

# Epidemic analysis of Covid-19 in Italy by dynamical modelling

Lorenzo Mangoni \*

Sant'Anna School of Advanced Studies

Marco Pistilli †

Sant'Anna School of Advanced Studies

April 3, 2020

## Abstract

Epidemic analysis by dynamical modelling is a reliable and insightful way to analyse epidemiological data in order to extract key indicators about the outbreak and to make predictions on its future course.

We develop a generalised SEIR model based on Peng et al. 2020 and estimate it on a national and regional level against the data published daily by the Italian *Dipartimento della Protezione Civile*.

We find the inflection point for Italy to have been on the 21st of March, a plausible end date to be on the 14th of May and expect the total number of infected people to be between 155 thousand and 185 thousand people.

## 1 Introduction

Since the 8th of December 2019, atypical pneumonia cases have emerged in Whuan City, Hubei province, China. The World Health Organization officially identifies a novel coronavirus, named Covid-19, as a cause of the Chinese outbreak, which first started in Whuan and rapidly spread out in the whole China.

In the following months, Covid-19 becomes a global pandemic due to its high infectious capacity with the first Italian cases emerging on the 21st of February, 2020. The virus spread out firstly in Lombardia and in northern regions and then to the rest of the Italian territory, forcing the Government to undertake tight preventive measures.

\*l.mangoni@santannapisa.it

†m.pistilli@santannapisa.it

Throughout the period from the 23rd of February to the 21st of March, the Government has implemented a series of increasingly stringent measures, such as the lockdown of all unnecessary productive activities and strong limitations on leaving one's domicile. These measures were driven by the high number of deaths caused by the virus and the risk of collapse of the National Sanitary System caused by the high propagation speed of the virus.

Medical staff and biologists are in the frontline against the coronavirus. In the meantime theoretical and empirical studies based on both mathematical and statistical modeling also play a fundamental role to support policymaking for curbing the spreading of the virus.

In the early stages of outbreak, several statistical studies have been carried out in order to estimate key epidemic parameters, such as the Basic Reproduction Number (often called  $R_0$ ) and doubling time, both for the outbreak in China (Y. Liu et al. 2020; Whu et al. 2020; T. Liu et al. 2020; Zhao et al. 2020; Li et al. 2020; Sanche et al. 2020) and the Italian outbreak (Raheem 2020; Agosto and Giudici 2020).

Thanks to an increasing amount of available data, as well as to ever stricter measures employed for flattening out the contagion curve, many researchers applied statistical methods (Giuliani et al. 2020, Brogi and Guardabascio 2020, Peracchi 2020) and mathematical models (Fanelli and Piazza 2020) aiming to estimate epidemic parameters and dynamics in order to offer forecasting tools for the evolution of the outbreak.

Among mathematical model based on dynamical equations, the classical SIR (Susceptible Infectious Recovered) model and its extensions are the most widely adopted ones. As shown in section 2, a generalized SEIR model

captures key epidemic parameters of tremendous value for predicting future dynamics and the potential tendency of the Covid-19 outbreak.

We develop a Generalized SEIR model based on Peng et al. 2020, aiming to analyze the dynamic epidemic characteristics of the outbreak in Italy. For this purpose we use the dataset made available by the Italian *Dipartimento della Protezione Civile*, which collects data on a national and regional level.

By adapting the standard SEIR model to the available data, such as a quarantined state and the effects of preventive measures, we estimate the epidemic parameters and make predictions about the future evolution of the outbreak. More specifically, we forecast the inflection point, ending time and total infected cases. In doing so, we also analyze the differences amongst regions. Furthermore, we compare our results with the findings of Peng et al. 2020 and outline the differences between the outbreaks in Italy and China.

The next sections are organized as follows. In Section 2 we describe our Generalized SEIR model and the process of the parameter estimation. In Section 3 we explain and discuss our results in a comparative perspective and present our forecasts for the Covid-19 epidemic in Italy, before concluding in Section 4 with some final remarks.

