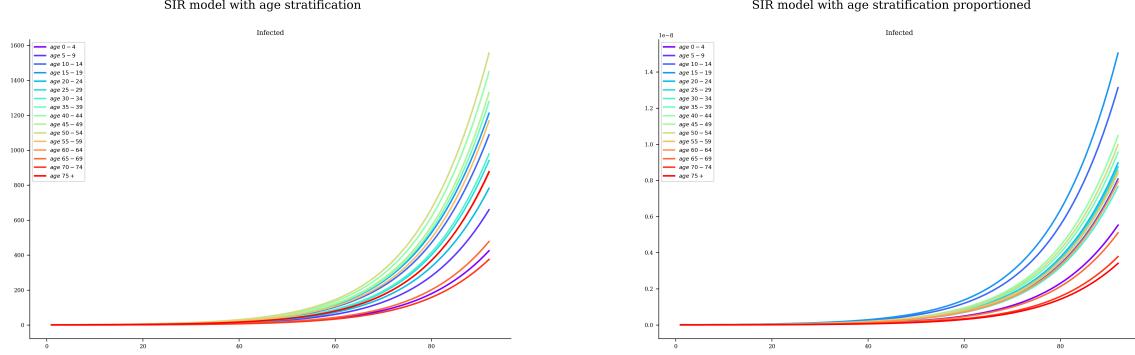


Figure 5: Left: Number of infected per age group, Right: Proportion of infected per age group

As already foreseeable from the visualization of the matrix, the younger age classes, especially age groups 15-19 and 10-14 have the most contacts and therefore the steepest infection rates. This is particularly observable in the proportioned plot, where e.g. age class 70-74's infections rise a lot slower and the time when the infections peak is significantly later compared to the younger ones. The resulting R_0 that is calculated from the eigenvalue of the contact matrix is 1.47.

V Exercise 3

Here, the age stratified-model from the previous exercise is extended with stochasticity meaning new infections do not occur deterministically but according to a binomial distribution with probability p . Thus, a variety of different realistic scenarios can be calculated as a number of simulations are run that all might be slightly different, basically creating best or worst-case scenarios comparable to confidence intervals. The mean from all these simulations with 'shadows' indicating the various simulations can be found in Figure 6. This time a longer period of time was used to better observe the rise and peak of infections.

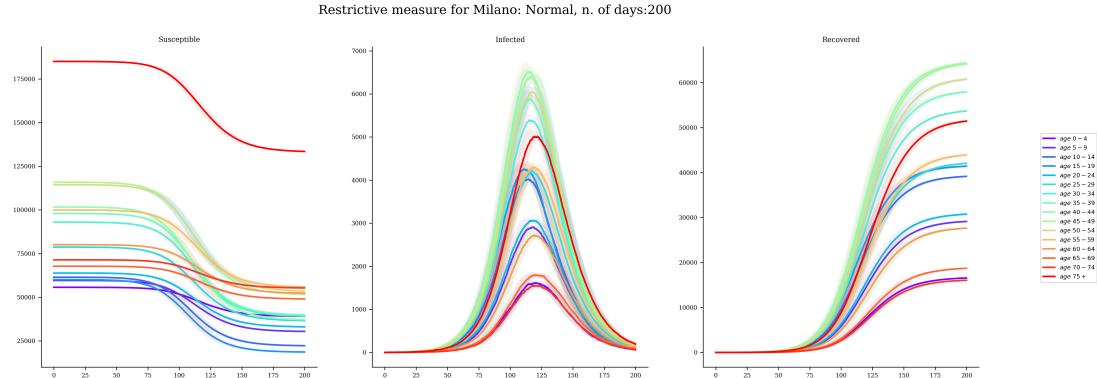


Figure 6: No restrictions

Next, new parameters were introduced that coincide with the following measures: wearing masks & social distancing, school closure, isolation of the elderly, and full lockdown. Each of them can be introduced at a specific time step and will reduce the number of new infections across the specified age groups by a certain percentage ('effectiveness of measure').

The measures introduced are as follows:

- Measure 1 - Mask/social distancing: time: 20, reduces propagation by 15%, all age groups
- Measure 2 - Schools close time: 60, reduces propagation by 75%, age group 5-9,10-14,15-19
- Measure 3 - Isolate elderly: time 30, reduces propagation by 75%, age group 65-69,70-74,75+
- Measure 4 - Full lockdown: time 90, reduces propagation by 60%, all age groups

The first measure introduced is the introduction of masks and social distancing after approximately three weeks (time 20) with a reduction of contacts (effectiveness) of 15% (see Figure 7). Compared to Figure 6 where no measures have been implemented the incline after time 20 becomes less steep and peaks later. This measure does not seem to stop the spread in any age group though, just slow it down.

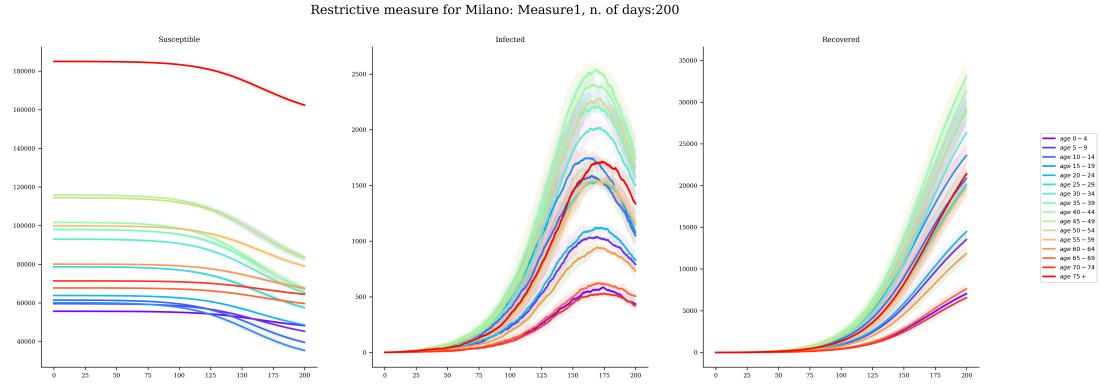


Figure 7: Measures 1

Then the closure of primary schools, middle schools, and high schools was simulated affecting the age groups 5-9, 10-14, 15-19 with an effectiveness of 75%. Graph 8 clearly shows this effect as the age groups infected show a significant decline in infections as soon as the measure is introduced at time 60.

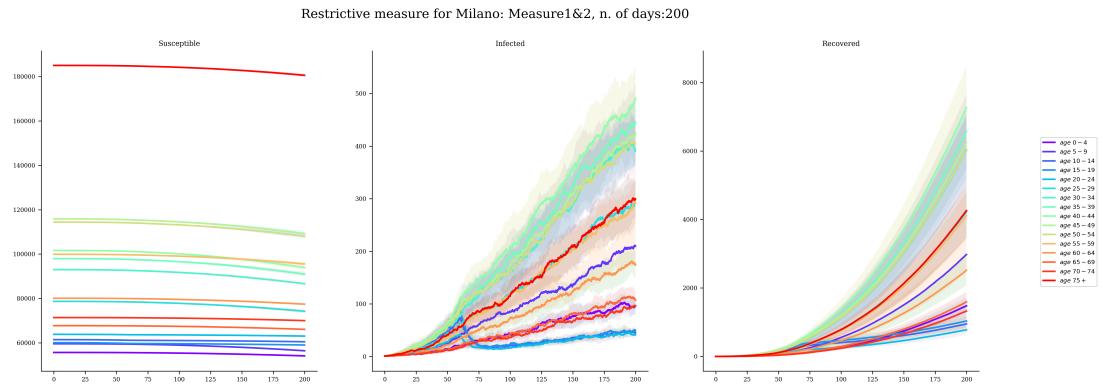


Figure 8: Measures 1 & 2

The same can be shown for elderly people if they were to be isolated and thus would have less contacts in Figure 9.

