

Predicting and analyzing the COVID-19 pandemic in Italy using SEIR-type and deep learning models: a comparative study

Simone Bodojra
s279635

Silvia Giammarinaro
s269991

April 9, 2021



Modelli Matematici per la Biomedicina
Politecnico di Torino
Academic Year 2020-2021

Contents

1	Introduction	1
1.1	General introduction to the problem	1
1.2	Phenomenological observation and demand	1
1.3	Related work	1
2	Methods	2
2.1	Model deduction	2
2.2	Qualitative analysis of mathematical models main features	4
2.2.1	Disease-free equilibrium point	5
2.2.2	Dimensionless models	7
3	Results	8
3.1	Model simulations	8
3.2	Sensitivity Analysis	10
3.3	Numerical simulations	11
3.3.1	Italy predictions	14
3.3.2	Molise predictions	14
3.3.3	Sardegna predictions	15
3.3.4	Summary of numerical simulations	16
3.4	LSTM: a deep learning model for predictions	16
3.5	A comparison between SEIIRHD and LSTM models	20
3.6	When the pandemic will end?	21
4	Conclusions and further work	22
5	Appendix	28
5.1	SIR, SEIR and SEIIR models fitting	28
5.1.1	Italy	28
5.1.2	Molise	30
5.1.3	Sardegna	32

1 Introduction

1.1 General introduction to the problem

At the end of 2019, the city of Wuhan, China, reported cases of pneumonia of unknown cause [1], whose symptoms include fever, cough, and shortness of breath [2]. The virus causing this disease will be later named as "SARS-CoV-2". While the disease was rapidly growing in many countries, the WHO (World Health Organization) shared the news with the entire world on January 5th. China reported the first death on January 11th, 2020. The first case outside China was declared in Thailand on January 13th. The first confirmed cases in Europe are reported in France on January 24th. On January 30th, 2020, the WHO declared the novel coronavirus outbreak a public health emergency of international concern (PHEIC) [1]. At the time of writing, there are still many infections around the world, so predicting the trend of the epidemic is fundamental to alleviate the pressure on the health system and activate control strategies (e.g. quarantines, lock-downs, and suspension of travel) aiming at containing the disease and delaying the spread.

1.2 Phenomenological observation and demand

This COVID-19 pandemic has spread worldwide and the scientific community came across to study the disease from different points of view. Those aspects include pathology, sociology, infection mechanism, and prediction.

The goal of this work is to analyze how the epidemic is spreading in Italy by using two different methods: mathematical and deep learning models. Mathematical problems use differential equations in which the population is divided into compartments, each of these representing a state each individual can assume. Instead, deep learning models apply advanced computing techniques on big datasets to make predictions or classifications.

A new epidemiological model SEIIRHD will be proposed by taking into account the asymptomatic and symptomatic cases, hospitalized and deaths. It is a system of coupled ordinary differential equations (ODEs) that are driven by a set of parameters that describes how the virus interacts with the different compartments. It will be compared to a deep learning model using LSTM cells.

1.3 Related work

Mathematical models are vital to make predictions. The first epidemiological model was formulated back in 1760 by Daniel Bernoulli to support the smallpox vaccination [3]. Several generalization of the SIR model [4], introduced by Kermack and McKendrick in 1927, have been formulated over the years by enlarging the number of compartments. For example, the SEIR model [5] introduces a compartment for individuals that have contracted the virus but are not yet infectious (Exposed).

The particular characteristics of the COVID-19 ask for models better able to accurately portray its nature, so the recent period has seen a considerable flowering of epidemiological mathematical models. For example, the SEIIURD model [6] divides the infected into asymptomatic and symptomatic and distinguishes between recovered (R) and dead people (D); the SUIHTER model [7] estimates the undetected (U) compartment: infected people are both asymptomatic and symptomatic. The number of detected new daily infected depends on the different swab policies applied in each country.

In addition to mathematical models, researchers recently started to take advantage of the advances made in Deep Learning (DL) in order to create models based on available large datasets, many of which are based on LSTM [8]. This method is preferred over basic machine learning models as the architecture is designed to deal with time series [9].

2 Methods

2.1 Model deduction

In the **SIR model** [4], the population is divided in three compartments: *Susceptibles* (S), *Infected* (I) and *Removed* (R). A susceptible is an individual that might get infected with infection rate β and then he recovers with recovery rate γ . It follows a system of differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}\tag{1}$$

where the total population $N = S + I + R$, which will always be assumed constant. The removed individuals are supposed to be immune to life, even if for the SARS-CoV-2 virus there are still studies going on about the immunity period. One clinical study analyzed different corona viruses and it identified a natural reinfection by the virus after a period of 12 months [10].



Figure 1: **Block diagram SIR.** Arrows represent infection flow; β is the infection rate and γ is the recovery rate

In the **SEIR model** [5] the *Exposed* (E) compartment is added: they have contracted the virus but are not yet infectious and, after the incubation period, they become infected. Therefore, the model can be rewritten as:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dE}{dt} &= \beta SI - \alpha E \\ \frac{dI}{dt} &= \alpha E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}\tag{2}$$

where the total population $N = S + E + I + R$. In this model α is the inverse of the incubation period, which gives information for monitoring and modeling the disease to effectively use health limited resources.

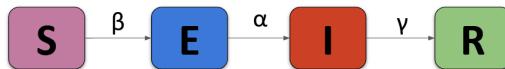


Figure 2: **Block diagram SEIR.** The flow is similar to the SIR one, the only new parameter is α which describes the inverse of the incubation period.

An extension of the classical SEIR model adapted to COVID-19 is the **SEIIR model**. It addresses the possibility of having an infected individual showing no symptoms by splitting the infected into two compartments: *Symptomatic Infected* (I_s) and *Asymptomatic Infected* (I_a). The system of equations can be defined in the following way:

$$\begin{aligned}\frac{dS}{dt} &= -(\beta_s I_s + \beta_a I_a)S \\ \frac{dE}{dt} &= (\beta_s I_s + \beta_a I_a)S - \alpha E \\ \frac{dI_a}{dt} &= (1-f)\alpha E - \gamma I_a \\ \frac{dI_s}{dt} &= f\alpha E - \gamma I_s \\ \frac{dR}{dt} &= \gamma(I_s + I_a)\end{aligned}\tag{3}$$

A susceptible can be infected by symptomatic infected and asymptomatic infected individuals with infection rates β_s and β_a . An infected can be symptomatic or asymptomatic with probability f and $1-f$.

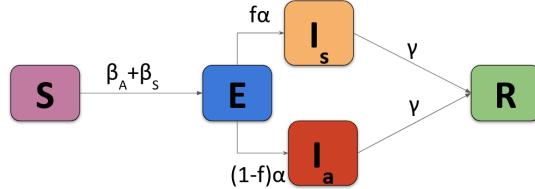


Figure 3: **Block diagram SEIIR.** The infected compartment is splitted into asymptomatic I_a and symptomatic I_s , β_a and β_s are the infection rates, f is the probability of being symptomatic.

The model proposed in this work is the **SEIIRHD model**. Two other compartments are added: *Hospitalized* (H) and *Deaths* (D). There is also a *Recovered* (R) compartment, instead of a Removed one, where all gathered all individuals recovered from the virus. The dynamics is described by the following system of differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -(\beta_s I_s + \beta_a I_a)S \\ \frac{dE}{dt} &= (\beta_s I_s + \beta_a I_a)S - \alpha E \\ \frac{dI_a}{dt} &= (1-f)\alpha E - \gamma I_a \\ \frac{dI_s}{dt} &= f\alpha E - (\gamma + \mu + \nu_s)I_s \\ \frac{dR}{dt} &= \gamma(I_s + I_a + H)\end{aligned}\tag{4}$$

$$\begin{aligned}\frac{dH}{dt} &= \nu_s I_s - (\gamma + \mu)H \\ \frac{dD}{dt} &= \mu(I_s + H)\end{aligned}$$

If the symptoms are serious, a symptomatic individual is hospitalized with a rate ν_s . An infected individual could recover with a recovery rate γ , or die with mortality rate μ .