## 2 Model

### 2.1 A generalised SEIR model

Replicating Peng et al. 2020, we use a generalised version of a SEIR epidemiological model, with the following states at time  $t$ :

- $S(t)$ , the number of susceptible individuals,
- $P(t)$ , the number of insusceptible individuals, who have protected themselves and cannot contract the illness
- $E(t)$ , the number of exposed individuals, not yet contagious,
- $I(t)$ , the number of infected individuals, who may infect others,
- $Q(t)$ , the number of “quarantined” individuals,

who have been tested and resulted positive, and have been put in a condition which makes it impossible for them to infect others,

- $H(t)$ , the number of cured individuals,
- $D(t)$ , the number of deceased individuals,

Defining  $N$  as the population of the considered territorial unit, of course, at any time  $t$ , it holds:

$$S(t) + P(t) + E(t) + I(t) + Q(t) + H(t) + D(t) = N \quad (1)$$

The introduction of the  $Q$  state is motivated by the nature of the available data: the Italian *Dipartimento della Protezione Civile* publishes daily updates on the number of people who have tested positive for COVID-19. Of course, this leaves out the number of people who unknowingly still have the illness and continue to spread it, which here we are calling  $I$ . Our model seeks to reconstruct the cardinality of such a contingent in data-consistent fashion.

The dynamics of the model and the “zero-sum” (due to Eq. 1) transitions between groups are determined by a group of Ordinary Differential Equations:

$$\begin{cases} \dot{S}(t) = -\beta \frac{S(t)I(t)}{N} - \alpha S(t) \\ \dot{P}(t) = \alpha S(t) \\ \dot{E}(t) = \beta \frac{S(t)I(t)}{N} - \gamma E(t) \\ \dot{I}(t) = \gamma E(t) - \delta I(t) \\ \dot{Q}(t) = \delta I(t) - \kappa(t)Q(t) - \lambda(t)Q(t) \\ \dot{H}(t) = \lambda(t)Q(t) \\ \dot{D}(t) = \kappa(t)Q(t) \end{cases} \quad (2)$$

The coefficients  $\{\alpha, \beta, \gamma^{-1}, \delta^{-1}, \lambda(t), \kappa(t)\}$  represent respectively the protection rate, infection rate, average latent time, average quarantine time, cure rate and mortality rate. To take into account improvements in the public health, and the increasingly widespread use of protection devices like face masks, we introduce the coefficient  $\alpha > 0$ , which captures a decrease in the susceptible population. Like Peng et al. 2020, we let cure and mortality rate be time-dependent, so as to capture both the discovery of new drugs and techniques to assist the poor, and conversely, an overburden of the national health service.

From the specification of the model, we have that the Basic Reproduction Number ( $R_0$ ) of the virus is governed by the following dynamic law:

$$R_0(t) = \beta\delta^{-1}(1 - \alpha)^t \quad (3)$$

The inflection point of the epidemiological curve ( $t^*$ , or, equivalently, the peak point in the curve of daily new cases will be the one after which the BRN falls under 1:

$$t^* = \frac{\ln \delta - \ln \beta}{\ln(1 - \alpha)} \quad (4)$$

## 2.2 Parameter estimation

In estimating our parameters we consider the discrete-time equivalent of the generalised SEIR model above described. Our state variables become  $\{S_t, P_t, E_t, Q_t, H_t, D_t\}$ , and their dynamics:

$$\begin{cases} S_{t+1} - S_t = -\beta \frac{S_t I_t}{N} - \alpha S_t \\ P_{t+1} - P_t = \alpha S_t \\ E_{t+1} - E_t = \beta \frac{S_t I_t}{N} - \gamma E_t \\ I_{t+1} - I_t = \gamma E_t - \delta I_t \\ Q_{t+1} - Q_t = \delta I_t - \kappa_t Q_t - \lambda_t Q_t \\ H_{t+1} - H_t = \lambda_t Q_t \\ D_{t+1} - D_t = \kappa_t Q_t \end{cases} \quad (5)$$

Like Peng et al. 2020, we fix  $\gamma^{-1}$  *a priori*, primarily to avoid overfitting. Peng et al. 2020 use  $\gamma^{-1} = 2$  in their model, however, we find that coefficient estimates using  $\gamma^{-1} = 1.6$  tend to have better external validation. Setting  $\gamma^{-1} = 2$  tends to overestimate the numbers of positive individuals at any given time after the end of training data, and consequently skews towards the right estimates for both the inflection date and the end of the epidemic.