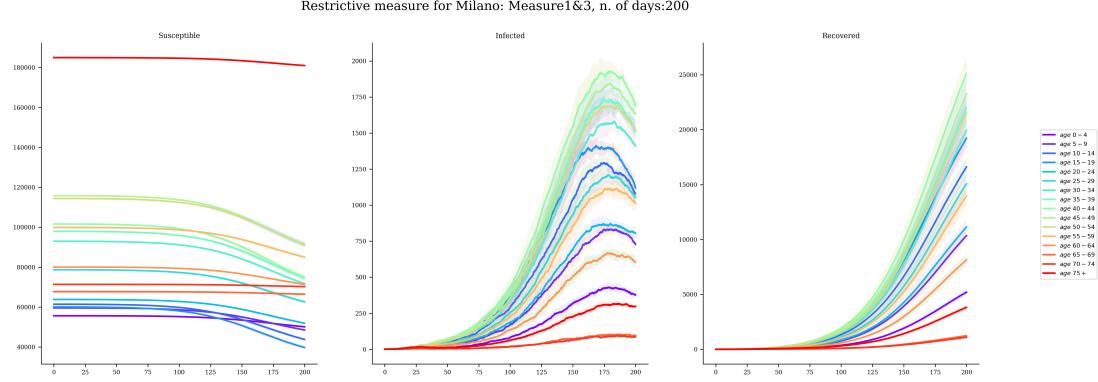


Figure 9: Measures 1 & 3

Finally a full lockdown at time 90 with 60% effectiveness together with all previous measures is simulated in 10. It clearly has the most drastic effects as it affects all age groups, particularly those unaffected by previous measures. It reduces the overall interactions significantly and thus steeply decreases newly infected people.

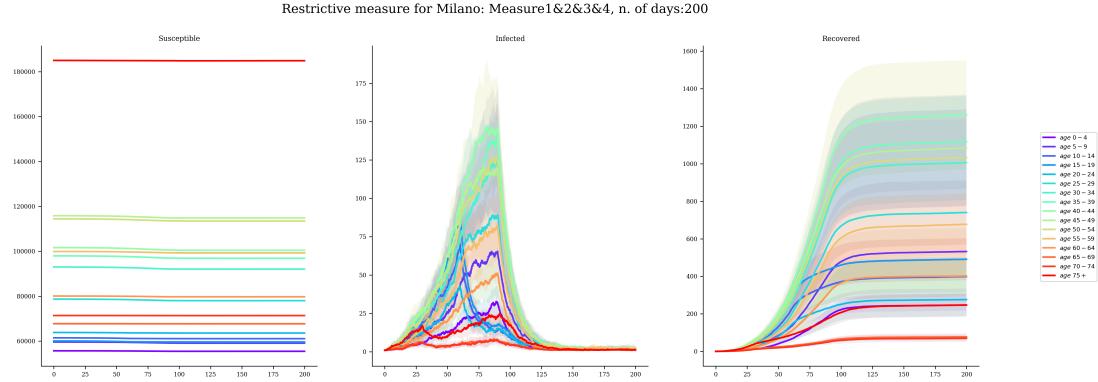


Figure 10: Measures 1 & 2 & 3 & 4

VI Conclusion

COVID-19 has been the most significant global challenge in the past decades. In this simulation, it was shown how even a small number of cases of a highly-infectious disease, driven by exponential growth, can infect entire populations in a matter of a few months. In Exercise 1, the spread of COVID-19 throughout the twelve provinces of the Italian region of Lombardy has been modeled with a *SEIR* model. The simulation shows that if no strong measures after two months had been imposed, dramatically more cases would have been reported. Exercise 2 extended the previous model by age-specific interaction numbers that proved to have a significant effect on the spread of the viral disease as the more interaction an age-group had, the higher its infections rose. In Exercise 3 different scenarios of the spread of COVID-19 were simulated and it was shown how measures, such as wearing masks, school closures, or lockdowns, affect the spread of the virus in the different age classes and the total population. Overall, it can be stated that exponential growth of a highly-infectious viral disease such as COVID-19 seems to be only stoppable with strong measures across all age classes.

References

- Coronavirus disease (covid-19).* (2021). World Health Organization. Retrieved from <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19>
- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., . . . others (2020). The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 5(5), e261–e270.
- Stokel-Walker, C. (2021). What we know about covid-19 reinfection so far. *bmj*, 372.
- Timeline of ecdc's response to covid-19.* (2021, Mar). Retrieved from <https://www.ecdc.europa.eu/en/covid-19/timeline-ecdc-response>

VII Appendix

I All regions data

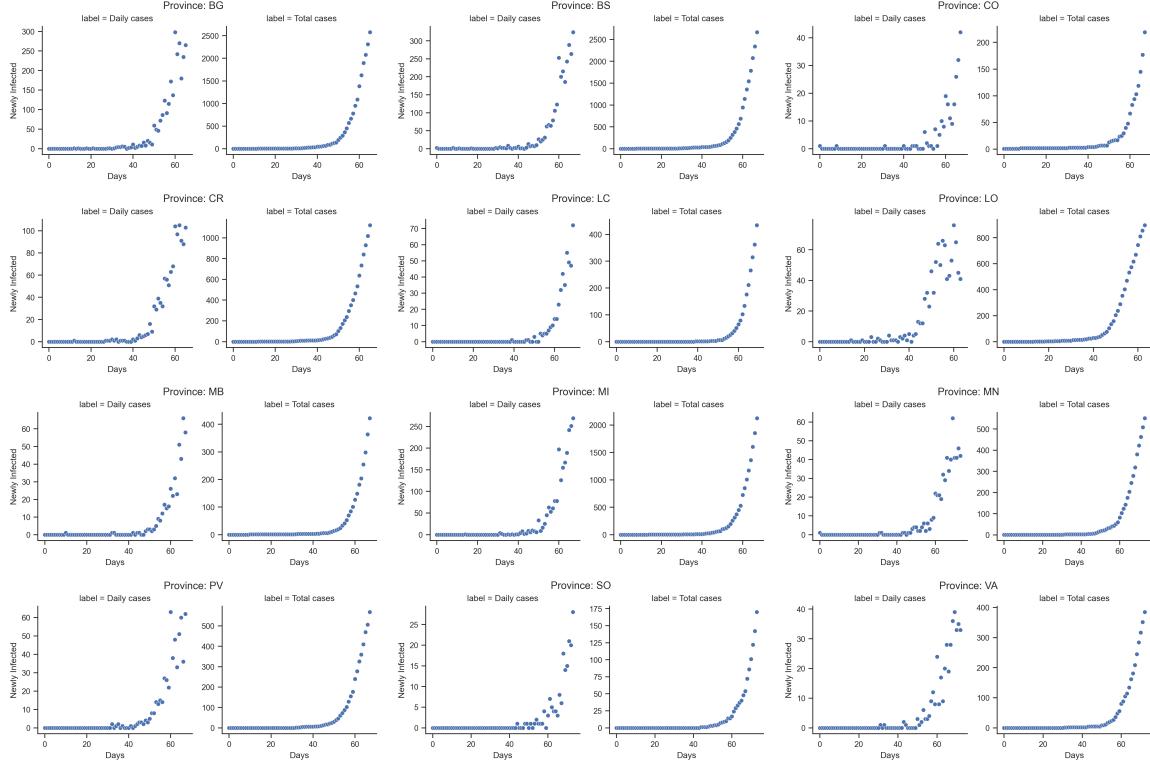


Figure 11: Infected during training period

II Calibration of all regions

Table 1: min RMSE and parameters for each region

Parameters	Regions of Italy											
	LO	PV	SO	MN	MI	CO	BS	MB	LC	CR	BG	VA
R0	1.73000	3.69000	3.99000	3.74000	3.83000	3.19000	3.94000	3.94000	3.84000	3.87000	3.80000	3.59000
alpha	0.96000	0.07000	0.04000	0.06000	0.09000	0.05000	0.09000	0.06000	0.06000	0.08000	0.10000	0.06000
beta	0.28840	0.61541	0.66568	0.62380	0.63869	0.53212	0.65703	0.65715	0.64048	0.64539	0.63367	0.59878
gamma	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667
b,m	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003

Table 2: max MLE and parameters for each region

Parameters	Regions of Italy											
	LO	PV	SO	MN	MI	CO	BS	MB	LC	CR	BG	VA
R0	3.85000	3.99000	3.83000	3.66000	3.77000	3.74000	3.89000	3.79000	3.68000	3.83000	3.99000	3.93000
alpha	0.08000	0.06000	0.04000	0.06000	0.09000	0.05000	0.09000	0.06000	0.06000	0.08000	0.09000	0.05000
beta	0.64206	0.66549	0.63898	0.61045	0.62869	0.62386	0.64870	0.63214	0.61379	0.63872	0.66537	0.65556
gamma	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667
b,m	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003