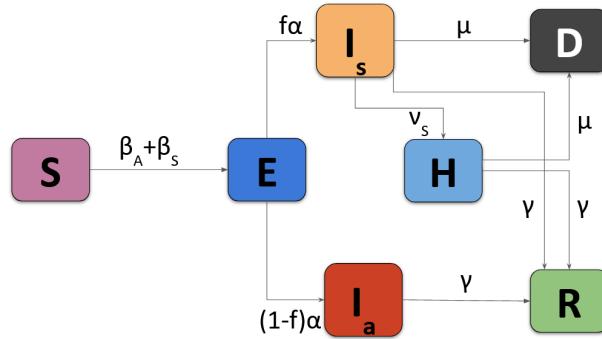


Figure 4: **Block diagram SEIIRHD.** The infected are divided in symptomatic and asymptomatic. The symptomatic individuals can be hospitalized, recover or die; instead the asymptomatic ones can only recover. The parameters γ and α are the same as SEIR model; β_a and β_s are the infection rates; f is the probability of being symptomatic; ν_s is the intervention rate; μ is the mortality rate.

2.2 Qualitative analysis of mathematical models main features

It is assumed that the total population N is constant and the state parameters are non-negative $\forall t \geq 0$.

The models will be studied in the following feasible regions:

$$\begin{aligned}\Omega_{\text{SIR}} &= \{(S(t), I(t), R(t)) \in R_+^3 : 0 \leq N(t) \leq N_0\} \\ \Omega_{\text{SEIR}} &= \{(S(t), E(t), I(t), R(t)) \in R_+^4 : 0 \leq N(t) \leq N_0\} \\ \Omega_{\text{SEIIR}} &= \{(S(t), E(t), I_a(t), I_s(t), R(t)) \in R_+^5 : 0 \leq N(t) \leq N_0\} \\ \Omega_{\text{SEIIRHD}} &= \{(S(t), E(t), I_a(t), I_s(t), H(t), R(t), D(t)) \in R_+^7 : 0 \leq N(t) \leq N_0\}\end{aligned}$$

As soon as the initial conditions are positive, the proposed models are well posed and the solutions of the systems live in the respective Ω set.

Inclusion of a demographic dynamics may permit the disease to persist in the population for a longer period of time [11], one or two decades. As the COVID-19 pandemic started one year ago, it is not enough time to consider this scenario. Furthermore, the endemic equilibrium (END) can be defined in this case. Therefore, the only equilibrium point defined is the disease-free equilibrium (DFE). It describes a scenario where no disease is present in the population.

2.2.1 Disease-free equilibrium point

The local stability of the DFE point is proved in this section after introducing the basic reproduction number.

In mathematical epidemiology, a fundamental quantity is the **basic reproduction number** \mathcal{R}_0 : it is the expected number of secondary cases produced by a single infection in a completely susceptible population. It is a dimensionless number, not a rate and it can be obtained for *any* model using the next generation method.

Let us assume that there are m compartments of which n are infected. Let X be a n -dimensional vector of infected classes (exposed, infected, asymptomatic...):

$$\frac{dX}{dt} = \mathcal{F}(X) - \mathcal{V}(X) \quad (5)$$

where

- $\mathcal{F}(X)$ is the **vector of new infection rates**, its i -th element describe the rate of appearance of new infections in compartment i ;
- $\mathcal{V}(X)$ is the **vector of all other rates**, for example recovery rates or death rates. It is equal to $\mathcal{V}^-(X) - \mathcal{V}^+(X)$, where $\mathcal{V}^+(X)$ is the rate of transfer of individuals into compartment i by all other means and $\mathcal{V}^-(X)$ is the rate of transfer out of the i -th compartment.

Note that $\mathcal{F}(X)$ should include only infections that are newly arising, but does not include terms which describe the transfer of infectious individuals from one infected compartment to another.

Assuming that \mathcal{F} and \mathcal{V} meet the conditions outlined in [12], the next generation matrix (operator) is defined FV^{-1} from matrices of partial derivatives of \mathcal{F} and \mathcal{V} . Specifically,

$$F = \left[\frac{\partial \mathcal{F}}{\partial X} \right] \quad V = \left[\frac{\partial \mathcal{V}}{\partial X} \right] \quad (6)$$

evaluated in the disease-free equilibrium point.

The next generation operator FV^{-1} gives rate at which individuals in compartment j generate new infections in compartment i times average length of time individual spends in single visit to compartment j . \mathcal{R}_0 is given by the dominant eigenvalue of FV^{-1} written as $\mathcal{R}_0 = \rho(FV^{-1})$.

Example 2.1 (SIR). This model has 3 compartments and $X = I$. The DFE is given by

$$(S_{DFE}^*, I_{DFE}^*, R_{DFE}^*) = (1, 0, 0) \quad (7)$$

$$\mathcal{F}_1 = \beta SI \quad \mathcal{V}_1 = \gamma I$$

$$F = \frac{\partial \mathcal{F}_1}{\partial I} = \beta S_{DFE}^* = \beta \quad V = \frac{\partial \mathcal{V}_1}{\partial I} = \gamma$$

Hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \quad (8)$$

Example 2.2 (SEIR). The model has 4 compartments and $X = [E, I]$. The DFE is given by

$$(S_{DFE}^*, E_{DFE}^*, I_{DFE}^*, R_{DFE}^*) = (1, 0, 0, 0) \quad (9)$$

$$\begin{aligned} \mathcal{F} &= \begin{bmatrix} \beta S I \\ 0 \end{bmatrix} & \mathcal{V} &= \begin{bmatrix} \alpha E \\ \gamma I - \alpha E \end{bmatrix} \\ F &= \begin{bmatrix} \frac{\partial \mathcal{F}_1}{\partial E} & \frac{\partial \mathcal{F}_1}{\partial I} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \beta S_{DFE}^* \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} & V &= \begin{bmatrix} \frac{\partial \mathcal{V}_1}{\partial E} & \frac{\partial \mathcal{V}_1}{\partial I} \\ \frac{\partial \mathcal{V}_2}{\partial E} & \frac{\partial \mathcal{V}_2}{\partial I} \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ -\alpha & \gamma \end{bmatrix} \end{aligned}$$

Therefore,

$$FV^{-1} = \frac{1}{\alpha\gamma} \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ \alpha & \alpha \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\gamma} & \frac{\beta}{\gamma} \\ 0 & 0 \end{bmatrix} \quad (10)$$

Recalling that \mathcal{R}_0 is the maximum eigenvalue of the matrix FV^{-1} :

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \quad (11)$$

Example 2.3 (SEIIRHD). The model has 7 compartments and $X = [E, I_a, I_s]$. The DFE is:

$$(S_{DFE}^*, E_{DFE}^*, I_{a,DFE}^*, I_{s,DFE}^*, R_{DFE}^*, H_{DFE}^*, D_{DFE}^*) = (1, 0, 0, 0, 0, 0, 0) \quad (12)$$

$$\begin{aligned} \mathcal{F} &= \begin{bmatrix} \beta_a I_a S + \beta_s I_s S \\ 0 \\ 0 \end{bmatrix} & \mathcal{V} &= \begin{bmatrix} \alpha E \\ \gamma I_a - (1-f)\alpha E \\ (\gamma + \mu + \nu_s) I_s - f\alpha E \end{bmatrix} \\ F &= \begin{bmatrix} \frac{\partial \mathcal{F}_1}{\partial E} & \frac{\partial \mathcal{F}_1}{\partial I_a} & \frac{\partial \mathcal{F}_1}{\partial I_s} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \beta_a & \beta_s \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & V &= \begin{bmatrix} \frac{\partial \mathcal{V}_1}{\partial E} & \frac{\partial \mathcal{V}_1}{\partial I_a} & \frac{\partial \mathcal{V}_1}{\partial I_s} \\ \frac{\partial \mathcal{V}_2}{\partial E} & \frac{\partial \mathcal{V}_2}{\partial I_a} & \frac{\partial \mathcal{V}_2}{\partial I_s} \\ \frac{\partial \mathcal{V}_3}{\partial E} & \frac{\partial \mathcal{V}_3}{\partial I_a} & \frac{\partial \mathcal{V}_3}{\partial I_s} \end{bmatrix} = \begin{bmatrix} \alpha & 0 & 0 \\ (f-1)\alpha & 0 & \gamma \\ -\alpha f & \gamma + \mu + \nu_s & 0 \end{bmatrix} \end{aligned}$$

Therefore,

$$FV^{-1} = \begin{bmatrix} \frac{f\beta_s}{\gamma + \mu + \nu_s} + \frac{(1-f)\beta_a}{\gamma} & \frac{\beta_s}{\gamma + \mu + \nu_s} & \frac{\beta_a}{\gamma} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (13)$$

Therefore:

$$\mathcal{R}_0 = \frac{f\beta_s}{\gamma + \mu + \nu_s} + \frac{(1-f)\beta_a}{\gamma} \quad (14)$$

The basic reproduction number is also a threshold value, as showed in Theorem 2.5. To prove this theorem, the following lemma is introduced:

Lemma 2.4. Let H be a non-singular N -matrix and suppose B and BH^{-1} have the Z sign pattern. Then B is a non-singular N -matrix if and only if BH^{-1} is a non-singular N -matrix.