Considering that we let  $\kappa_t$  and  $\lambda_t$  be time dependent, the coefficients left to estimate are  $\{\alpha, \beta, \delta\}$ , along with the initial state variables for infected and exposed individuals  $\{E_0, I_0\}$ . To this end, we use the data that the Italian *Dipartimento della Protezione Civile* updates

Italy	
$\beta$	1.0
$\alpha$	0.04
$\delta$	0.28
$\gamma$	0.61
$I_0$	295.51
$E_0$	2016.86
Inflection Date	22/03/2020
End Date	14/05/2020
Total Cases	155914

Table 1: Summary of (i) all constant parameters for the generalized SEIR model ( $E_0$  and  $I_0$  denote the initial values for exposed cases and infectious cases separately) and of (ii) estimated inflection date, end date and final total infected cases

daily<sup>1</sup>, containing the cumulative number of positive, cured and deceased individuals. We run a *simulated annealing* algorithm to find the coefficients and initial values that best fit the available data, minimising the sum of squared deviations from the actual value. We do so, both for the national and regional level.

## 3 Results and forecasts

We train our model on the publicly accessible data as released by *Protezione Civile* on a national and regional level, excluding the last three days, as to keep some out-of-sample data to check for external validation.

The coefficients which best fit our observations for Italy in its entirety are reported in table 1, while coefficients for the regions of Italy are reported in table 2, in the appendix.

On the National level, these coefficients produce the actual vs. fit graph reported in Fig. 1. The model has been trained on the blue dots, while the red dots are shown to offer some external validation. Relative errors are on average 0.7%, which is very promising.

We find the infection rate ( $\beta$ ) to be precisely 1, replicating the findings of Peng et al. 2020, showing the great infective potential for the disease. The Italian protection rate, however, is found to be significantly lower than the Chinese one, to which delayed and less stringent public health measures in Italy might have strongly

<sup>1</sup> Accessible at the link: <https://github.com/pcm-dpc/COVID-19>

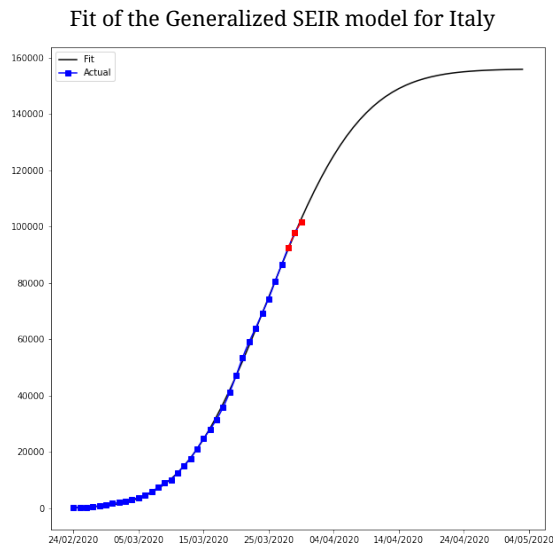


Figure 1: Prediction of the Generalized SEIRD model made on 27th March on the cumulative number of quarantined cases in Italy over 30 days. Blue points are actual values used to estimate the Model, whereas red spots represent the reported data of Italy from 28th March to 30th March, as a validation (they are in 95% confidence interval). These data are publically available and published by the Italian *Dipartimento della Protezione Civile*.

contributed. However, as a “counterbalance”, the average quarantine time for Italy is found to be about half – the one Peng et al. 2020 find for China (Hubei exluded) is 6.6, while the one for Italy is 3.57.

The values for  $E_0$  and  $I_0$  are particularly noteworthy as well. The estimated number of total infected people on the 21st of February is decisively high, especially in Lombardia, suggesting that before the first reported occurrences, the virus had already been circulating for at least some days in the region.

Let us now turn to the infectivity of the disease. Using equation 3 we can compute an estimate for the Basic Reproduction Number of the disease as 3.59, a number perfectly in line with current research (see Flaxman et al. 2020 and Y. Liu et al. 2020). Its dynamics, according to the same law, are shown in figure 2.

As the figure shows, the point at which the BRN fell below one – i.e. the *inflection point* – was approximately the 21st of March, and so we are now on a diminishing trend in terms of daily new cases. It is also important to underline how Peng et al. 2020 estimate the inflection point for Wuhan to also have been thirty days after the first confirmed cases.