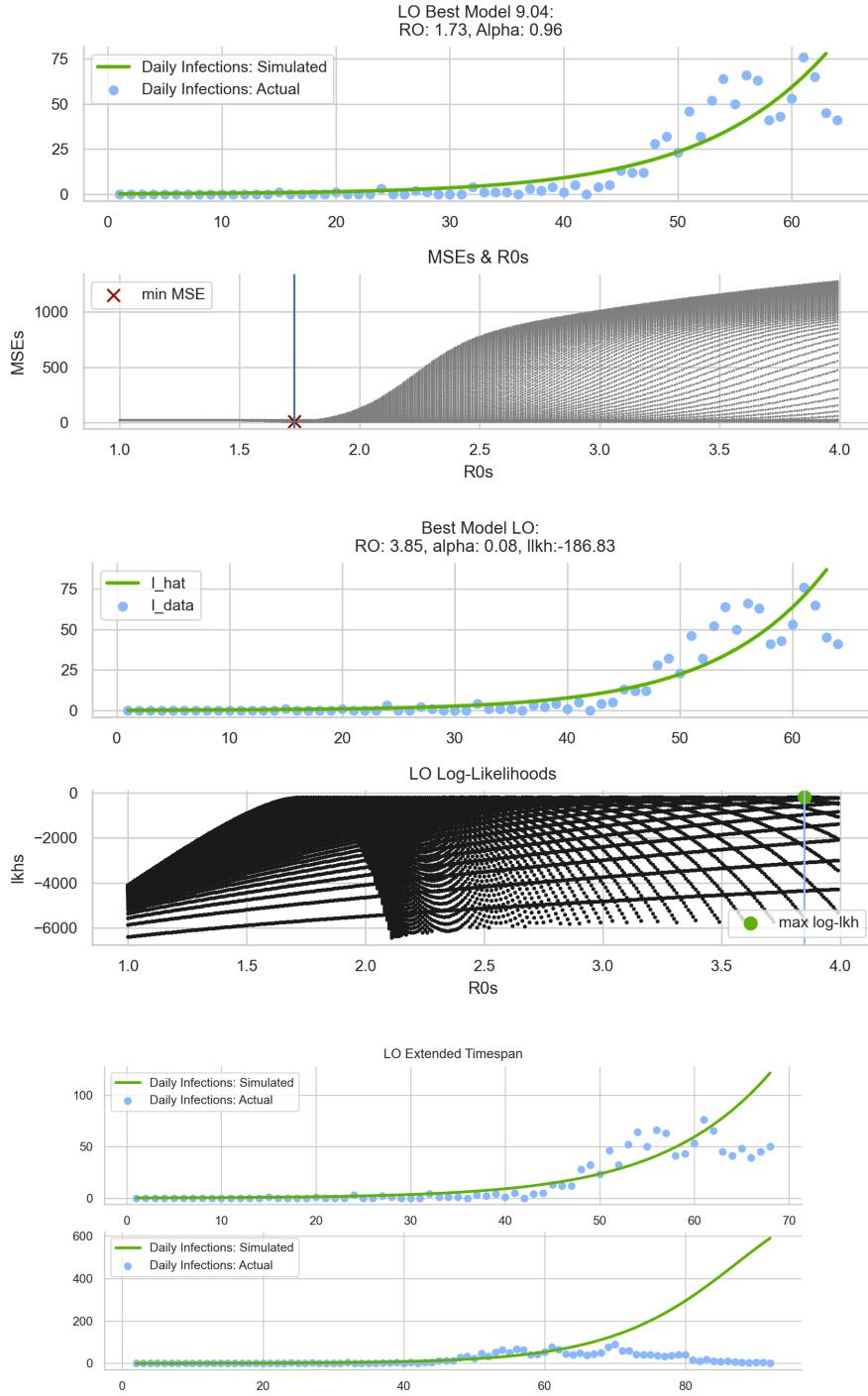


Figure 12: Lombardy infections

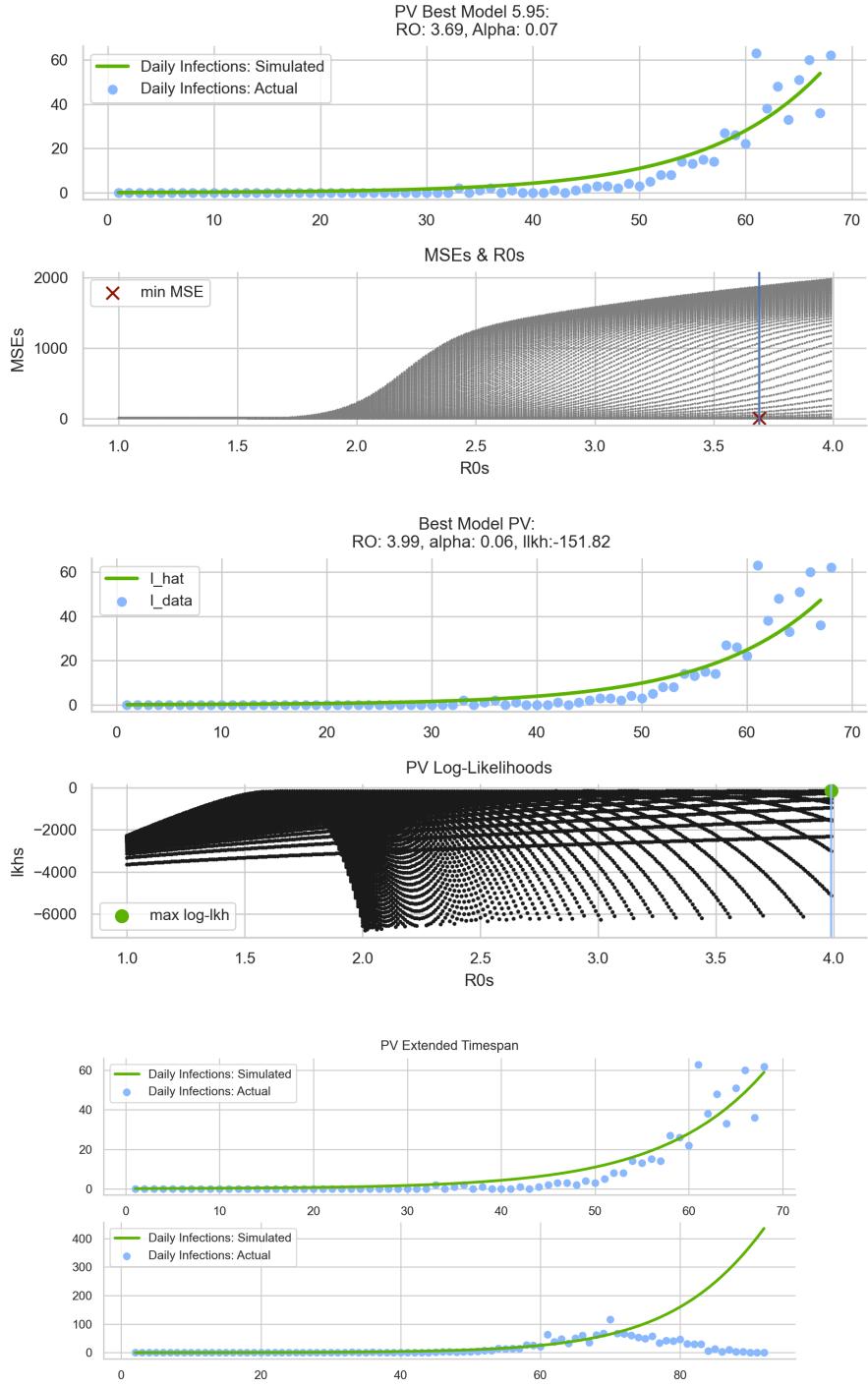


Figure 13: Pavia infections

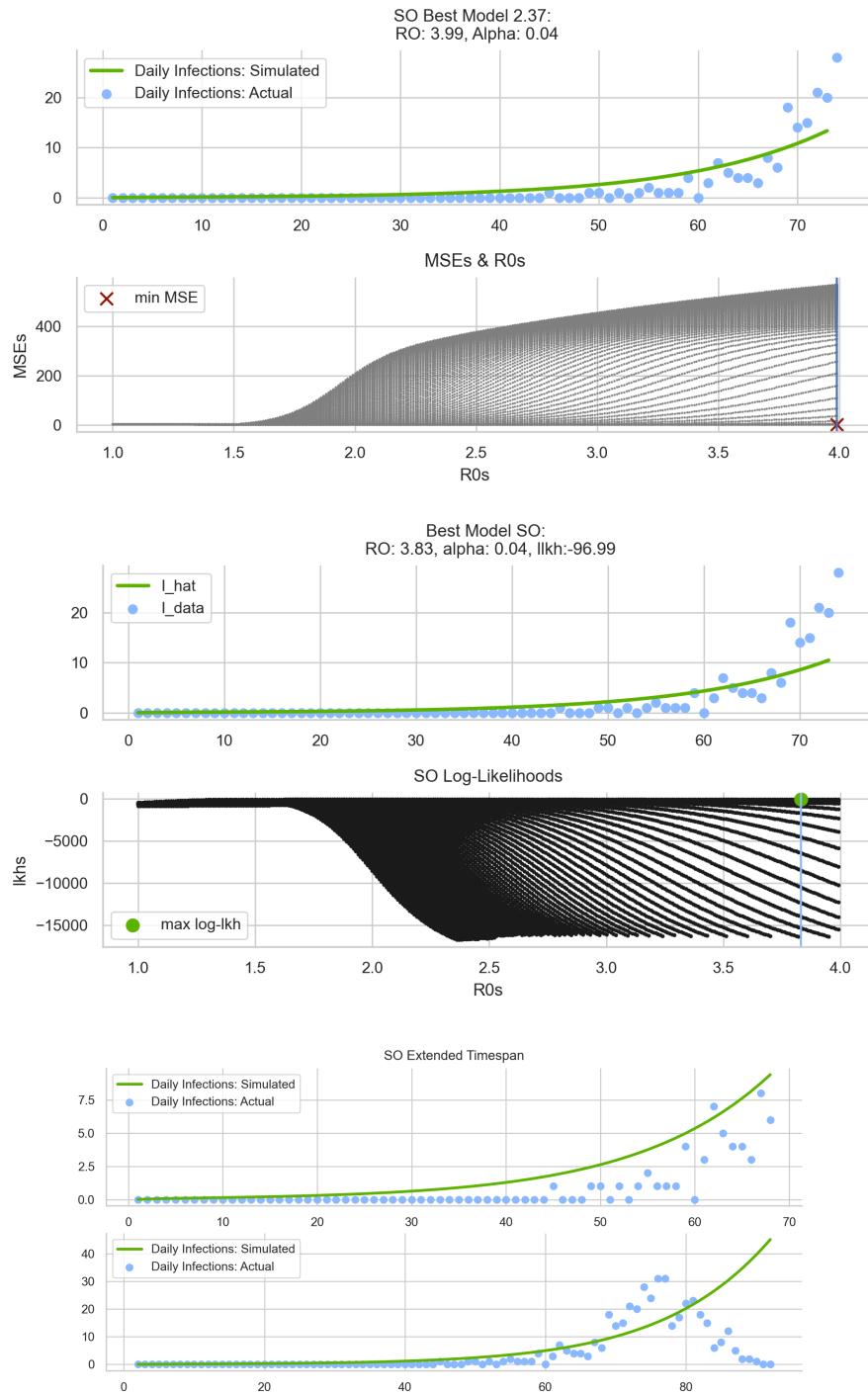


