

# Modeling the Spread of COVID-19 in Lombardy: Adaptation of the SEIQRS Mathematical Framework

Pratheek Tirunagari<sup>1</sup>, Naman Vashishta<sup>2</sup>, and Anshita Bhardwaj<sup>3</sup>

<sup>1</sup> [pratheektirunagari@gmail.com](mailto:pratheektirunagari@gmail.com)

<sup>2</sup> [namannaman.v32@gmail.com](mailto:namannaman.v32@gmail.com)

<sup>3</sup> [bhardwaj.anshi@gmail.com](mailto:bhardwaj.anshi@gmail.com)

**Abstract.** This document uses the mathematical model Susceptible-Exposed-Infected-Quarantined-Recovered-Susceptible (SEIQRS) in the Small-World approximation. This mathematical framework is used to describe the outbreak of the COVID-19 in Italy, in particular in the most affected region, the Lombardy. To achieve a correct description of the data, the model is slightly modified. The interactions between people are modulated against time to mimic the social-distancing. Moreover corrections based on the number of performed tests are implemented. The model parameters are then measured through Profile Likelihood Fit against data of the outbreak in Lombardy. The fitted parameters are the used to create a baseline simulation of the outbreak and to provide estimations of the outbreak features. These features are then also validated against the actual evolution of the virus. The model provides a robust estimation of the evolution in particular of their shape and evolution. Estimations of the absolute numbers are affected by the large errors due to the fit on a limited time (until April 4th).

## 1 Introduction

The document studies and develop the model Susceptible-Exposed-Infected-Quarantined-Recovered-Susceptible (SEIQRS) described in literature [1] in a Small-World approximation. This model classifies the population into five connected categories and studies their evolution in time. Susceptible people consist of people not got in touch with the virus. The exposed category is made of people in contact with an infected person. These people are not contagious and the virus are in incubation. After an incubation time an exposed person becomes infected, becoming contagious but without heavy symptoms. Part of the infected category evolves heavy symptoms and they are hospitalised and quarantined. Those people may recover or die. However a recovered person may not be immune, therefore are still susceptible. Section 2 provides a detailed mathematical representation of the model, elucidating its underlying principles and equations. Given are coupled equations describing evolution of the categories are introduced, following the notation in [1]. The results are compared with the

literature and discussed.

In late December a cluster of this new pneumonia virus COVID-19 has been recorded in China, specifically in Wuhan city, spreading in all the Hubei region [3]. Despite the strict measures adopted in this region to contain the outbreak, the virus has spread worldwide, starting from Europe. The first region strongly affected by the virus was Lombardy, in Italy, where around 1000 people resulted infected on March 1st and currently counting more than 60000 cases [4]. Currently, on April 15th 2020, the most affected area is USA, while Spain is the country counting the largest amount of infected people in Europe [5]. More than two million infected people are in the world, counting, affecting more than 200 different countries with more than 100000 confirmed deaths [5]. The data of the evolution of the outbreak in Italy and Lombardy are reported in Section 3.

The dataset corresponding to Lombardy outbreak is used in examining model performance [4]. The construct has been adapted to fit the adopted social-distancing and testing campaigns. The parameters of the mathematical model are measured against the data through a Profile Likelihood Fit (PLF), helping our estimation of the outbreak evolution (refer to Section 4).

Concluding remarks along with further considerations find discussion in Section 5.

## 2 Model Explanation

In this section, the mathematical framework is described. The equations describing the evolution of the population categories are presented in Section 2.1. The current implementation of the equations are validated against the results in literature in Section 2.2.

### 2.1 Mathematical background

The notation and the model used in this text is taken from reference [1]. In this section, a brief overview, necessary to understand the document, is given. As mentioned in the introduction, the mathematical model is SEIQRS in a Small-World approximation. The population of  $N$  individuals is separated into five classes: Susceptible, Exposed, Infected, Quarantined and Recovered. In the following, variables noted with capital letters (such as  $S$ ,  $E$ ...) indicate the absolute numbers of individuals belonging to the given category, whether quantities given in lowercase letters (such as  $s$ ,  $e$ ...) correspond to fractions of total populations (i.e.  $s = S/N$ ).

The susceptible  $S$  category contain all people that may be potentially infected. They decrease in time, since people get exposed (with contact with an

infected individual). However, healed people not developing immunity pass from the recovered category to the susceptible. The change in time of  $s$ ,  $\dot{s} = \frac{ds}{dt}$ , is:

$$\dot{s}(t) = -\beta \langle k \rangle s(t)i(t) + \gamma r(t) \quad (1)$$

The exposed  $E$  category consist in susceptible individuals got in touch with an infected. An exposed person, after  $\tau$  days, turn into infected. The equation for the variation in time of  $e$ ,  $\dot{e}$ , is:

$$\dot{e}(t) = \beta \langle k \rangle s(t)i(t) - \beta \langle k \rangle s(t-\tau)i(t-\tau) \quad (2)$$

Once an exposed person becomes infected, belonging to  $I$ , it may either die or being quarantined, resulting in the following equation for the variation in time of  $i$ ,  $\dot{i}$ :

$$\dot{i}(t) = \beta \langle k \rangle s(t-\tau)i(t-\tau) - d_1 i(t) - \delta i(t) \quad (3)$$