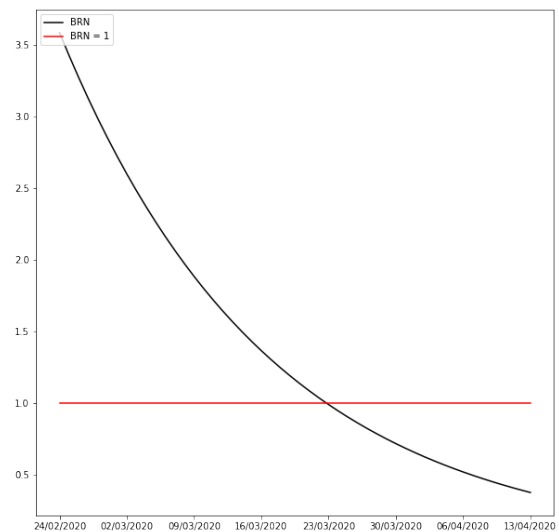


Figure 2: Estimated Italian basic reproduction number over time. On the 21st of March, it falls below one

On the aggregate national data, a likely end date for the epidemic<sup>2</sup> can plausibly be estimated to be around the 14th of May. This is in line with Peracchi (2020). The overall epidemic estimated duration is slight longer than the analogue estimation for China, probably because of less stringent policies.

However, breaking down the numbers on a regional level, the epidemic shows widely varying predicted end dates. In fact, the southern regions have on average a delayed predicted end date. For them, in effect, the outbreak started later on, as the lower  $I_0$  and  $E_0$  numbers show, and therefore also have a lag in end date. The regions which are going to see the end of the epidemic first are Marche and Lombardia.

As for the total number of infected individuals, running the model on the aggregate national data, we find it to be about 156 thousands.

Regionally, apart from the grim figure for total infected in Lombardia, we find moderate numbers for the other regions, showing the effectiveness of the measures taken to limit travel on the national territory.

Since we let  $\lambda_t$  and  $\kappa_t$  be time-dependent, we can plot the evolution of mortality rates and cure rates, and we do so in figure 3. As it is evident, both show an increasing trend through time.

This can be due to a variety of factors, but mostly:

<sup>2</sup>Defined as the day in which the daily new cases are 1 or less

Mortality and heal rates for Italy

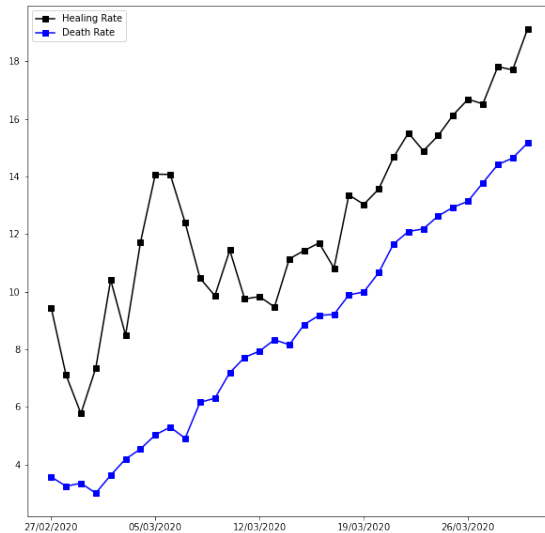


Figure 3: Daily mortality and cure rates

- The increase in death rate is probably due to an overburden of the health system, especially in the most affected areas of the country
- The increase in the cure rate is most likely due to innovative treatments and drugs that are currently being tested in Italy.

## 4 Concluding remarks

This research aims to forecast the potential trajectory of the Covid-19 outbreak, hoping to help policymakers in choosing the appropriate measures to defeat Covid-19 in Italy.

We develop a Generalized SEIR model to capture the dynamics of the outbreak, its features and mechanisms. Unlike many statistical analyses, our model properly incorporates the intrinsic effect of exposed and infectious cases on the entire epidemic dynamics. Moreover, introducing a quarantined state with a cured state, which together replace the original recovered state in classical SIR, the model accounts for the daily reported confirmed infected cases, healed cases and death cases.