Theorem 2.5. The disease-free equilibrium point is asymptotically stable if $\mathcal{R}_0 < 1$.

Proof. According to [13], the proof is based on the matrices F and V. Let $J_1 = F - V$.

$$J_1 = F - V = \begin{bmatrix} -\alpha & \beta_a & \beta_s \\ -(f-1)\alpha & 0 & -\gamma \\ \alpha f & -(\gamma + \mu + \nu_s) & 0 \end{bmatrix}$$

Since V is a non-singular N-matrix (N=3 in the SEIIRHD model) and F is non-negative, the matrix $-J_1$ has the Z sign pattern. A matrix has the Z sign pattern if $x_{ij} \leq 0 \forall i \neq j$. Thus:

$$s(J_1) < 0 \Leftrightarrow -J_1 \text{ non-singular N-matrix} \quad (15)$$

where $s(J_1)$ denotes the maximum real part among all eigenvalues of matrix J_1 . Furthermore, since FV^{-1} is non-negative, $-J_1 V^{-1} = I - FV^{-1}$ also has the Z sign pattern.

According to the lemma 2.4, setting $H = V$ and $B = -J_1 = V - F$, we obtain that $BH^{-1} = I - FV^{-1}$.

$$I - FV^{-1} = \begin{bmatrix} 1 - \frac{f\beta_s}{\gamma+\mu+\nu_s} + \frac{(1-f)\beta_a}{\gamma} & \frac{\beta_s}{\gamma+\mu+\nu_s} & \frac{\beta_a}{\gamma} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (16)$$

So, $I - FV^{-1}$ is a non-singular N-matrix. Going back to the definition of spectral radius, all eigenvalues of FV^{-1} have magnitude less than or equal to $\rho(FV^{-1}) = \mathcal{R}_0$. Thus

$$I - FV^{-1} \text{ is a non-singular N-matrix} \Leftrightarrow \rho(FV^{-1}) < 1 \quad (17)$$

Finally, $s(J_1) < 0$ if and only if $\mathcal{R}_0 < 1$. \square

2.2.2 Dimensionless models

To generalize the models, the variables are normalized by the conserved number N , as it could be useful to see the behaviors of the systems with different N . As the infection cannot evenly spread throughout Italy, for example, it does not make sense to set the question of how the entire population of Italy is infected: there are many high-risk areas such as densely populated areas. Since the spread of the infection in each place is similar to each other, dimensionless models will be considered, where fractions of each compartment are used. For example, the dimensionless SEIIRHD model is:

$$\begin{aligned} \frac{ds}{dt} &= -(\beta_s N \frac{I_s}{N} + \beta_a N \frac{I_a}{N}) \frac{S}{N} = -(\beta_s N i_s + \beta_a N i_a) s \\ \frac{de}{dt} &= (\beta_s N \frac{I_s}{N} + \beta_a N \frac{I_a}{N}) S - \alpha \frac{E}{N} = -(\beta N) s i - \alpha e \\ \frac{di_a}{dt} &= (1-f)\alpha \frac{E}{N} - \gamma \frac{I_s}{N} = (1-f)\alpha e i_s - \gamma i_a \\ \frac{di_s}{dt} &= f\alpha \frac{E}{N} - (\gamma + \mu + \nu_s) \frac{I_s}{N} = f\alpha e - (\gamma + \mu + \nu_s) i_s \\ \frac{dr}{dt} &= \gamma \left(\frac{I_s}{N} + \frac{I_a}{N} + \frac{H}{N} \right) = \gamma(i_s + i_a + h) \\ \frac{dh}{dt} &= \nu_s \frac{I_s}{N} - (\gamma + \mu) \frac{H}{N} = \nu_s i_s - (\gamma + \mu) h \\ \frac{dd}{dt} &= \mu \left(\frac{I_s}{N} + \frac{H}{N} \right) = \mu(i_s + h) \end{aligned}$$

3 Results

In this section, the numerical results of SEIIRHD and LSTM models are presented, while results for the other models are reported in the Appendix [5]. The code used for this project is available at <https://github.com/sigeek/fitting-covid-19>.

The flow of this work may be described as follows:

- *Model simulations* (3.1): to describe the model and to show how different values of the basic reproduction number \mathcal{R}_0 influence the evolution of the model;
- *Sensitivity Analysis* (3.2): to evaluate how the variation of a parameter affects the \mathcal{R}_0 value;
- *Estimating parameters from real data with SEIIRHD* (3.3): to check the forecasting ability of the model; the fittings used two different periods for Italy (3.3.1), Molise (3.3.2) and Sardegna (3.3.3);
- *LSTM: a deep learning model for predictions* (3.4): to analyze the forecasting abilities of a deep learning model;
- *Comparison between SEIIRHD and LSTM* (3.5): to further discuss the validity of the model with respect to a modern approach based on deep learning techniques;
- *Simulations of the end of the pandemic* (3.6): to analyze forecasting abilities on longer periods.

3.1 Model simulations

At first, it is examined how each compartment changes according to the value of the basic reproduction number \mathcal{R}_0 . As seen in Section 2.2.1, it is a threshold value used to understand the epidemic evolution. **NPI** (nonpharmaceutical interventions) have played a critical role in reducing transmission rates [14]. NPIs adopted during the pandemic include the use of face masks, hand sanitizers, teleworking, and school closure.

The simulations were conducted using 1000 values of parameters extracted from a uniform distribution $\mathcal{U}(a, b)$, where a and b indicate lower and upper values, as specified in Table 1. The initial conditions are the data reported on February 24th 2020 (i.e., the day Italy started collecting data). For each variable, first quartile, third quartile and median were obtained and three different \mathcal{R}_0 were computed. Results for 50 days are shown in Figure 5.

Parameter	Description	Distribution interval
f	Probability of being symptomatic	(0.3, 0.9)
β_a	Transmission rate from S to E from contact with I_a	(0.0, 0.6)
β_s	Transmission rate from S to E from contact with I_s	(0.0, 0.6)
γ	Recovery rate	(0.0, 0.4)
α	Inverse of the incubation period	(0.15, 0.35)
ν_s	Hospitalization rate from state I_s	(0.0, 0.4)
μ	Death Rate	(0.0, 0.2)

Table 1: **Values of the parameters of the SEIIRHD model.** For each parameter, it is provided a description along with the interval used to define the uniform distributions for the simulations.