Quarantined people  $Q$  may die or recover, therefore the variation of the quarantined category fraction  $q$  follows the equation:

$$\dot{q} = \delta i(t) - d_2 i(t) - \mu i(t) \quad (4)$$

Eventually, recovered people may actually turn back into susceptible if they do not develop immunity. Therefore the equation for the variation of  $r$  category,  $\dot{r}$ , is:

$$\dot{r} = \mu q(t) - \gamma r(t) \quad (5)$$

The small-world assumption makes all individuals interact with each other. The number of interactions per day is measured by the  $\langle k \rangle$  parameter. The initial conditions are  $i(0) = i_0 \ll 1$ ,  $e(0) = e_0 = \langle k \rangle i_0$  and  $s(0) = s_0 = 1 - i_0 - e_0$ . The coupled equations are solved with a second-order Runge-Kutta Method. No substantial discrepancy has been observed by the use of the Euler method.

## 2.2 Validation against literature

The obtained results from the solution of equations above are cross-checked with the literature [1,2]. The input parameters are set the same as in Section 4.1 of reference [1], where a numerical example of the model prediction is shown. The specifics of the variables are documented in Table 1.

The evolution of the population categories obtained by solving the set of coupled equation in Section 2.1 is shown in Figure 1a. A clear difference of the evolution of the populations is evident with respect to figures in reference [1], where the outbreak actually occurs with a peak of about 20% of population infected after 35 days. The results with the implementation used in this document, the infected population dies down in about ten days. The motivation is that the initial infected population is not high enough with respect to the outbreak parameter ( $\beta \langle k \rangle$ ) and the fraction of quarantined people.

Values of parameters in literature	
$\beta$	$2 \times 10^{-5}$
$\langle k \rangle$	6
$\gamma$	$2 \times 10^{-4}$
$\delta$	0.3
$d_1$	$5 \times 10^{-3}$
$d_2$	$1 \times 10^{-3}$
$\tau$	5
$N$	$10^4$
$i_0$	$1 \times 10^{-3}$

Table 1: Set of parameters used in Section 4.1 in reference [1,2].

The obtained behaviour of the outbreak is actually expected from the used equations, shown in Section 2. A necessary condition to let the outbreak spread is a positive increase of the infected population, so  $\dot{i}(0) > 0$ . This provides an equation connecting the model parameters and the initial conditions:

$$i_0 > \frac{(\delta + d_1) - \beta \langle k \rangle}{\beta \langle k \rangle (1 + \langle k \rangle)} \quad (6)$$

Therefore, there is a threshold to the initial infected population to make the outbreak occur. From the Equation 6, a higher quarantine fraction and mortality ask for higher initial infected population. Also, since  $\beta \langle k \rangle \ll (\delta + d_1)$ , a smaller  $\beta \langle k \rangle$  asks for higher initial infected population. Setting the model parameters to Table 1, the minumum initial fraction of infected people  $i_0 > 360$ , meaningless, since it is expected to be lower than 1. By imposing this condition, a relation between the different parameters is also found:

$$\beta > \frac{\delta + d_1}{\langle k \rangle (\langle k \rangle + 2)} \quad (7)$$

This condition is necessary to provide meaningful values of the population fractions. Using  $\delta$ ,  $d_1$  and  $\langle k \rangle$  from Table 1, the conditions in Equations ?? are  $\beta > 6 \times 10^{-3}$  and  $i_0 \approx 1$ , that is a limit case. An explanation of the discrepancy with respect to the literature is that probably the  $\beta$  parameter is scaled up of orders of magnitudes according with the initial conditions or the total population. To find a reasonable set of parameters, the  $\beta$  parameter is changed while all the other parameters to Table 1. Since one wants  $i_0 \ll 1$ , then:

$$i_0 < 1 \Rightarrow \beta \gg \frac{\delta + d_1}{\langle k \rangle (\langle k \rangle + 2)} \approx \frac{\delta}{\langle k \rangle (\langle k \rangle + 2)} \approx \frac{0.3}{48} \approx 6 \times 10^{-3} \quad (8)$$

But also:

$$\dot{i}(0) > 1 \Rightarrow \beta > \frac{\delta + d_1}{\langle k \rangle [1 + (1 + \langle k \rangle) i_0]} \approx \frac{\delta}{\langle k \rangle} \approx \frac{0.3}{6} \approx 5 \times 10^{-2} \quad (9)$$

Assuming the difference is a matter of order of magnitude in the  $\beta$  parameter,  $\beta$  is set to 0.2 ( $N$  times the original value). The corresponding evolution is shown in Figure 1b: a similar evolution as in reference [1] is finally obtained.