Figure 14: Sondrio infections

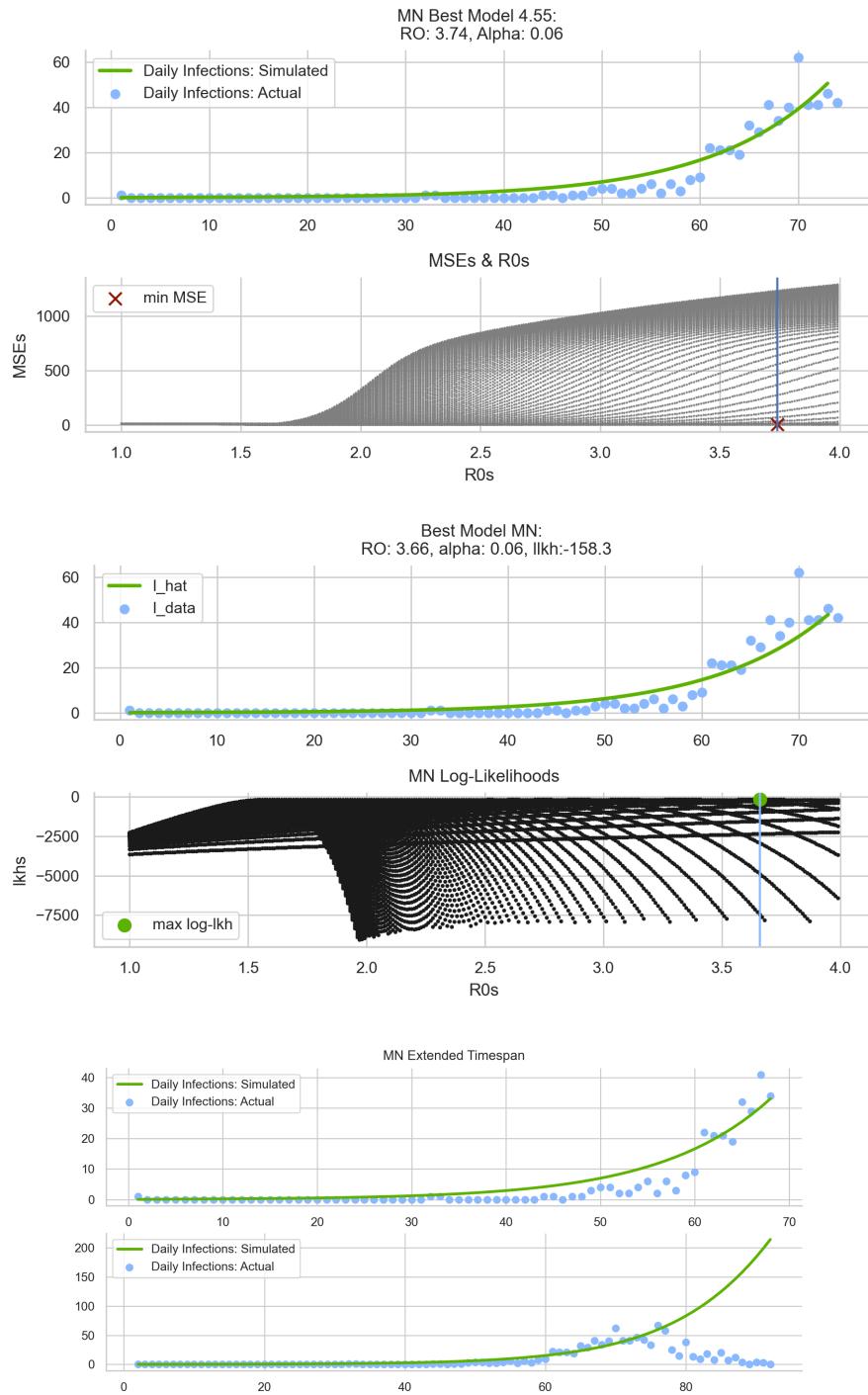


Figure 15: Mantua infections

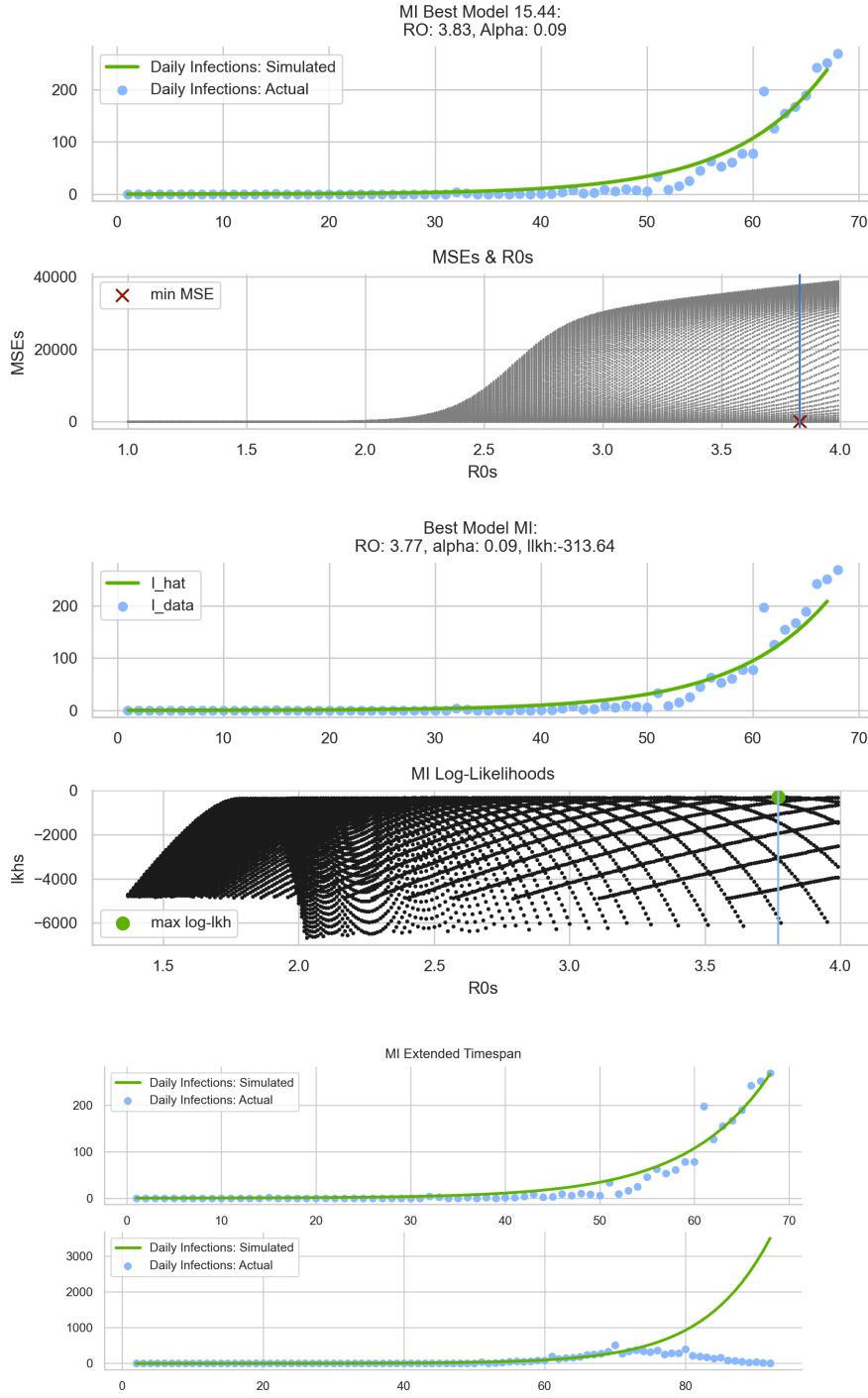


Figure 16: Milano infections