Analyzing data made available by the Italian *Dipartimento della Protezione Civile*, we estimated several key parameters for Covid-19 epidemic both in Italy and in every region, like the latent time, the quarantine time and the basic reproduction number. In this way, we

make predictions evolution of the epidemic in Italy, here summarised:

- According to national estimates, the figure for final total infected cases will be around 155 thousands. The sum of the figures for final total infected cases as estimated for individual regions is equal to 185 thousands. We claim that the actual number will most likely be between these two values.
- The estimated number of total infected cases in Lombardia, about 57 thousands, will not be reached by other regions because of anti-epidemic measures.
- We find a starting basic reproduction number equal to 3.58. It decreases throughout time due to the contrasting measures and from the 21st March (inflection date) it is below 1, consequently the Covid-19 epidemic is on a declining path. In the meantime, many regions – especially those where the first confirmed cases were much delayed compared the first-hit regions – are yet to reach the inflection date.
- The epidemic of Covid-19 in Italy will be mostly under control by the 14th of May. This finding is in line with Peracchi (2020). The overall estimated duration of the epidemic is slightly longer than in China, probably because of delayed and less stringent public health policies in Italy.

## References

- Agosto, Arianna and Paolo Giudici (2020). “A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics”. In: *Working paper*. URL: <https://ssrn.com/abstract=3551626>.
- Brogi, Federico and Barbara Guardabascio (2020). “Un modello per la stima dell’andamento del contagio da COVID-19 in Italia”. In: *Working paper*.
- Fanelli, Duccio and Francesco Piazza (2020). “Analysis and forecast of COVID-19 spreading in China, Italy and France”. In: *Chaos, Solitons & Fractals* 134, p. 109761.
- Flaxman, Seth, Swapnil Mishra, and Axel Gandy (2020). “Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries”. In: *MRC Centre for Global Infectious Disease Analysis - Imperial College London working paper*. URL: <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/working-papers/2020-04-29-estimating-the-number-of-infections-and-the-impact-of-non-pharmaceutical-interventions-on-covid-19-in-11-european-countries/>

ac.uk/mrc-global-infectious-disease-analysis/covid-19/.

phase of the outbreak". In: *International Journal of Infectious Disease*. DOI: 10.1101/2020.01.23.916395.

- Giuliani, Diego, Maria Michela Dickson, Giuseppe Espa, and Flavio Santi (2020). "Modelling and predicting the spread of Coronavirus (COVID-19) infection in NUTS-3 Italian regions". In: *Working paper*.
- Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy Leung, Eric Lau, Jessica Y Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Li, and Zijian Feng (2020). "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". In: *New England Journal of Medicine* 382. DOI: 10.1056/NEJMoa2001316.
- Liu, Tao, Jianxiong Hu, Jianpeng Xiao, Guanhao He, Min Kang, Zuhua Rong, Lifeng Lin, Haojie Zhong, Qiong Huang, Aiping Deng, Weilin Zeng, Xiaohua Tan, Siqing Zeng, Zhihua Zhu, Jiansen Li, Dexin Gong, Donghua Wan, Shaowei Chen, Lingchuan Guo, Yan Li, Limei Sun, Wenjia Liang, Tie Song, Jianfeng He, and Wenjun Ma (2020). "Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China". In: *bioRxiv*. DOI: 10.1101/2020.01.25.919787.
- Liu, Ying, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv (2020). "The reproductive number of COVID-19 is higher compared to SARS coronavirus". In: *Journal of Travel Medicine* 27.2. ISSN: 1195-1982. DOI: 10.1093/jtm/taaa021. URL: <https://doi.org/10.1093/jtm/taaa021>.
- Peng, Liangrong, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong (2020). "Epidemic analysis of COVID-19 in China by dynamical modeling". In: *Working paper*.
- Peracchi, Francesco (2020). "The Covid-19 pandemic in Italy". In: *Working paper*.
- Raheem, Ali (2020). "Estimating cases of COVID-19 from Daily Death Data in Italy". In: *medRxiv*. DOI: 10.1101/2020.03.17.20037697.
- Whu, Joseph T, Kathy Leung, and Gabriel M Leung (2020). "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". In: *The Lancet* 395.10225. ISSN: 689-697. DOI: [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- Zhao, Shi, Jinjun Ran, Salihu Musa, Guangpu Yang, Yijun Lou, Daozhou Gao, Lin Yang, and Daihai He (2020). "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early