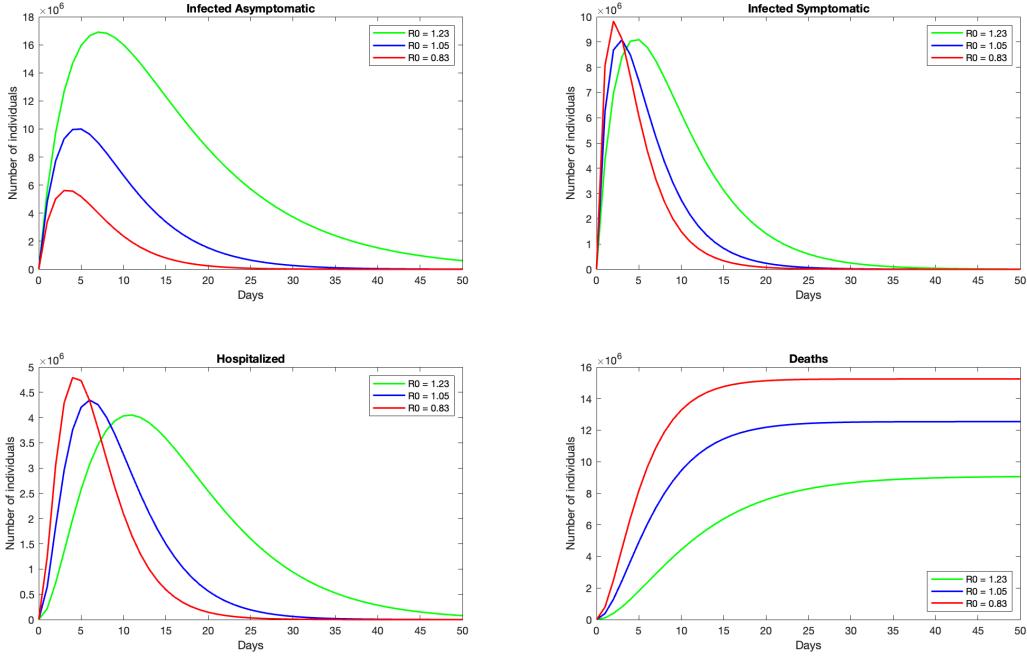


Figure 5: **Compartment evolution of the SEIIRHD model.** First quartile (green line), median (blue line) and third quartile (red line) were obtained from 1000 simulations conducted using values of parameters extracted from uniform distributions reported in Table 1.

The results can be described by the three following scenarios:

- **Scenario 1 ($R_0 = 1.23$)**. The number of asymptomatic individuals is higher than the number of symptomatic ones due to the higher value of f . The decrease of symptomatic infected individuals is slow and this happens because the asymptomatic individuals will continue to infect susceptibles, who may later become symptomatic. Even though, the death rate and the hospitalization rate are quite small, this means that still the majority of the infected is asymptomatic. After 50 days, there are still many asymptomatic infected individuals and the disease is still present in the population. This happens because $R_0 > 1$ and, overtime, the endemic equilibrium will be reached. Even if it is the scenario with a higher value of R_0 , the number of severe cases is lower than the mild cases;
- **Scenario 2 ($R_0 = 1.05$)**. Even if the probability of being symptomatic is higher than before, the number of asymptomatic is comparable to the symptomatic one. The symptomatics will exit faster than the previous scenario from the compartment. Both the infected compartments will recover with the same rate γ , so the difference is that symptomatic individuals are linked to the hospitalized and death compartments, which have higher entry rates;
- **Scenario 3 ($R_0 = 0.83$)**. This scenario describes a faster and more lethal disease. The symptomatic individuals are more than the asymptomatic ones and hospitals can only handle a fixed number of patients. This results in a high number of dead individuals. However, thanks to the high hospitalization rate and high recovery rate, the epidemic is quickly controlled after 25 days. Thanks to the small number of asymptomatic individuals, the population will be free from the disease before the other two scenarios.

From this analysis, the importance of the asymptomatic compartment is clear: if its individuals are more than the symptomatic ones, the disease is less lethal but more difficult to eradicate; on the contrary, the disease will cause a higher number of deaths, but the population will be free from it faster.

Parameters			
	$\mathcal{R}_0 = 1.23$	$\mathcal{R}_0 = 1.05$	$\mathcal{R}_0 = 0.83$
f	0.4528	0.6013	0.7479
β_a	0.1587	0.3176	0.4434
β_s	0.1504	0.3210	0.4442
γ	0.0919	0.1911	0.2912
α	0.1982	0.2515	0.3003
ν_s	0.0991	0.2027	0.3096
μ	0.0462	0.1018	0.1498
Maximum value for each compartment			
I_a	16.90M	9.99M	5.61M
I_s	9.09M	9.08M	9.82M
H	4.05M	4.34M	4.79M
D	9.05M	12.54M	15.24M

Table 2: **Simulations’ results.** Three different values of \mathcal{R}_0 are computed using first quartile, median and third quartile of each parameter distribution. Here, the exact values of these parameters are listed along with the maximum value reached by each of the four compartments in Figure 5.

3.2 Sensitivity Analysis

In this section, the sensitivity index of each parameter is computed. This index correlates the parameter with the basic reproductive number to further identify the major factors that contribute to the spread of COVID-19 [15]. The sensitivity index is given computing the partial derivative of the \mathcal{R}_0 expression (Equation 14) with respect to the parameter p :

$$C_p^{\mathcal{R}_0} = \frac{\partial \mathcal{R}_0}{\partial p} \times \frac{p}{\mathcal{R}_0} \quad (18)$$

Given the parameters for $\mathcal{R}_0 = 1.23$ from Table 2, the sensitivity indexes are computed.

$$\begin{aligned} C_{\beta_a}^{\mathcal{R}_0} &= \left(\frac{(1-f)}{\gamma} \right) \frac{\beta_a}{\mathcal{R}_0} = 0.7670 \\ C_{\beta_s}^{\mathcal{R}_0} &= \left(\frac{f}{\gamma + \mu + \nu_s} \right) \frac{\beta_s}{\mathcal{R}_0} = 0.2330 \\ C_f^{\mathcal{R}_0} &= \left(\frac{\beta_s}{\gamma + \mu + \nu_s} - \frac{\beta_a}{\gamma} \right) \frac{f}{\mathcal{R}_0} = -0.4017 \\ C_\gamma^{\mathcal{R}_0} &= \left(- \left(\frac{f\beta_s}{\gamma + \mu + \nu_s} \right)^2 - \left(\frac{(1-f)\beta_a}{\gamma} \right)^2 \right) \frac{\gamma}{\mathcal{R}_0} = -0.0728 \end{aligned}$$

$$C_{\nu_s}^{\mathcal{R}_0} = \left(- \left(\frac{f\beta_s}{\gamma + \mu + \nu_s} \right)^2 \right) \frac{\nu_s}{\mathcal{R}_0} = -0.0066$$

$$C_{\mu}^{\mathcal{R}_0} = \left(- \left(\frac{f\beta_s}{\gamma + \mu + \nu_s} \right)^2 \right) \frac{\mu}{\mathcal{R}_0} = -0.0031$$

The parameters β_a and β_s have a positive sensitivity index which means positive significance in the increase of the basic reproduction number: increasing (or decreasing) the value of these parameters, while the other parameters' values remains the same, will contribute to the increase (or decrease) in \mathcal{R}_0 . Instead the recovery rate (γ), the hospitalization rate (ν_s) and the death rate (μ) have a negative sensitivity index, which means that increasing (or decreasing) them will contribute to a decrease (or increase) in \mathcal{R}_0 . So, for example, decreasing the asymptomatic transmission rate by 10% will result in a decrease of \mathcal{R}_0 by 7.67%.

This analysis confirms that asymptomatic individuals need to be controlled to eradicate the virus, as β_a is the more significant parameter for SEIIRHD model. Thus, adequate social distancing, quarantine, contact tracing, and swab policies are needed to decrease the transmission of COVID-19.

3.3 Numerical simulations

The SEIIRHD model has been used to simulate two different phases of the pandemic which have been identified according to the occurrence of some crucial events [16]:

- October 8th: new rules imposing mandatory use of maks in all locations (either indoor or outdoor) accesible to public;
- November 6th: stricter confinement rules including distance learning, further restrictions on commercial activities, limitations on the circulation outside the own municipality (for some Italian regions, classified as *red* regions);
- November 19th: additional confinement rules as more regions turned to red color;
- December 11th: relaxation of confinement rules in some regions turned to yellow color;
- December 20th: first UK variant case in Italy;
- January 11th: relaxation of confinement rules after Christmas holidays, most regions turned to orange and yellow color;
- February 4th: around 20% of Italian cases are linked to the new versions of the virus;
- February 22th: Molise turned to orange color due to many English variant cases;
- March 1st: Sardegna turned to white color, with almost no restrictions.

Considering a time lag to account for the incubation period, the two corresponding phases for which the model parameters are defined, are the following:

- **Second wave**: from October 8th 2020 to November 23rd 2020, to analyze if the trend has changed after the restrictions of November;
- **Third wave**: from January 21st 2021 to February 20th 2021 to observe how the UK variant made the parameters change.