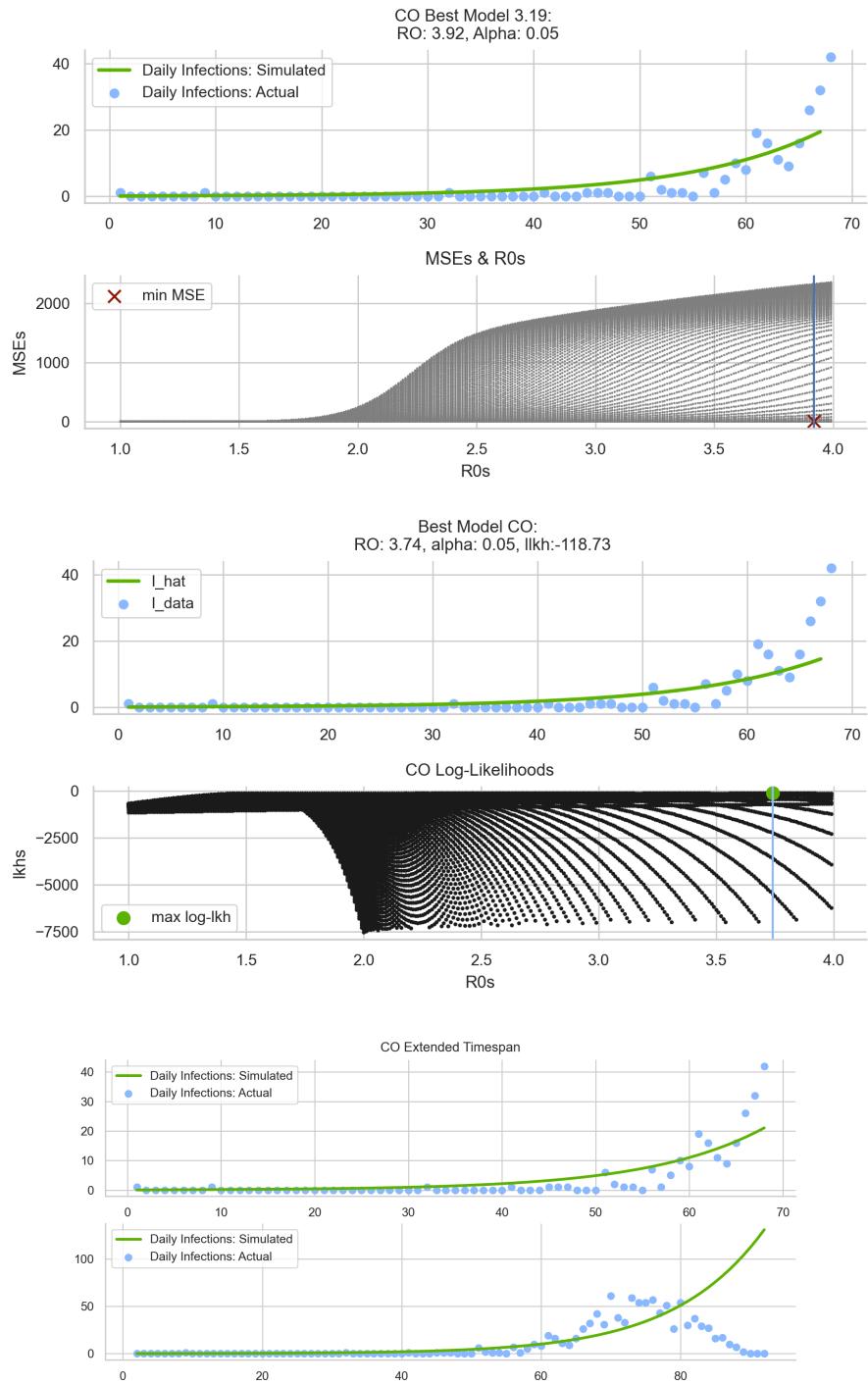


Figure 17: Como infections

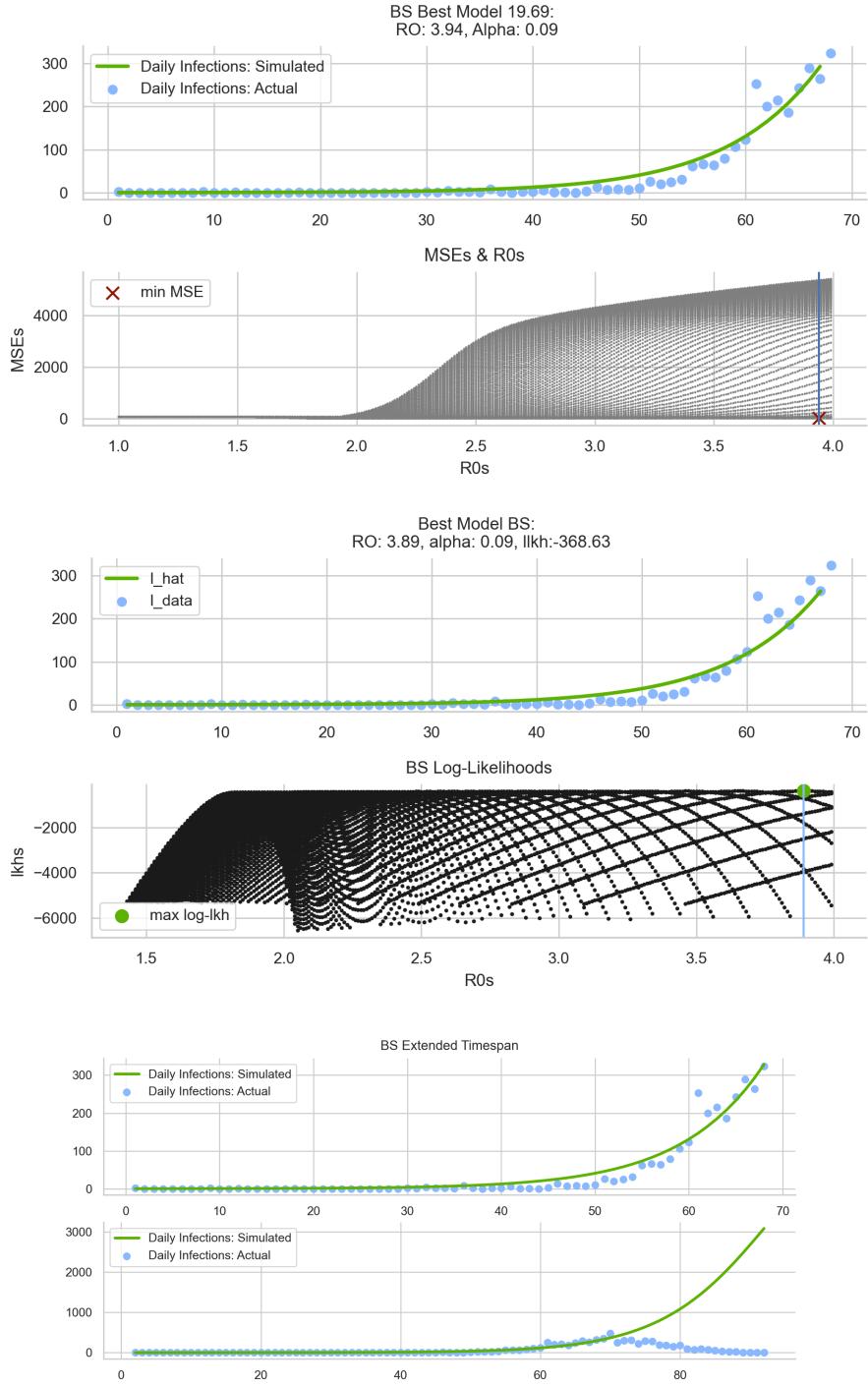


Figure 18: Brescia infections

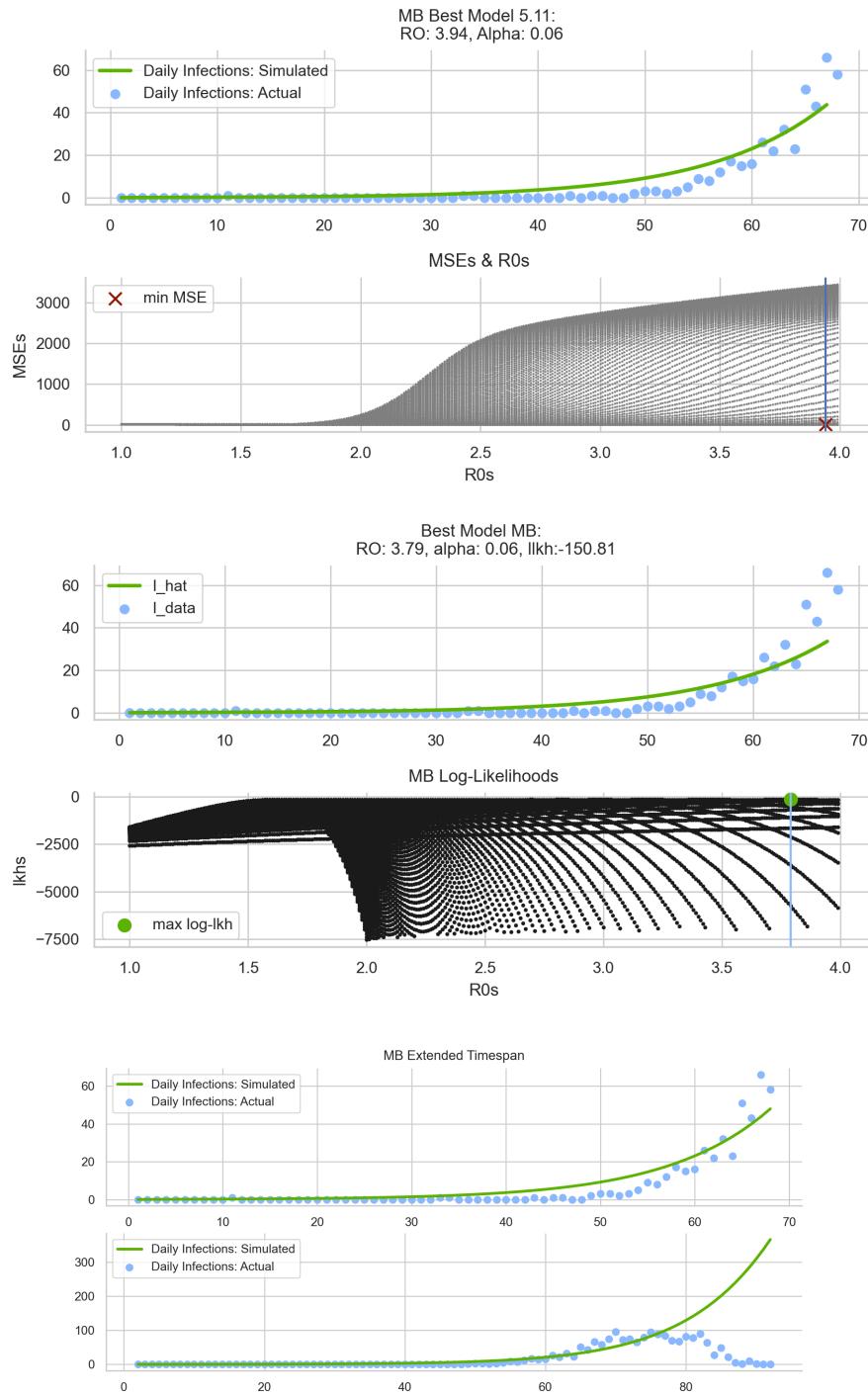