# A Appendix

## Fit of the Generalized SEIR model for every regions<sup>3</sup>

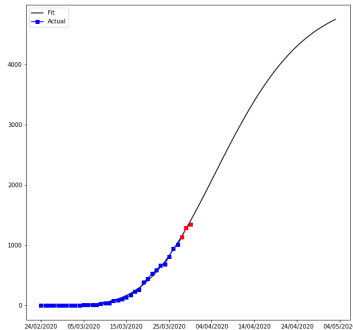


Figure 4: Abruzzo

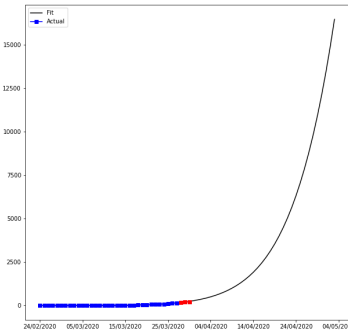


Figure 5: Basilicata

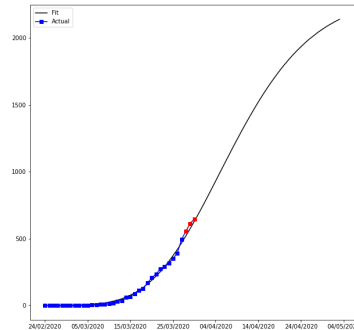


Figure 6: Calabria

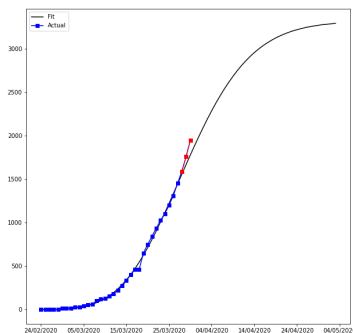


Figure 7: Campania

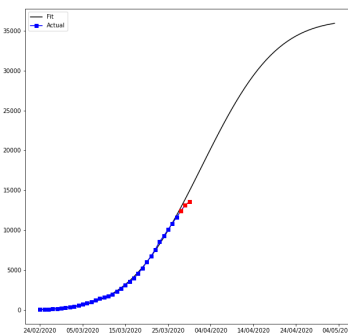


Figure 8: Emilia Romagna

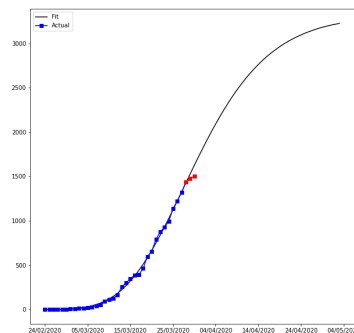


Figure 9: Friuli Venezia Giulia

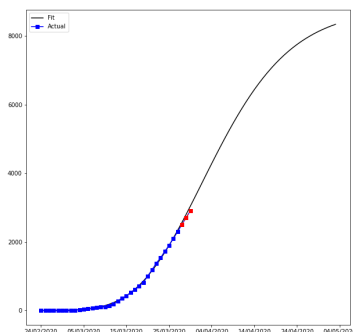


Figure 10: Lazio

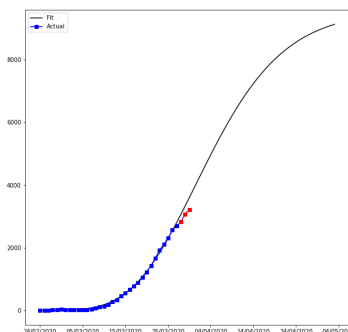


Figure 11: Liguria

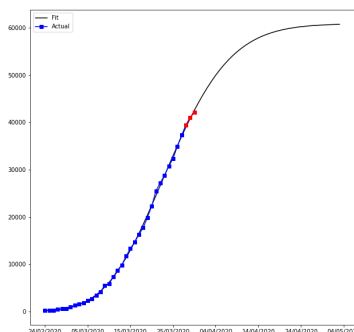


Figure 12: Lombardia

<sup>3</sup>Prediction of the Generalized SEIRD model made on 27th March on the cumulative number of quarantined cases in every Italian regions over 30 days. Blue points are actual values used to estimate the Model, whereas red spots represent the reported data from 28th March to 30th March, as a validation (they are in 95% confidence interval). These data are made available in public from the Italian *Dipartimento della Protezione Civile*. Parameters are taken in accordance with Table 2.