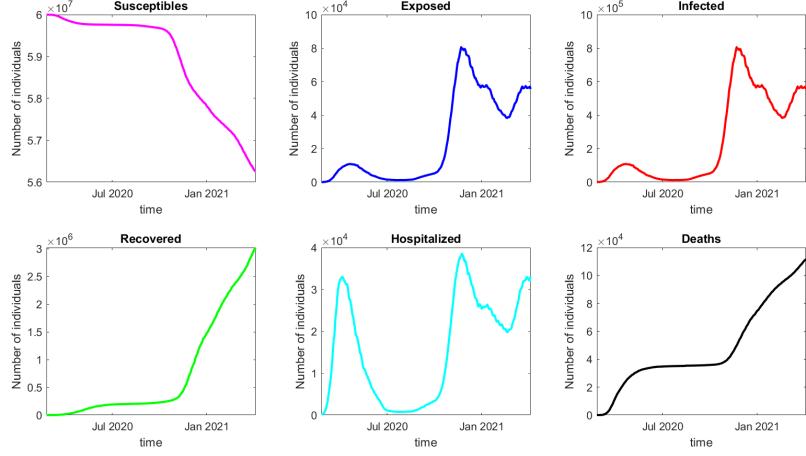


Figure 6: **Overview of COVID-19 pandemic in Italy until 6th April, 2021.** The exposed compartment is defined starting from the infected compartment, for this reason they have a similar behaviour. The three waves are clearly distinguishable.

Along with the simulations of Italy, two regions are considered: Molise and Sardegna. In February, Molise started to deal with the more aggressive UK virus variant, which lead to a high hospitalization rate. Instead, Sardegna was on the road of becoming the first and only "white" region in Italy, with weaker NPIs.

These simulations used the data provided by the Dipartimento della Protezione Civile [17], which report daily and cumulative data such as the number of infected, hospitalized, dead people, swabs and rapid test results. This dataset has been daily updated since February 24th, 2020, which marks the date of the first Italian NPI measure: some towns turned into "red zones" in Lombardia and in Veneto to limit the spread of the virus [16].

The accuracy of these data is highly questioned, in particular concerning the estimate of the total number of infections (strongly dependent on the daily screening effort). From this dataset, four columns were employed: *totale_positivi* for the infected (I); *dimessi_guariti* for the recovered (R); *totale_ospedalizzati* for the hospitalized (H); *deceduti* for deaths (D). The compartments missing in the SEIIRHD model are retrieved as follows:

- susceptibles (S): subtracting from the total population N the values of all other compartments;
- exposed (E): approximating it with the 30% of the current infected individuals;
- infected asymptomatic (I_a) and infected symptomatic (I_s): computed with the parameter f , which is the probability of being symptomatic (i.e $(1 - f)$ is the probability of being asymptomatic).

In Section 2.1, the SEIIRHD model was deduced using three other models: SIR, SEIR and SEIIR. Those models share some parameters with SEIIRHD and they were used in a concatenated fitting process to provide more precise initial conditions. The fitting will be performed in order, SIR, SEIR, SEIIR and SEIIRHD and, if the estimated parameters are present in the next model they are used as initial conditions. For example, the estimated parameters β and γ obtained from the fitting of

the SIR model are used as β_0 and γ_0 for the SEIR model, while α_0 will be initialised (Table 3). To better understand this algorithm, a pseudocode is provided.

Algorithm 1: Concatenated SEIIRHD fitting

Result: $f, \alpha, \gamma, \beta_a, \beta_s, \nu_s, \mu$

Initialize β_0, γ_0 ;
 $[\beta_{SIR}, \gamma_{SIR}] = \text{SIR}(\beta_0, \gamma_0)$;

Initialize α_0 ;
 $[\beta_{SEIR}, \gamma_{SEIR}, \alpha_{SEIR}] = \text{SEIR}(\beta_{SIR}, \gamma_{SIR}, \alpha_0)$;

Initialize $f_0, \beta_{a,0}, \beta_{s,0}$;
 $[f_{SEIIR}, \alpha_{SEIIR}, \gamma_{SEIIR}, \beta_{a,SEIIR}, \beta_{s,SEIIR}] = \text{SEIIR}(f_0, \alpha_{SEIR}, \gamma_{SEIR}, \beta_{a,0}, \beta_{s,0})$;

Initialize ν_0, μ_0 ;
 $\text{SEIIRHD}(f_{SEIIR}, \alpha_{SEIIR}, \gamma_{SEIIR}, \beta_{a,SEIIR}, \beta_{s,SEIIR}, \nu_0, \mu_0)$;

Each fitting is performed using Matlab's optimization `fmincon` function, which minimizes the $L2$ loss function between the reported data and the fitted data at each iteration. This function can assign constraints on the parameters to optimize. A lower and an upper bound are set for all parameters:

$$f \in [0, 0.8] \quad \alpha \in [0, 1] \quad \gamma \in [0, 1] \quad \beta_a \in [0, 1] \quad \beta_s \in [0, 1] \quad \nu_s \in [0, 0.1] \quad \mu \in [0, 0.1]$$

To simulate a realistic scenario, some assumptions are made for f, ν_s and μ :

- the probability of being symptomatic f needs to be smaller than 1 in order not to have the asymptotic compartment empty;
- the hospitalization rate ν_s is bounded not to be higher than the value fitted during the first wave (see Table 3), as it is linked to limited medical resources that are difficult to scale;
- the death rate μ is bounded according to mortality rate observed in different countries [18].

After the concatenated fitting is performed, a forecast is made for the next 14 and 21 days.

Parameters	Second wave		Third wave		
	Italy, Sardegna, Molise		Italy	Sardegna	Molise
	Initial Value	Reference			
β_0	0.9620	[5]	0.1068	0.0637	0.9620
γ_0	0.3700	[5]	0.0134	0.0123	0.3700
α_0	0.1960	[19]	0.2741	0.0884	0.1960
f_0	0.4264	[6]	0.4600	0.5164	0.4264
$\beta_{a,0}$	0.0500	[6]	0.0833	0.0217	0.0500
$\beta_{s,0}$	0.6753	[6]	0.1281	0.0253	0.6753
$\nu_{s,0}$	0.0800	[20]	0.0102	0.0039	0.0800
μ_0	0.0204	[6]	0.0025	0.0022	0.0204

Table 3: **Initial conditions for the second and third wave fittings.** For Italy, the I.C. of the third wave are obtained from the fitting on the first period in Italy (Table 4); the initial conditions for Sardegna are estimated from the fitting on this second period in Italy (Table 4); for Molise initial conditions are the same for both periods.

3.3.1 Italy predictions

As explained in Section 3.3, restrictions on mobility, schools, businesses, and partial lockdowns were introduced in late October at regional and national levels. From Figure 7 (on the left), it is possible to see how these measures have shown good results only after more than two weeks after their introduction. In forecasting, the model overestimated the data and failed to detect the peak reached around the beginning of November.

In January, the trend in Italy was decreasing thanks to the stricter NPIs introduced for Christmas holidays (Figure 7, right). Even if in some regions, like Molise, the percentage of cases linked to the variants was increasing. The model is not able to predict the third wave as it is not able to differentiate between COVID-19 and its UK variant. The SEIIRHD matches correctly during the fitting period, but it underestimates the numbers from February 20st and forecasts a decreasing trend.