Figure 19: Monza and Brianza infections

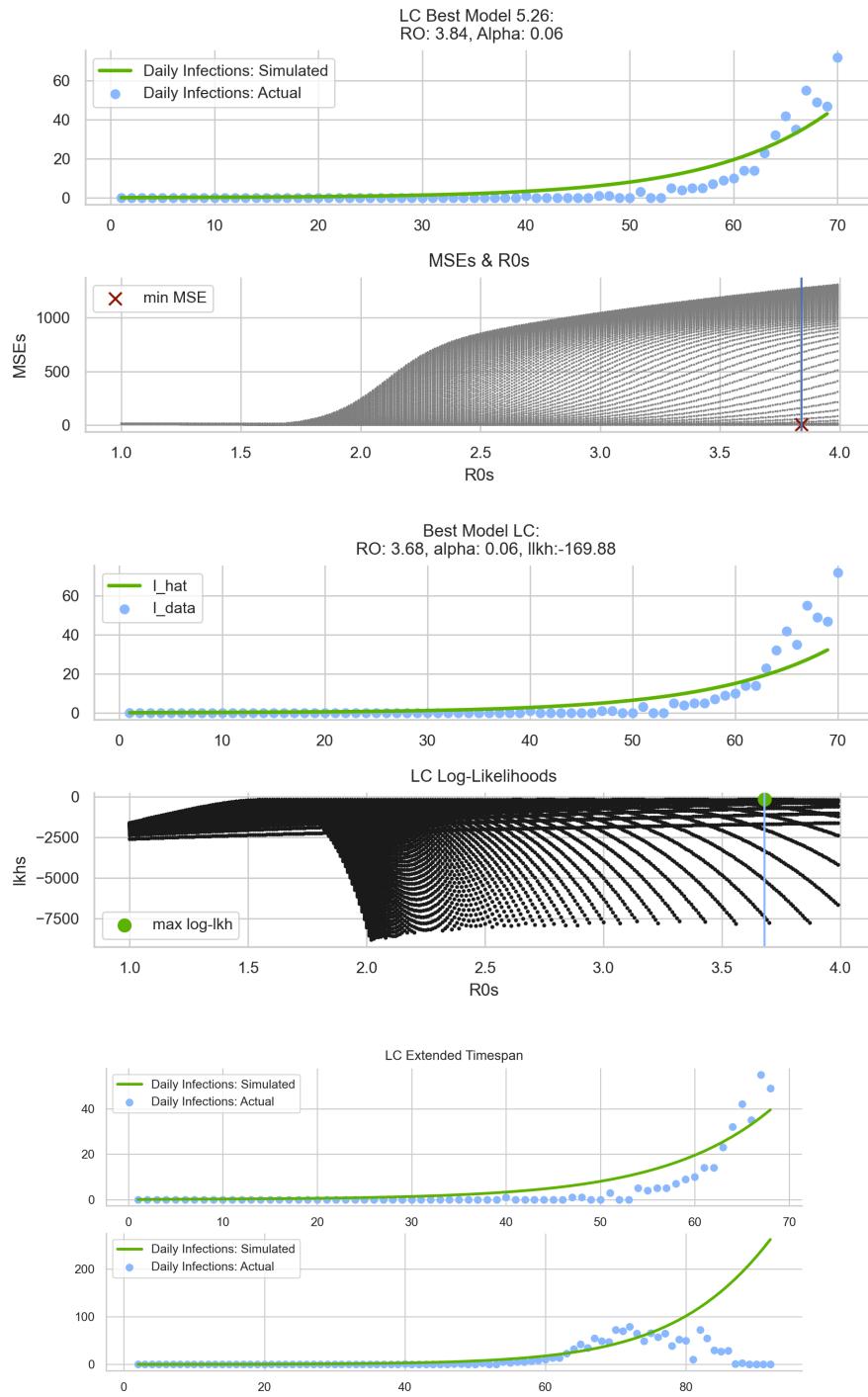


Figure 20: Lecco infections

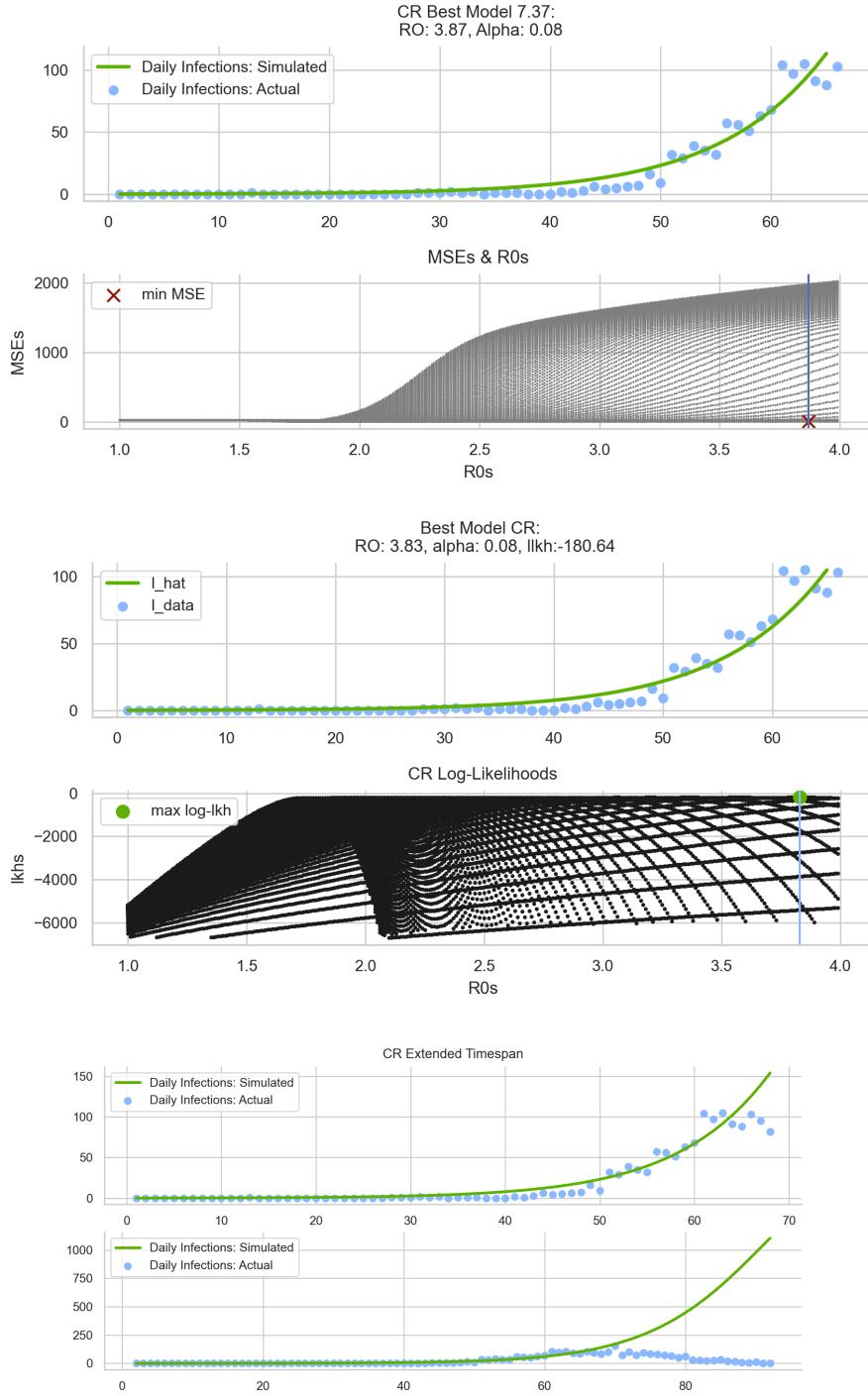