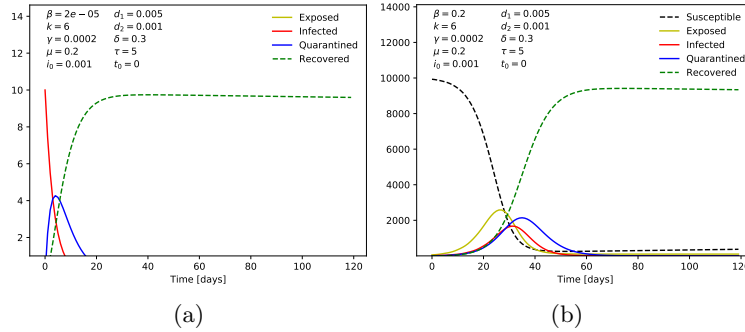


Fig. 1: Numerical prediction of the model with different settings of the parameters. Parameters in Table 1 are used in (a), as well as in (b) but for  $\beta$  parameter set to 0.2.

### 3 Dataset

As mentioned in the introduction, the dataset of the COVID-19 outbreak in Italy is used to test the mathematical model. The Italian spreading of the virus is measured through different quantities, reported in Figure 2a. The numbers refer to people that has been tested for the virus. In the following, a person with a positive outcome of the test, meaning infected by the virus, is also called as a *positive case*.

Tests have been performed on population showing heavy symptoms and hospitalised, but also on the medical staff. In some cases, also close connections to those individuals are also tested. This strategy has been quite restrictive and constant, due to a lack of available test kits. As the Figure 2a shows, a positive person dies or recover. The total cases per day are given by the sum of current positively tested individuals, deaths and recovered.

However, the dataset used to test the model refers are collected only in Lombardy. This region has been the first region where the virus spread and the most affected one. Moreover most of the cases, especially at the beginning, were registered in this region, being the largest contribution in the overall Italian dataset. The data of the total cases versus time in the Lombardy area are shown in Figure 2b. The shown data are extracted from the web-site of *Il Sole 24 Ore* [4], the main economical italian newspaper, that has been one of the first sources of data of the outbreak in Italy divided by regions.

The reason because only Lombardy has been chosen to test the model are several. As mentioned above, it is not restrictive since it is by far the most affected region in Italy. Moreover, sanitary procedures are managed at regional level,

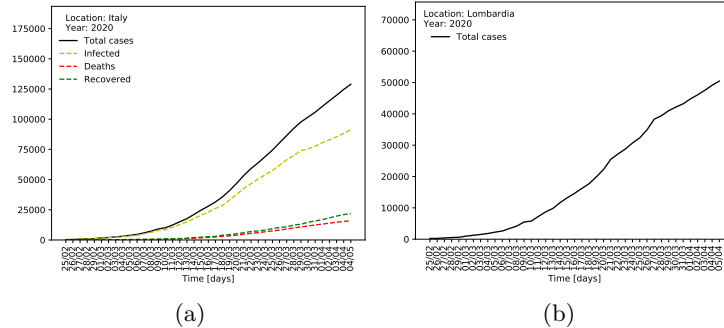


Fig. 2: Dataset collected in Italy (a) and Lombardy (b).

therefore there is a more consistent choice of strategy to contain the outbreak. Finally, it is smaller than the whole country, high-populated, very dynamic and connected area, where small-world approximation is a better assumption.

## 4 Application to the COVID-19 outbreak

The mathematical model described in Section 2 is checked against data of the outbreak in Lombardy in Figure 2b.

The dataset indicates only the total cases, corresponding to positive tested population, deaths and recovered people. In the model, this quantity is estimated by the sum of quarantined  $Q$ , deaths  $D$  and recovered  $R$  categories. The infected and exposed population is here not considered due to the testing campaign adopted in Lombardy. Only people with a chance of having contracted the virus (medical staff, people close to positive cases) and infected people developing heavy symptoms to need hospitalisation. Since exposed and infected categories contain individuals with the virus but with at most mild symptoms, they are not tested and are not included in the numbers of positive cases. The quarantined category contain people getting hospitalised and therefore being immediately quarantined by the medical structure.

Since the plateau of the total cases is not reached in data, the most important parameters to tune to agree with the collected data are  $\beta$ ,  $\langle k \rangle$ ,  $\delta$  and  $\tau$  parameters. On top of the parameters described in Section 2, an additional parameter  $t_0$  has been added to shift the distribution in time. The prediction that matches at most with data, after tuning the model parameters, is shown in Figure 3. Parameters modifying the trailing edge of the infection is set to approximate values of the outbreak. Mortality is set to 2%, reinfection probability to  $10^{-5}$  since yet no cases have been observed. Population is set to  $10^7$ , corresponding

to Lombardy population.

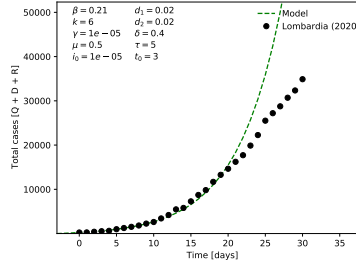


Fig. 3: Best match of parameters to make the model described in Section 2 fit with data il Lombardy.

As shown in Figure 3 the model predicts an exponential growth until the peak, that is not what is measured, since different slopes of curve are present in data. The hypothesis is that the parameters of the model changes in time due to the several restrictions applied to social activity to contain the virus spreading (non-essential activities closed, cancellations of events, etc.). These regulations cause a decrease of the total cases slope, impossible to predict with the mathematical implant of Section 2.