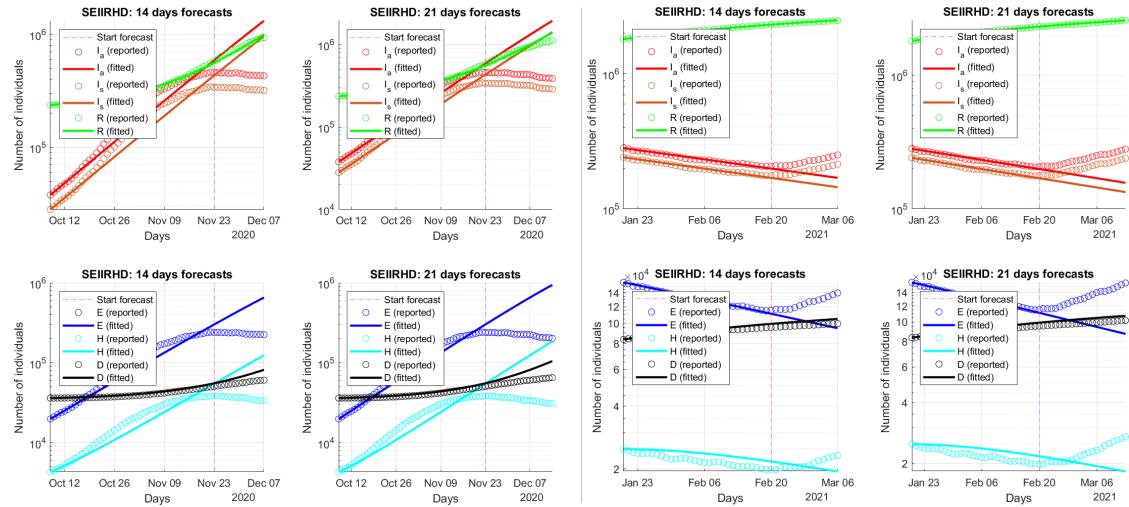


Figure 7: **Fittings Italy, second and third wave.** The plot on the left show the fitting and the forecasting of the second wave, while on the right the third wave. The first forecast starts from November 23th while the second starts from February 20st.

3.3.2 Molise predictions

Differently from Italy, in Molise both periods present an increasing trend. At the beginning of 2021, it was one of the first regions where COVID-19 mutations start spreading. The dominant one is Lineage B.1.1.7, commonly called the *UK variant*. This mutation was sequenced back in September 2020 in the UK and it is becoming the dominant one in Italy (as seen in Section 3.3.1). Recent studies have demonstrated that this variant is more transmissible and can cause more severe symptoms [21].

While the trends for the second outbreak are similar to the Italian one, on the left, in Figure 8, it can be noticed a different tendency that can be associated with the presence of the UK variant. The hospitalized compartment shows fluctuations, probably due to an incorrect data recording of the data, that the model is not able to fit. This fluctuations can be also seen in the infected compartments. Besides this behaviour, the model is able to recognize the increasing trend from the data and its predictions are more accurate than the Italian ones even if the model is not able to identify the UK variant.

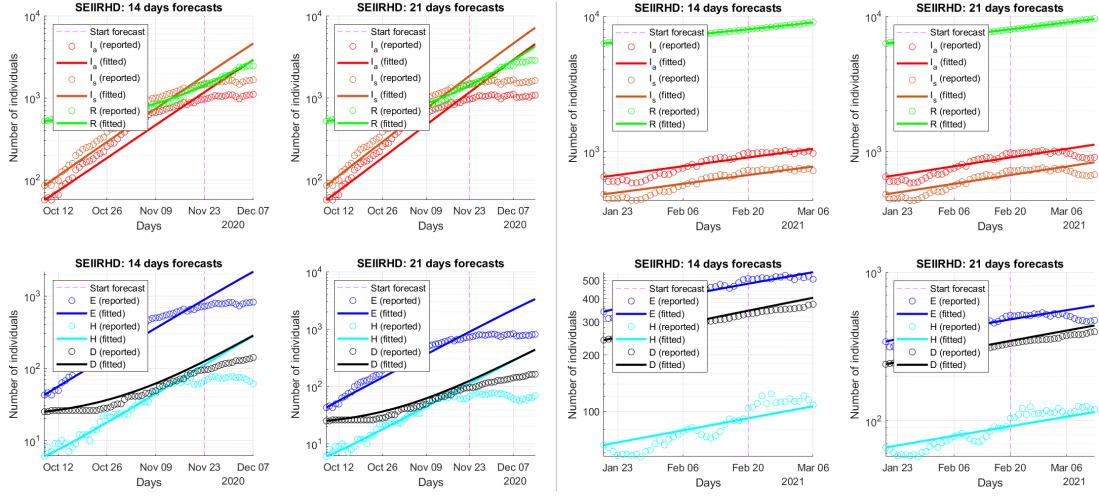


Figure 8: **Fittings Molise, second and third wave.** The plots on the left show the fittings and the forecastings for the second wave, while on the right there are the ones for the third wave.

3.3.3 Sardegna predictions

For Sardegna the SEIIRHD model has a good fit until November 23rd and it overestimates the numbers in the forecasting period, as it can be seen for the left of Figure 9.

Regarding the third wave, the model is able to fit and to make good predictions for five out of six compartments of the SEIIRHD model. Several attempts were made to improve the fitting of the hospitalized compartment, but they all proved unsuccessful. This decreasing trend also for the following 21 days shows that the UK variant is still not predominant in this region. This tendency will allow the region to turn to the "white" category (lower NPIs) by the beginning of March.

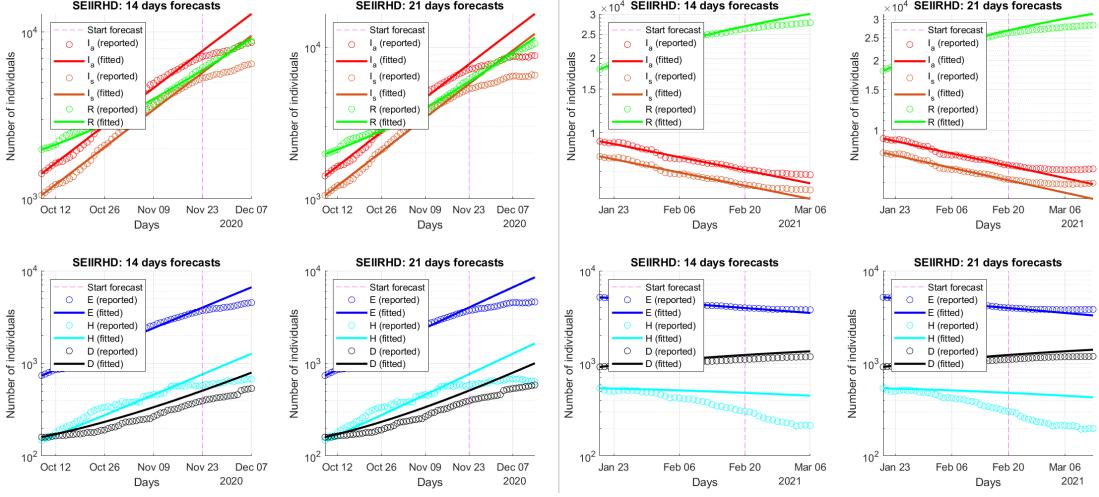


Figure 9: **Fittings Sardegna, second and third wave.** The plot on the left show the fitting and the forecasting of the second wave, on the right the third wave.

3.3.4 Summary of numerical simulations

In Table 4, the results obtained from the SEIIRHD model and the corresponding \mathcal{R}_0 values are reported. The basic reproduction numbers of the second wave are quite high, in fact the number of infected was increasing really fast due to higher transmission rates. For example, $\mathcal{R}_0 = 4.85$ for Italy means that one infected individual will transmit it to an average of 4 to 5 people, which is really high. This is the reason why Italian Government introduced stricter NPIs. Molise has a larger \mathcal{R}_0 value than Italy, even with similar transmission rates. It is probably connected with an higher probability of being symptomatic, but a lower hospitalization rate. So there are more symptomatic infected individuals and hospitals in Molise have some difficulties in taking care all of them.

The \mathcal{R}_0 of Molise is higher than Italy and Sardegna also during the third wave. By recalling the sensitivity analysis performed in Section 3.2, this is connected to higher transmission rates, β_a and β_s , and to a lower probability f of being symptomatic. Moreover, the death rate is the highest among all fitted values, which confirms that these parameters, quite different from the one of Italy and Sardegna, are related to the dominant presence of the UK variant in Molise.

In Table 4, the τ parameter is introduced. It is nothing but the inverse of α and it describes the incubation period of COVID-19 in days. When the trend is decreasing, this parameter becomes very large. When the infected compartment is decreasing, the model wrongly interprets this trend by guessing that the exposed individuals are showing symptoms more slowly.