Figure 21: Cremona infections

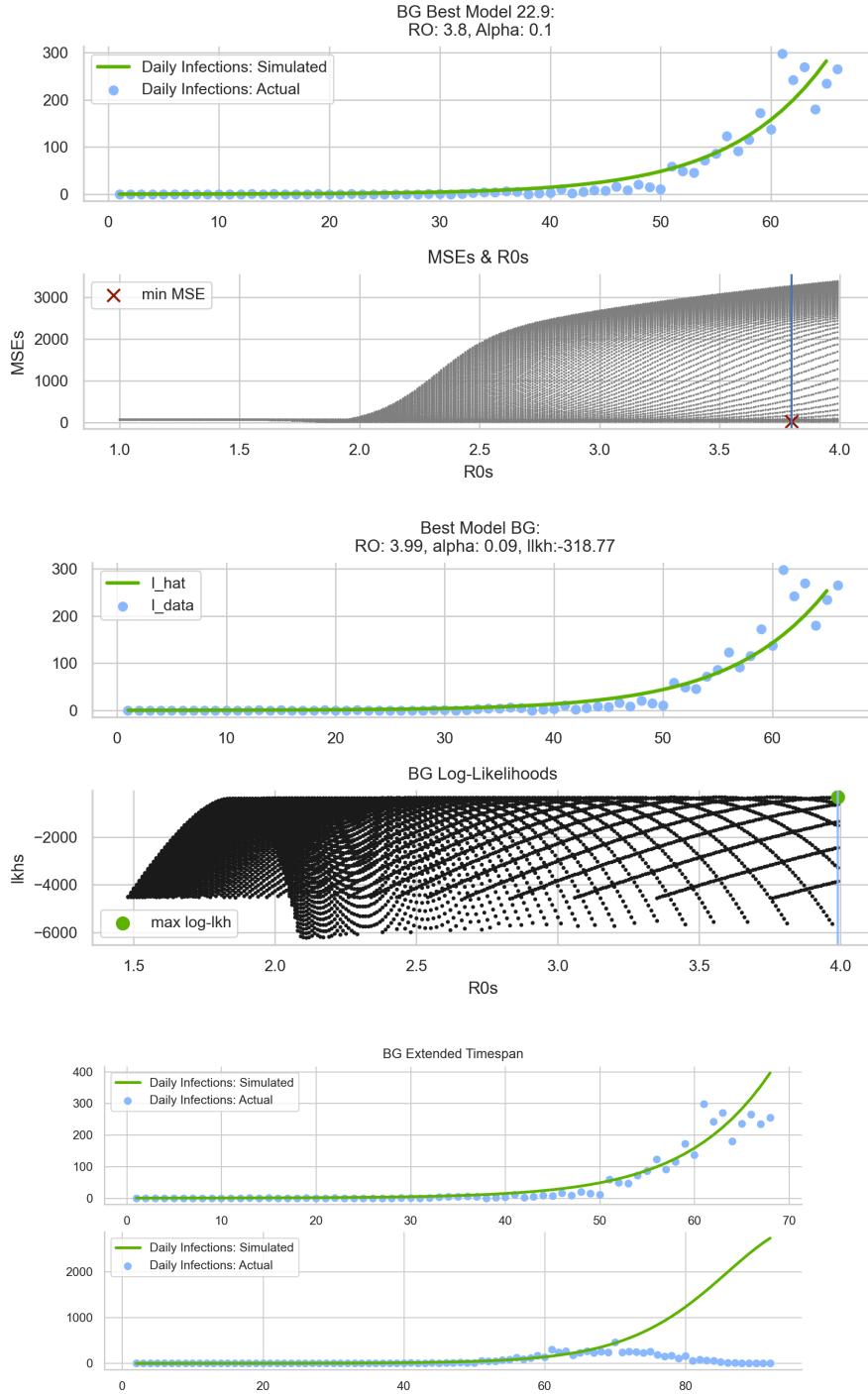


Figure 22: Bergamo infections

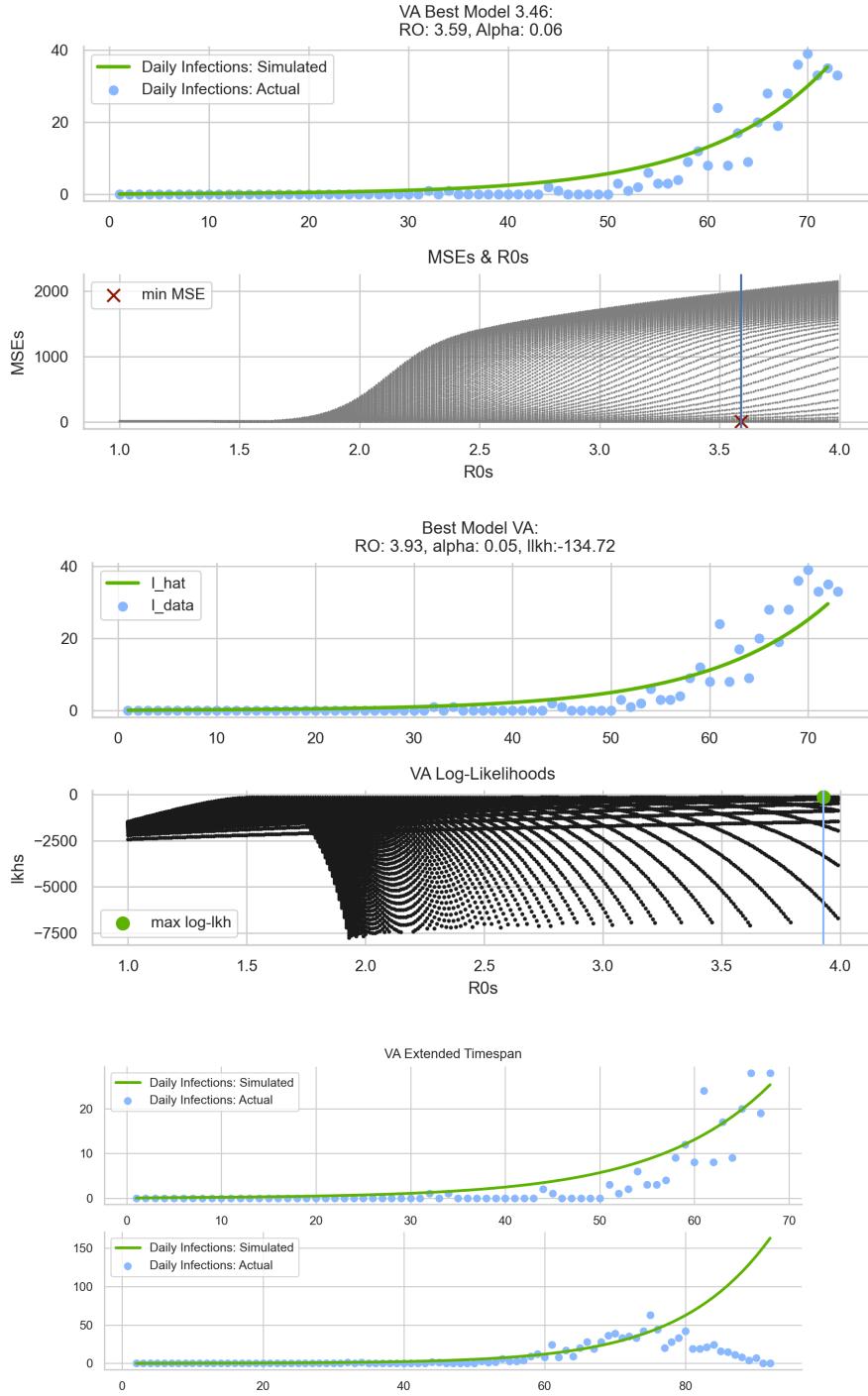


Figure 23: Varese infections