Two main measures to contain the virus have been introduced since March 1st, when the outbreak started to significantly spread. The first strong limitations have been introduced on March 9th [6] and 11th [7], the so-called *#iorestoacasa* laws, forcing non-essential activities to close and encouraging home-quarantine. More stringent laws, prohibiting any movement outside the city of residence and forcing the lock-down except for emergencies, have been released on March 20th [8] and 22th [9]. This regulations are still valid at the time of writing (April 15th). Indeed, around those dates, the slopes of the total cases curves change, confirming that these procedures have been important to contain the spreading of the virus. The mathematical model described in Section 2 is then modified to reflects these changes and to model better the data behaviour.

#### 4.1 Adapting the model to real-world scenario

To adapt the model to the real world scenario, few improvements have been implemented:

- Inclusion of a recovery rate for the infected category evolution.
- Fluctuation of  $\delta$  parameter: this reflects the fluctuation of the number of tests performed per day.
- Scheduling of  $\langle k \rangle$  parameter: this reflects the social-distancing measures.

A scheduling of the  $\delta$  parameter is not introduced. This implies that the fraction of the infected people getting tested is almost constant over time. This assumption is reasonable since the testing strategy has been kept consistent over time. Indeed, the mortality rate, normalised by the number of tests, is quite flat.

The original model does not include a recovery rate for the infected category. Therefore an infected person is only expected to develop symptoms and get quarantined. A person heals only if quarantined before. This is not the real case, the infected people are identified as non-symptomatic contagious infected individuals, and they can recover from the virus without being hospitalised, tested and quarantined. Therefore the variation  $\dot{i}$  is modified as follows:

$$\dot{i}(t) = \beta \langle k \rangle s(t - \tau) i(t - \tau) - d_1 i(t) - \delta i(t) - \mu i(t) \quad (10)$$

This part of recovered people are not counted in the recovered category.

Corrections of the  $\delta$  parameter over time are derived to account for the non-constant rate of performed test and shown in Figure 4. A functional model has been fitted to the data points (taken from reference [4]) and the corrections are computed as ratio of the observed points with respect to the function. The function is consistent with the choice of a constant  $\delta$  since it approximately describe the data trend in Figure 2b.

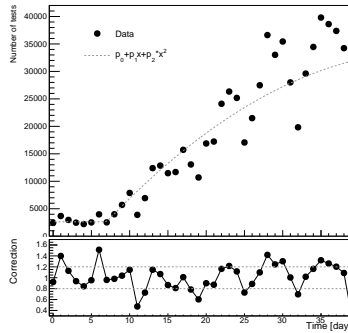


Fig. 4: Numbers of performed tests over time in Italy. A constant plus a third degree polynomial is fitted and corrections (plot below) are extracted as a difference of the measured points to the functional model.

This correction is introduced to describe the non-smooth behaviour of the curve. However, since this data are available national-wise, the numbers for Lombardy may differ. The potential inconsistency is considered and analyzed inside Section 4.2.



The  $\langle k \rangle$  scheduling represents the introduction of social-distancing, varying from a freely-communicating system (where  $\langle k \rangle$  is set to the initial value) to the lock-down situation in which an infected person interact only with few susceptible individuals (family).

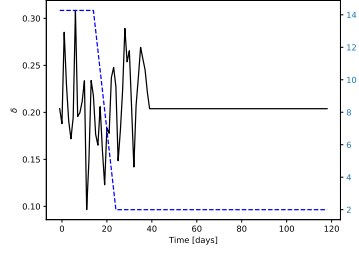


Fig. 5: Variations of  $\delta$  and  $\langle k \rangle$  parameters over time. The reference values of the parameters are set the same as in Figure 3.

The values of  $\delta$  and  $\langle k \rangle$  parameters over time is shown in Figure 5. A new set parameters is tuned to approximately match data shape. This new tuned prediction is shown in Figure 6a, while the comparison with data is presented in Figure 6b.

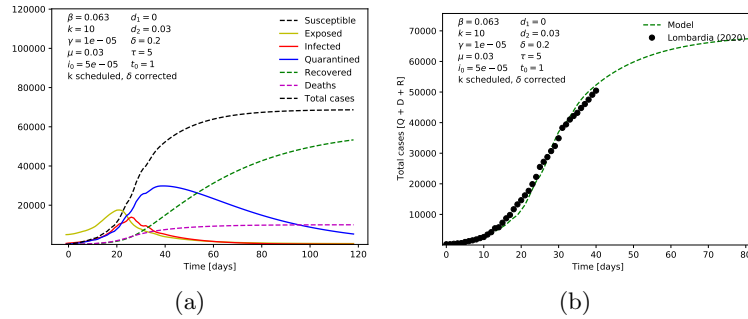


Fig. 6: Prediction of the improved model for the population classes with the correction with a new set of tuned parameters (a). Comparison of the model in (a) with the data collected in Lombardy (b).

## 4.2 Likelihood Profiling

Model parameters are measured with a Profile Likelihood fit. The fit is performed using ROOT framework [11], in particular with the *HistFactory* framework to create statistical models [12].