Parameters	Second wave			Third wave		
	Italy	Molise	Sardegna	Italy	Molise	Sardegna
f	0.4600	0.6363	0.4679	0.5164	0.4820	0.5211
β_a	0.0833	0.0564	0.0708	0.0217	0.0508	0.0076
β_s	0.1281	0.0754	0.0103	0.0253	0.0692	0.0087
γ	0.0200	0.0198	0.0132	0.0351	0.0391	0.0183
α	0.2741	0.3061	0.1774	0.0884	0.1803	0.0356
$\tau(\frac{1}{\alpha})$	3.6483	3.2662	5.6369	11.3122	5.2500	28.080
ν_s	0.0102	0.0053	0.0068	0.0039	0.0075	0.0012
μ	0.0025	0.0035	0.0024	0.0022	0.0052	0.0014
\mathcal{R}_0	4.8531	5.5499	4.4216	0.6197	1.4386	0.4199

Table 4: **SEIIRHD results.** These are the numerical results obtained from fittings described in the previous sections (Figure 7 for Italy, Figure 8 for Molise, Figure 9 for Sardegna).

3.4 LSTM: a deep learning model for predictions

The methods for forecasting time-series data can be mainly classified into two types, machine learning and deep learning methods. Deep learning models are superior over the statistical machine learning models for dealing with non-linear functions, just like the case of COVID-19 pandemic. Recurrent Neural Networks (RNNs) are a powerful and robust type of neural networks that uses existing time-series data to predict future data over a specified period. In general, they can process a sequence of vectors x by applying a recurrence formula at every time step:

$$h_{(t)} = f_W(h_{(t-1)}, x_{(t)})$$

where $h_{(t)}$ is the new (hidden) state, $h_{(t-1)}$ is the old (hidden) state, $x_{(t)}$ is an input vector at some time step and f_W is some function with parameters W , that are the weights of the layers.

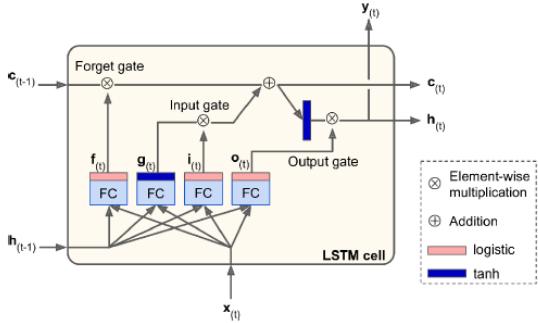


Figure 10: **LSTM cell.** The figure show the block diagram with all the layers and operations inside a LSTM cell [9].

In these experiments **LSTM cells** are used and a detailed representation of their architecture can be seen in figure 10.

It maintains two hidden states at every time step: $h_{(t)}$ and $c_{(t)}$, short-term state and long-term state [9]. The previous long-term state $c_{(t-1)}$ crosses the network from left to right. The current input vector $x_{(t)}$ and the previous short-term state $h_{(t-1)}$ are fed to four different fully connected layers: three of them are gates. Each gate receives a number between 0 and 1 as it is the output of the logistic function: if the value is close to 1, the gate opens.

The behaviour of the cell can be described in the following way:

- **output** $g_{(t)}$ filters the relevant input from $x_{(t)}$ and $h_{(t-1)}$;
- **forget gate** $f_{(t)}$ controls which parts of the long-term state $c_{(t-1)}$ should be erased;
- **input gate** $i_{(t)}$ controls which parts of $g_{(t)}$ should be added to the long-term state $c_{(t-1)}$;
- **output gate** $o_{(t)}$ controls which parts of the long-term state $c_{(t-1)}$ should be read and output at this time step.

These four main components of LSTM will work and interact in a special manner, as it accepts long-term memory, short-term memory, input sequences at a given time step and generates new long-term memory, new short-term memory and output sequence at a given time step. In this report, the implementation of LSTM uses a sequential approach: the first two layers are LSTM layers with 50 inner neurons each, a fully connected layer and the output layer with a linear activation function.

Deep learning methods require a lot of data as input for training. However, training one single model for all countries across the world raises a lot of problems: not all countries show the same trend in infection and testing, while registration is different for multiple countries, so the data could show a lot of different fluctuations and peaks. A solution to this problem is to train a model using countries' data with similar trends. The European countries which had exceeded 1 million COVID-19 cases were chosen: *Russia, United Kingdom, France, Spain, Germany, Poland, Ukraine, Czech Republic and Netherlands*. The data used for these experiments are obtained from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [22]. The dataset for each country comprises of three time-series data sets: infected cases, cumulative recovered cases, and cumulative fatalities. Instead, the results for the hospitalized compartment were obtained using only Italian data.

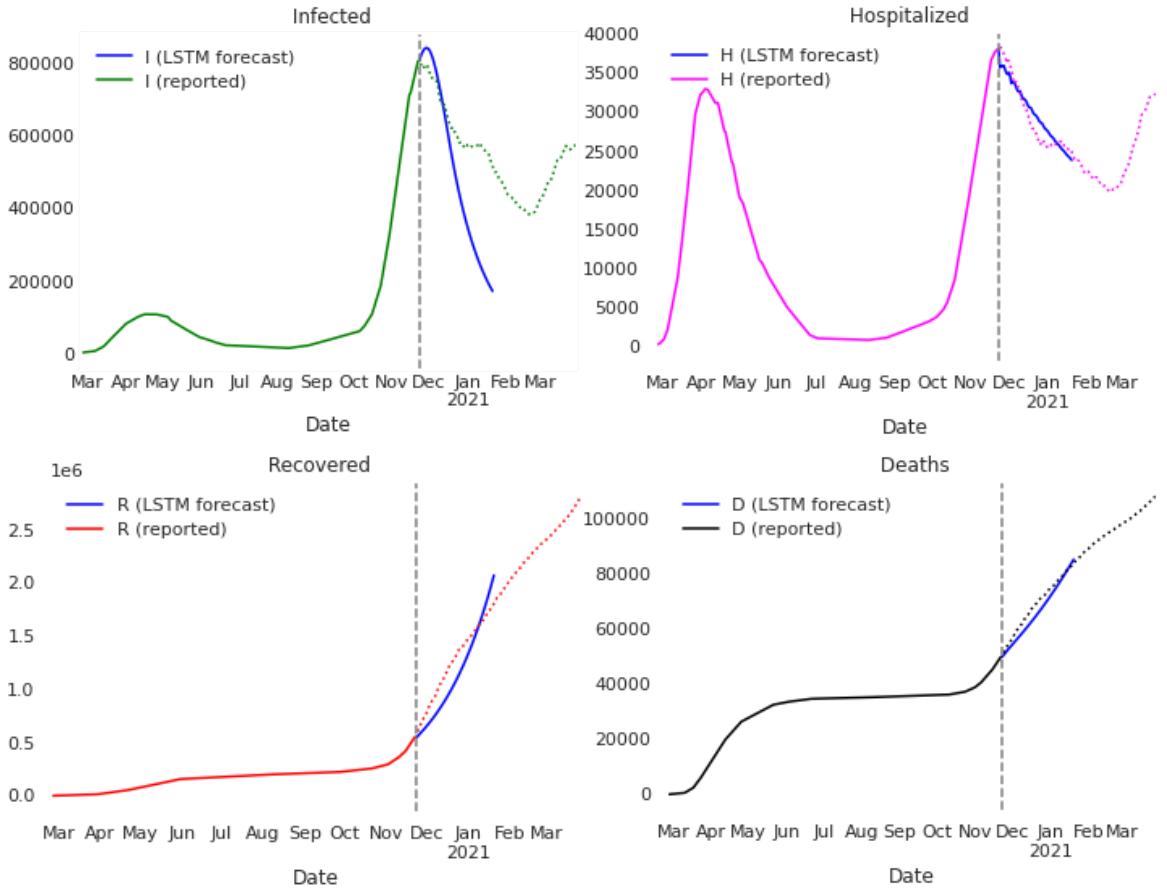


Figure 11: LSTM predictions second wave These plots describes a 60 days forecasts starting on November 23th. The dotted lines are the reported data for the forecast period, while the blue line is the forecast provided by the LSTM.

Two forecasting periods were considered: 60 days forecasting starting from November 23rd (Figure 11) and a 40 days forecast starting from February 20th (Figure 12). During these tests, four hyperparameters were tuned: the number of epochs, the batch size, the number of neurons in the fully connected layer, and the time window used by LSTM to make its predictions.