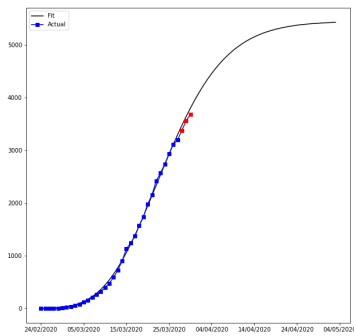


Figure 13: Marche

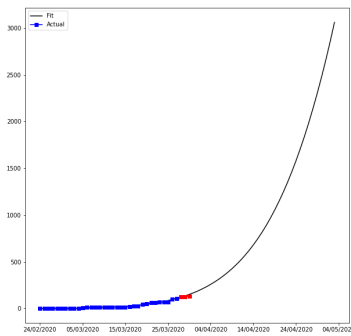


Figure 14: Molise

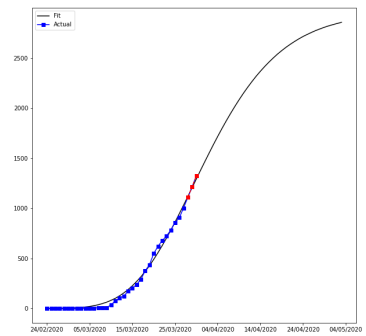


Figure 15: P.A. Bolzano

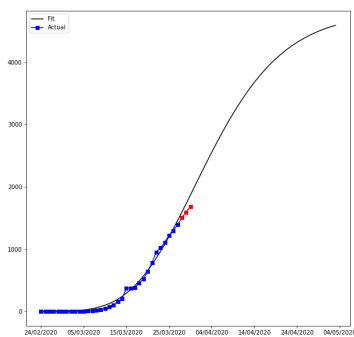


Figure 16: P.A. Trento

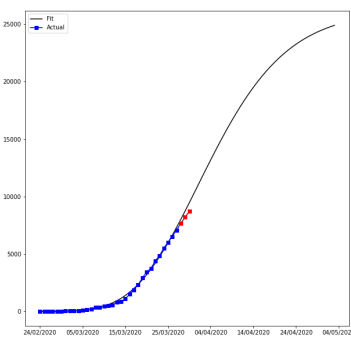


Figure 17: Piemonte

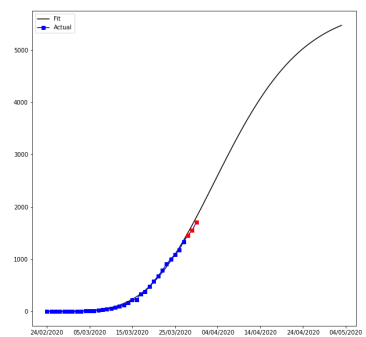


Figure 18: Puglia

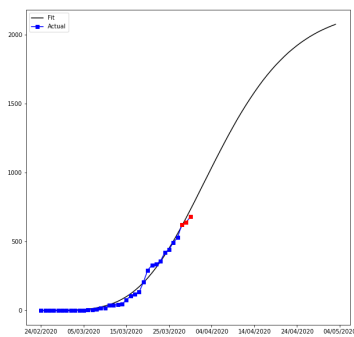


Figure 19: Sardegna

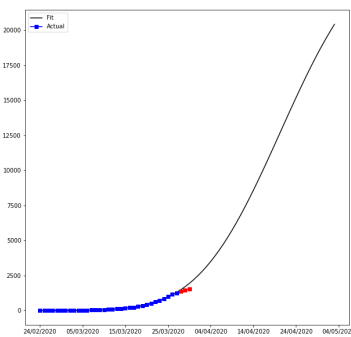


Figure 20: Sicilia

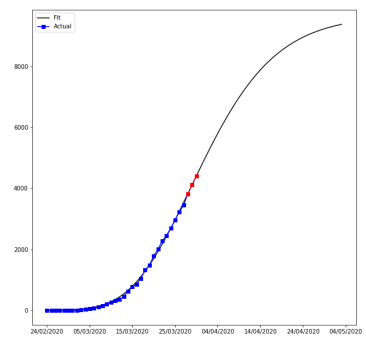


Figure 21: Toscana



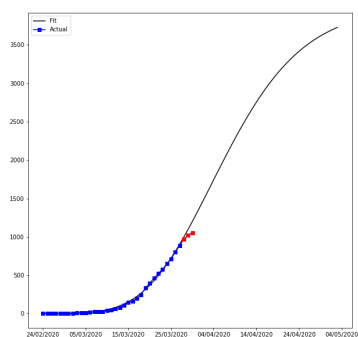


Figure 22: Umbria

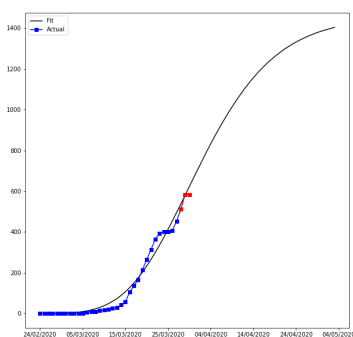


Figure 23: Valle d'Aosta

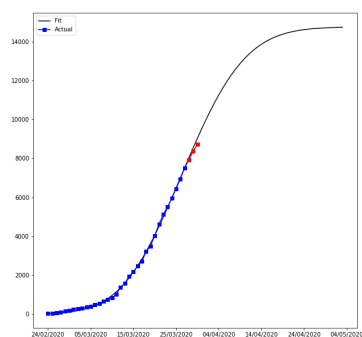


Figure 24: Veneto

Region	Sardegna	Sicilia	Campania	Marche	Veneto	P.A. Trento	Liguria
beta	1	1	0.707862	1	1	1	1
alpha	0.0644789	0.0537757	0.0528773	0.0730891	0.0401963	0.0662973	0.0643163
delta	0.1	0.116648	0.1	0.156815	0.317253	0.116033	0.123106
gamma	0.61	0.61	0.61	0.61	0.61	0.61	0.61
I0	1	1.8608	6.01036	10.8718	26.9818	2.68171	4.84472
E0	1.58516	1.30258	12.0207	21.7437	18.8873	5.36342	9.68944
Infection Date	34.5466	38.8705	36.0243	24.4103	27.9833	31.3989	31.5099
End Date	16/05/2020	29/05/2020	31/05/2020	30/04/2020	04/05/2020	15/05/2020	19/05/2020
Total Cases	1912	6134	5579	5132	13465	3721	7107
RE	0.823781	0.818134	0.80816	0.765858	0.694652	0.81127	0.794362
	Lazio	Umbria	Toscana	Lombardia	Abruzzo	Puglia	Calabria