The observable used in the fit is the daily number of total cases (deaths, quarantined and recovered) to use uncorrelated quantities for each bin of the fitted distribution. The distribution is rebinned to increase statistics per bin. The Parameter of Interest (PoI) is a global normalisation factor of the distribution, introduced as a free floating parameter. This corresponds to a variation of the initial condition,  $i_0$ , that is not included in the set of parameters to fit. The profile likelihood fit optimizes the model coefficients to maximize the probability of the data, given the assumed model [12].

Since no functional model is available, templates of the total cases distribution are generated for various model settings and used in the fit. For each parameter of the model, two templates of the distribution are generated, corresponding of increasing and decreasing of a 10% the parameter value. A nuisance parameter (NP) of the model, with Normal Gaussian prior, is associated to each model parameter used in the fit. The variation of each template from the nominal predictions given by the variation of a single parameter is associated to the variation of the associated NP, whose  $\pm 1$  standard variation correspond to the two differences of the templates in input. For middle values of the NP, an interpolation is used bin-by-bin, to get the corresponding effect of the parameter on the distribution. An example of the effect of the variation of a model parameter is shown in Figure 7a. To improve the fit stability the effect of the NPs are symmetrised against the nominal prediction. In case both variations are on one side, as in Figure 7b, the variation is one-sided symmetrised.

To give enough degrees of freedom to the fit, additional NPs are used to take into account uncertainties on the  $\delta$  correction (refer to Section 4.1). From Figure 4, a variance of 20% is estimated, and a NP for each bin is added, assuming that the variations are uncorrelated bin-by-bin. The example for the two highest bins are shown in Figure 8.

The pre-fit prediction is shown in Figure 9, The data demonstrates a positive correspondence with the model within the relatively large uncertainties. The summary of the fitted model parameters and their correlations are shown in Figure 10. Figure 10a shows the pre-fit NPs value, equal to zero, following a Normal Gaussian distribution. The points indicate the post-fit values and associated uncertainties. The uncertainties on the parameters are constrained, meaning that the assumed systematics model is conservative and the NPs are measured with a higher precision than the assumptions. This is acceptable, since the used 10% uncertainty on the parameters is set arbitrarily, choosing the smallest uncertainty range for all the parameters that cover the discrepancies of the model

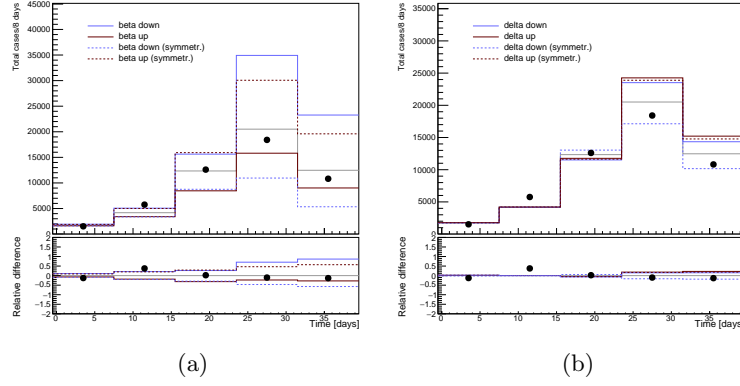


Fig. 7: Two examples of NPs included in the fit: variation of  $\beta$  (a) and  $\delta$  (b) parameters. In the  $\delta$  parameter case, a one direction effect of the parameter is expected. A higher value of  $\delta$  increases the quarantined people per day. A lower value of  $\delta$  instead makes the spreading faster, leading to an increase in the infected and subsequently the quarantined people. The effect of the symmetrised variation is depicted in the provided charts.

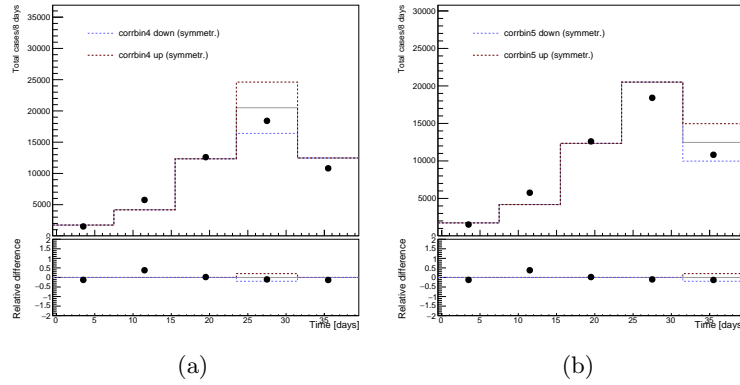


Fig. 8: Variations associated to the test-based  $\delta$  parameter corrections on the most significant bins in terms of the shape of the outbreak spreading.

with data. Large pulls, over one standard deviations are observed in just two NPs, corresponding to the corrections on the bin contents. This is acceptable as well since the uncertainty on those two parameters are estimated assuming the Gaussian fluctuations on the fitted model in Figure 4. In case the Lombardy tests are way different from the country-wise data in some periods, this assumption is not valid and pulls may appear. The parameter of interest, denoted as  $SF$ , is expected to be unitary and still fitted consistently to the expectation within one standard deviation. Correlations between parameters are understood: the scale factor,  $\beta$ ,  $\tau$  and the bin corrections are the most correlated, since they are the most important parameter to change the distribution shape in the fit range. The correlations are coherent with the expectations since they tend to counter-balance the different effects among themselves to agree with the data distribution.