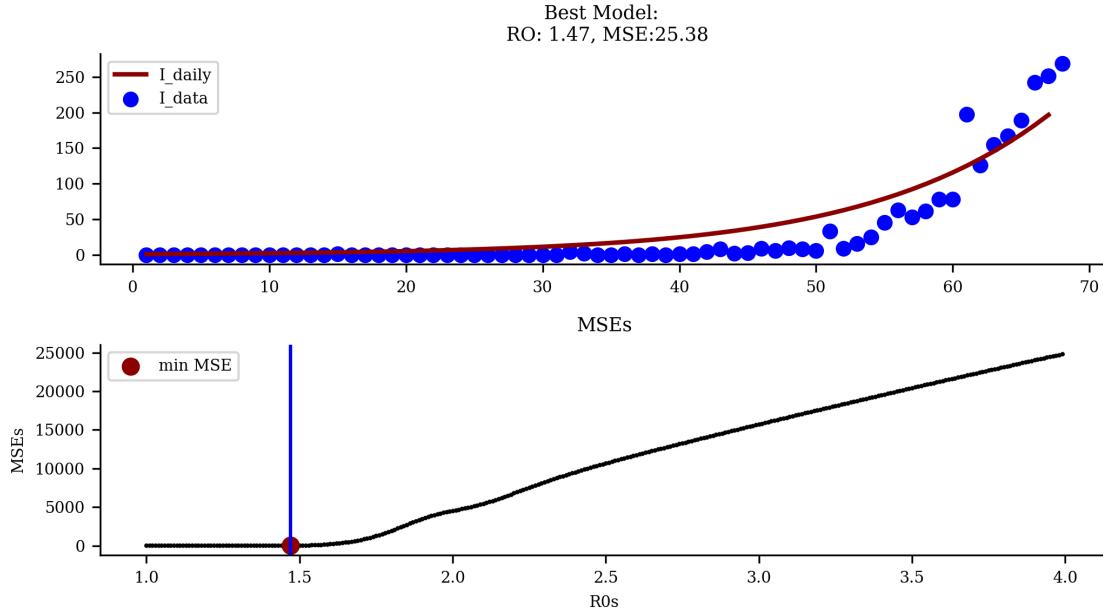


Figure 24: Best model for multi-age

III Other restrictive measures

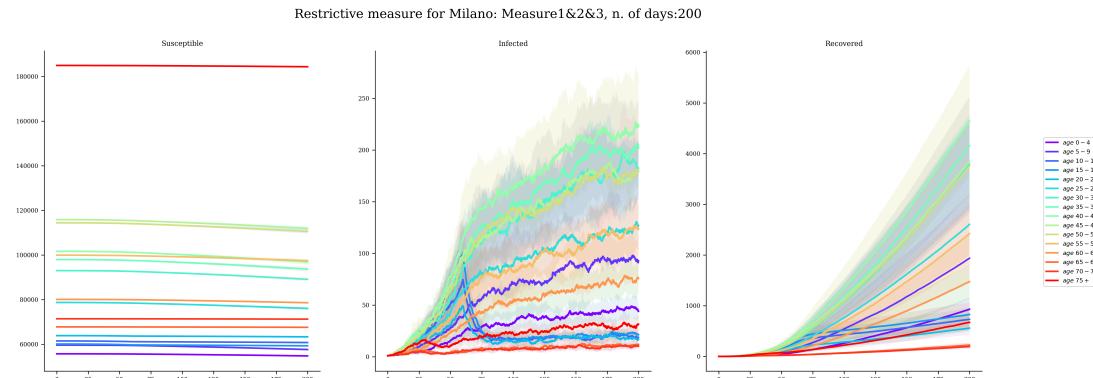


Figure 25: Measures 1 & 2 &3