beta	1	1	1	0.905868	1	1	0.919481
alpha	0.0634545	0.0625818	0.06997	0.0429557	0.0593729	0.063092	0.0532966
delta	0.106681	0.108118	0.102768	0.297842	0.1	0.1	0.1
gamma	0.61	0.61	0.61	0.61	0.61	0.61	0.61
I0	3.65884	1.32845	7.55246	141.528	1.30044	2.01753	1
E0	7.31769	2.6569	15.1049	283.055	2.60087	4.03507	0.7
Infection Date	34.1368	34.4218	31.3666	25.3346	37.6187	35.3319	40.5087
End Date	26/05/2020	17/05/2020	26/05/2020	06/05/2020	29/05/2020	26/05/2020	28/05/2020
Total Cases	7741	2962	9658	57684	4764	4998	2833
RE	0.802539	0.819977	0.78194	0.191243	0.819313	0.814901	0.825349
	Emilia Romagna	Valle d'Aosta	P.A. Bolzano	Molise	Friuli Venezia Giulia	Basilicata	Piemonte
beta	1	1	1	0.501562	1	0.454803	1
alpha	0.0318313	0.0681708	0.0690921	0.0202546	0.0656247	0.0151776	0.0639718
delta	0.377541	0.110132	0.1	0.168036	0.141999	0.1	0.11534
gamma	0.61	0.61	0.61	0.61	0.61	0.61	0.61
I0	42.4062	1	2.10294	1	2.82484	1	12.0982
E0	29.6844	1.95484	4.20589	0.7	5.64969	0.7	24.1964
Infection Date	30.1115	31.245	32.1613	53.4414	28.7569	99.0389	32.6711
End Date	09/05/2020	07/05/2020	17/05/2020	06/03/2020	05/05/2020	06/03/2020	31/05/2020
Total Cases	23856	1279	3010	5	2742	4	20916
RE	0.624993	0.823645	0.816507	0.828516	0.81019	0.828669	0.741237

Table 2: Summary of (i) all constant parameters for the generalized SEIR model ( $E_0$  and  $I_0$  denote the initial values for exposed cases and infectious cases separately) and of (ii) estimated inflection date, end date and final total infected cases