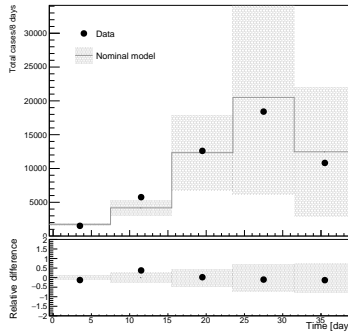


Fig. 9: Pre-fit prediction of the model with the corrections described in Figure 4.1 and the parameters in Figure 6a. The total uncertainty is given by summing all the sources of uncertainties described in Section 4.2

The post-fit prediction is shown in Figure 11. The uncertainties are shrinked at higher bins, while enlarged in the lower bins because of the pulls and shows good agreement of the model with data within the fitted uncertainties.

### 4.3 Predictions

The fitted set of parameters is used to predict the evolution of the population categories evolutions, as shown in Figure 12a. In Figure 12b the evolution in time of  $R_0$  is shown.  $R_0$  stands for the infection power of each single individual. Here, it is estimated by calculating the ratio of the quarantined and infected (or solely quarantined) population at  $t$  in time to that at another point in time  $t - \tau$ , where  $\tau$  represents the incubation time.

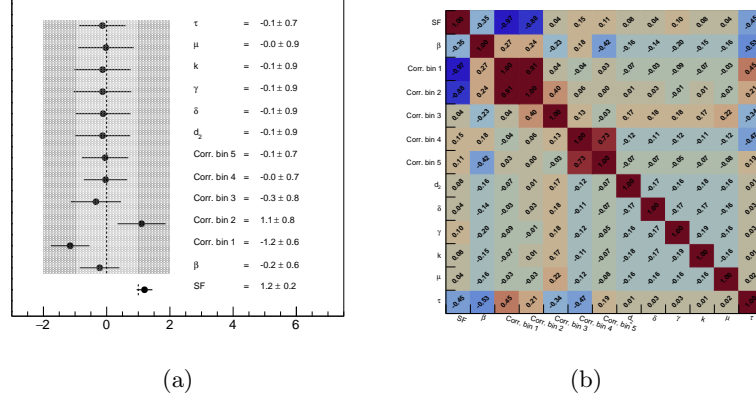


Fig.10: Fitted values of the model parameters (a). Correlations between the fitted parameters (b).

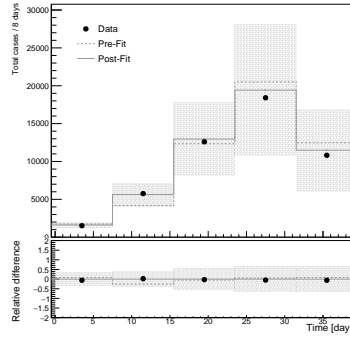


Fig.11: Post-fit prediction of the model with the corrections described in Figure 4.1, with parameters pulled as in Figure 10a. The total uncertainty is given by summing all the post-fit uncertainties, correlated as in Figure 10b.

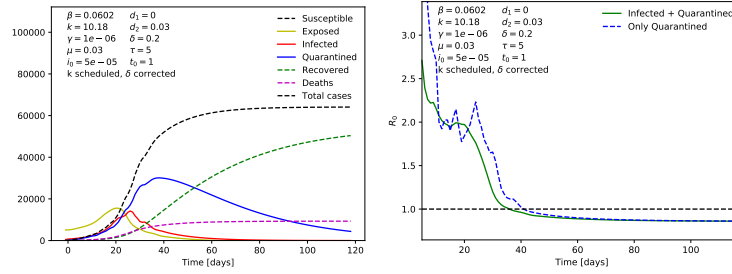


Fig.12: Prediction of the model tuned with the fitted parameters in Figure 10a. Prediction of  $R_0$  parameter against time (b).

The fitted model is used to predict some features of the virus spreading and to test them against the observations. To estimate the uncertainties on these estimations, a set of 1000 pseudo-experiments is generated, obtained by extracting values of the model parameters from the fitted distributions, taking into account the fitted correlations. Predictions on a given observable are made by estimating mean and variance of the generated distribution from the pseudo-experiments. Since the fit is performed on the total cases distributions, predictions on the total cases related observables are more reliable. Given that the number of quarantined individuals plays a crucial role in shaping the total cases, especially within the range of the fitted data, the estimates of observables related to quarantined individuals must possess strong reliability and consistency. The predictions on the number of deaths may be affected by the small contribution to the total cases in the fitted range, and therefore not very reliable. However, in the initial tuning (Figure 6b), the ratios between the contributions of the different classes is taken into account, so the obtained numbers should be in the same order of magnitude of the observations. A summary of the predictions on some observables of the outbreak evolution is given in Table 2.

Quantity	Value
<i>Quarantined people</i>	
Peak	$36000 \pm 21000$
Peak time	$38 \pm 1$ days
<i>Deaths</i>	
Total deaths	$13000 \pm 9000$
<i>Total cases</i>	
Total	$88000 \pm 67000$
Saturation time	$56 \pm 7$ days

Table 2: Predictions of the model on some observables of the outbreak. The model is tuned with the fitted parameters in Figure 10a. Dates are expressed in days after February 27th, the first day of available data.

Currently (on 16/04/2020), from the updated data on reference [4], there are 62153 total cases, with 11377 deaths. Both numbers are compatible with the predictions within the large uncertainties, but at least in the correct order of magnitude. The model predicts the peak of quarantined (positive) cases on April 2nd, however not a simple comparison with data is possible due to a lack of data documentation. The observed peak of daily new total cases is observed between March 21st and 26th, 2020: this is the highest slope of the total cases distribution against time, so expected earlier than the peak of quarantined population. A first decrease of total intensive in Italy care has been observed on

April 4th, withing 2 standard deviations from the expectation. A comparison in terms of  $R_0$  parameter is also possible. The  $R_0$  evolution, shown in Figure 12, can be tested as well. Considering the  $R_0$  estimated only using quarantined population yields, an  $R_0 < 1$  occurs around the 42th day, so on April 6th. The first indication of a  $R_0 < 1$  in Italy is reported on April 9th [14], quite close to the expectations. Unfortunately both previous data are available only national-wise and not for Lombardy only, so results may differ. Finally, the estimated time of saturation of the total cases, corresponding to reaching the 90% of the maximum total cases, is expected in the week around April 20th.

## 5 Conclusions

The mathematical model of SEIQRS [1,?] with a Small-World approximation for describing outbreak evolutions is here implemented and studied. Results present in literature are here validated and cross-checked.

The model is used to describe the COVID-19 virus spreading in Lombardy, Italy. Since March 1st, Lombardy's population is heavily hit by the spreading of this virus, causing thousands of deaths, and tens of thousands of hospitalised people. The Italian government has deployed strong laws to limit the sociality and contain the contagion. The model presented above is tuned and modified to better fit the Lombardy scenario, and describe better the collected data about the COVID-19 outbreak in the region.

A Profile Likelihood Fit is used to measure the model parameters against the collected data. This provides measurements for all the parameter, with associated uncertainties and correlations. The fit results for the parameters are used to generate a set of pseudo-experiments and estimate some features of the outbreak evolutions. The obtained predictions are compatible with data collected in Lombardy so far. Moreover the model predicts a saturation of the total cases of COVID-19 in Lombardy within a week around April 20th.

The Italian government strategy is to prolongue the lock-down measures until May 3rd [10]. The model used for the estimations above is used to estimate the effect of finishing the lock-down on that date. This is done by resetting the  $\langle k \rangle$  parameter to the initial value on May 3rd in the model. The prediction is shown in Figure 13a. The prediction for a gradual reopening over six months is instead shown in Figure 13b. Since the levels of exposed and infected people do not go below the threshold for the outbreak (see Section 2.2), the outbreak occur again when the social interactions are restored to the nominal values. Therefore, without a vaccination campaign that could stop the contagion, the model indicates that the pre-outbreak level of interaction can be hardly restored even in a longer time.

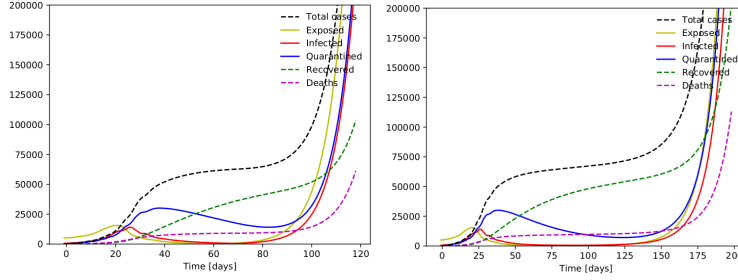


Fig. 13: Two exit strategies are here simulated: immediate reopening of all the activities on May 3rd (a) and gradual reopening over six months time (b).

## 1 Influence of Parameters on Model Forecasts

The most important parameters for the evolution of the infected population and the occurrence of the outbreak are  $\beta$  and  $\langle k \rangle$  parameters. The  $\beta$  parameters is connected to the infection factor, so the probability of infecting a susceptible person in contact with and infected one. The  $\langle k \rangle$  parameters indicate instead the number of connection for each individual, so measures the level of interaction of the Small-World. Since the two parameters always come together their effect is the same on the evolution and visible in Figure 14a, the larger those parameters are the steeper the curve of the infected population is, as expected from Equation 3. A large impact on the outbreak evolution is also given by the fraction of quarantined infected population  $\delta$ , as shown in Figures 14b and 14c. The outbreak is significantly contained in case a larger fraction of infected population is spotted and quarantined.

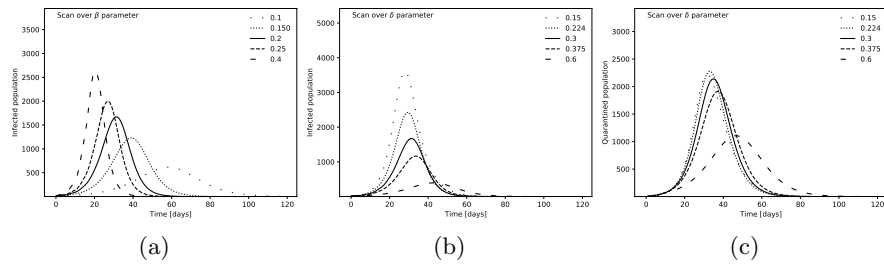


Fig. 14: Effect on the evolution of the infected population given by varying the  $\beta$  parameter (a). Effect on infected (b) and quarantined (c) population given by varying the  $\delta$  parameter.



Other effects are given by  $\gamma$ ,  $\mu$  and  $d_{1/2}$  parameters. Those control more the recovery curve. The  $\gamma$  parameter is related to the probability of a recovered individual to not develop immunity: as shown in Figure 15a, the recovered people has a decrease after the peak of the outbreak. The recovery curve is strongly affected by the  $\mu$  parameter that regulates the recovery probability of the infected population, as shown in Figure 15b. The mortality  $d_{1/2}$  instead increases the number of deaths, given by the difference of the recovered population and the initial 10000 susceptible people, shown in Figure 15c.

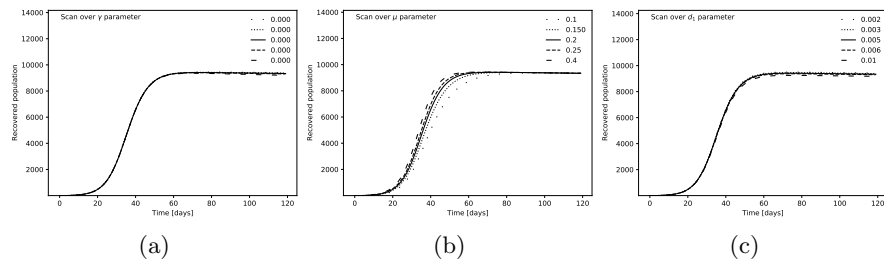


Fig. 15: Effect on the evolution of the recovered population given by varying the  $\gamma$  (a),  $\mu$  (b) and  $d_1$  (c) parameters.

The only temporal parameter is the incubation time  $\tau$ : it also strongly affects the evolution of the outbreak, shown in Figure 16. In case of long incubation time, the outbreak develops a multiple peak structure, while in case of a short incubation time a clean “gaussian” peak is recognised.

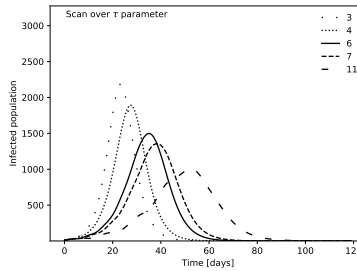


Fig. 16: Effect on the evolution of the infected population given by varying the  $\tau$  parameter.

## References

1. Z. Gao, M. Liu and Y. Xiao, *Modeling and Analysis of Epidemic Diffusion within Small-World Network*, Journal of Applied Mathematics (2012) 841531.
2. M. Liu and L. Zhao, *Analysis for epidemic diffusion and emergency demand in an anti-bioterrorism system*, International Journal of Mathematical Modelling and Numerical Optimisation (2011), vol. 2, no. 1, pp. 51–68.
3. Jiang, F., Deng, L., Zhang, L. et al., *Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19)*, J GEN INTERN MED (2020).
4. Il Sole 24 Ore, *Coronavirus in Italia, i dati e la mappa*, <https://lab24.ilsole24ore.com/coronavirus/>
5. World Health Organisation, *Coronavirus disease 2019 (COVID-19)*, Situation report, 85.
6. Presidenza del Consiglio, *Decreto del presidente del consiglio dei ministri*, Atto 73629, 09/04/2020.
7. Presidenza del Consiglio, *Decreto del presidente del consiglio dei ministri*, Atto 73643, 11/04/2020.
8. Presidenza del Consiglio, *Decreto del presidente del consiglio dei ministri*, Atto 73717, 20/04/2020.
9. Presidenza del Consiglio, *Decreto del presidente del consiglio dei ministri*, Atto 73728, 22/04/2020.
10. Presidenza del Consiglio, *Decreto del presidente del consiglio dei ministri*, , 10/04/2020.
11. I. Antcheva et al., *ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization*, Computer Physics Communications (2009), vol. 180, no. 12.
12. Cranmer, Kyle and Lewis, George and Moneta, Lorenzo and Shibata, Akira and Verkerke, Wouter, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016.
13. Agenzia Nazionale Stampa Associata, *Coronavirus: Calano malati in terapia intensiva, è la prima volta. In Lombardia in giro con le mascherine*, Salute&Benessere, 05 Aprile 2020.
14. Ministero della Salute, *Covid-19, Ministro Speranza: “Vicini a R con zero inferiore a 1, ma la sfida non è vinta”*, News, 9 Aprile 2020.