- **Epochs:** it defines the number of times that the algorithm will work through the entire training dataset. The tuning was done over 10, 20, 30, 50, 70 epochs;
- **Batch size:** for each epoch, the dataset is split in subsets, called batches. Their size is fixed to the batch size, and it affects the frequency of network weights updates. The values for the tuning are 8, 12, 24 and 32;
- **Number of neurons of the fully connected layer:** it affects the training process, in general, greater than the number of neurons makes the training process longer and more prone to overfitting. Four different values were analyzed 6, 8, 12, 14;
- **Time window:** the time window is the amount of data considered at a time. The following values were analyzed: 3, 5, 7, 10.

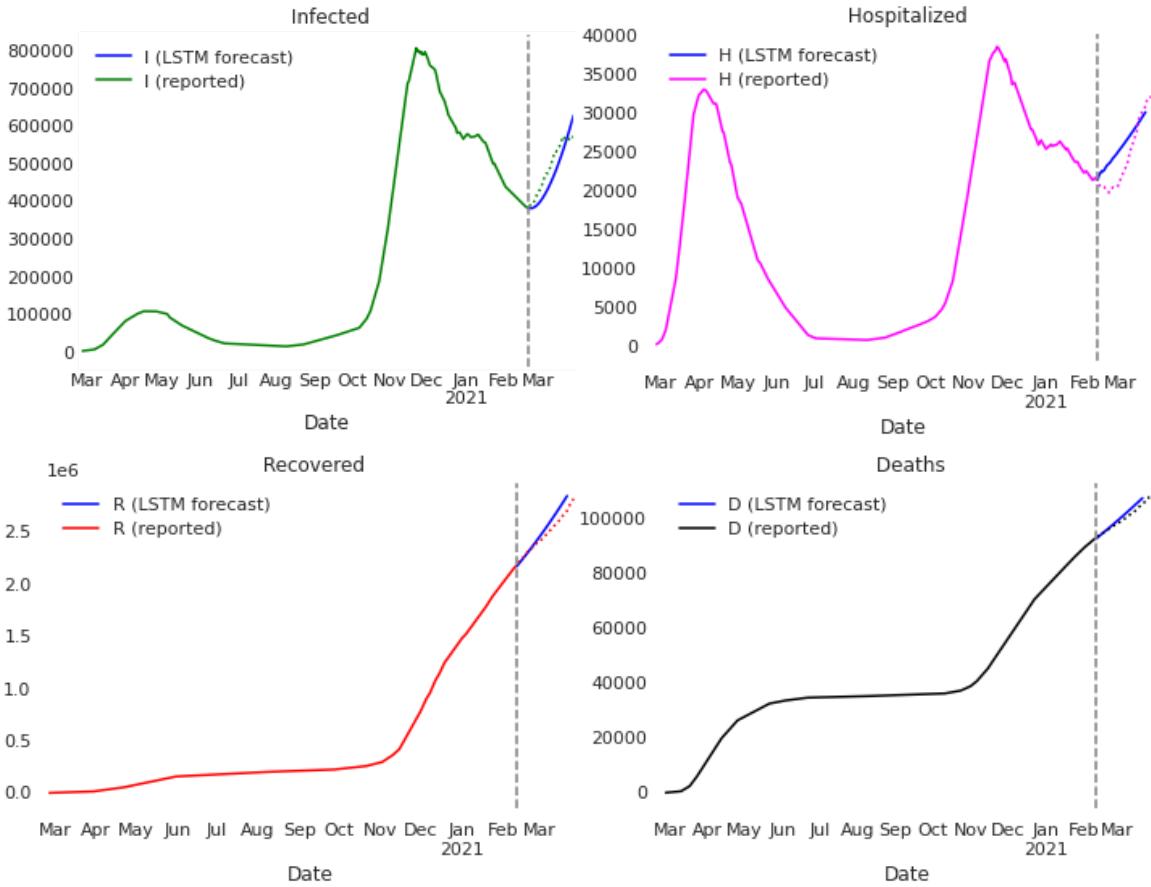


Figure 12: **LSTM predictions third wave.** These plots describes a 40 days forecasts starting on February 21th. The dotted lines are the reported data for the forecast period, while the blue line is the forecast provided by the LSTM.

The results for each compartment are shown in Table 5. For the infected compartment, the network is able to correctly predict the peak of the second wave and the concavity of the third one. Also for the deaths and recovered compartments, the model is able to fit quite well both waves. Instead, the hospitalized compartment suffers from lack of data, indeed the majority of deep learning models would have struggled to select the optimal parameters on a such small dataset.

Compartment	Second wave				Third wave			
	Epochs	Batch	Neurons	T. window	Epochs	Batch	Neurons	T. window
Infected	20	12	12	5	20	12	12	5
Hospitalized	20	8	8	5	20	8	8	7
Recovered	20	14	14	10	20	14	14	7
Deaths	50	14	14	7	50	14	14	10

Table 5: **LSTM hyper-parameters tuning results.** Every compartment was tuned for both the second and third wave. The hyper-parameters tuning was done over number of epochs, batch size, number of neurons of the Dense layer and the time window used by LSTM to forecast.

3.5 A comparison between SEIIRHD and LSTM models

Three performance measures are used to evaluate and compare the forecasting performances of the two proposed models for the second wave (60 days starting from November 23th) and for the third wave (40 days starting from February 21st). These are the mean absolute error (MAE) in Eq. 19, root means square error (RMSE) in Eq. 20 and R_2 score in Eq. 21:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (20)$$

$$R_2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

where y denotes the observed data, \bar{y} the mean of the observed data, \hat{y} the estimated value and n is the number of predictions, so it is equal to 60 for the second wave and to 40 for the third.

Both MAE and RMSE express the average model prediction error and they are negatively oriented scores, which means lower values are better. They can range from 0 to infinity, while R_2 is usually a value between 0 and 1. The closer R_2 is to 1, the better our model will be at predicting. If the chosen model fits worse than a horizontal line, then R_2 can be negative, it means that the chosen model does not follow the trend of the data, so fits worse than a horizontal line.

The obtained errors are shown in Table 6. Both models obtained good results for Recovered and Deaths compartments, with an R_2 score close to one, which is a considerable achievement for real-world data. LSTM had better results for Infected compartment, with lower MAE and RMSE scores than SEIIRHD and a positive R_2 . Even if it is not as close to one as before, an R_2 of 0.27 means that a full 27% of the variation is completely explained by the predictions, which is still a good result. The trend of the Hospitalized compartment was the more difficult to predict also for LSTM, due to the small size of the training dataset. The SEIIRHD had poor performance for the Infected and Hospitalized as it was not able to predict the corresponding peaks.

Compartment	Second wave					
	SEIIRHD			LSTM		
	MAE	RMSE	R_2	MAE	RMSE	R_2
Infected	2489632.79	3235121.77	-0.86	156992.10	201420.69	0.27
Hospitalized	378468.57	508199.33	-0.78	3702.55	3935.55	-0.25
Deaths	125326.99	178037.58	0.91	2171.30	2819.07	0.94
Recovered	1655269.79	2550386.67	0.91	138311.86	186830.11	0.87
Third wave						
Infected	71694.57	83659.48	-0.99	186830.11	94799.61	0.56
Hospitalized	5784.88	7492.38	-0.98	1945.79	2302.81	0.55
Deaths	4603.20	4629.50	0.99	2671.78	3393.29	0.67
Recovered	17408.67	25839.88	0.99	20282.58	20282.58	0.98

Table 6: **LSTM and SEIIRHD performance measures.** These results are obtained from 60 days of forecast starting from November 23th for the second wave and from a 30 days of forecast starting from February 21st. The infected compartment considered for the SEIIRHD model is I_s .

Even if LSTM proved to have better forecasting abilities, it has some limitations:

- it makes poor predictions without enough data available, a problem that makes it useless in the first stages of a new disease;
- it is not able to distinguish between asymptomatic and symptomatic, a key distinction for COVID-19;
- it does not provide any information about rates or \mathcal{R}_0 value, so it is more difficult to understand, for example, why the number of infected individuals is increasing or to decide which NPI should be introduced to contain the transmission.

The last two issues are linked to the best-known disadvantage of neural networks: their “black box” nature. As it is not clear why the neural network came up with a certain output, it is very hard to understand what caused it to arrive at this prediction.

3.6 When the pandemic will end?

To further analyze these models, simulations of the end of the pandemic are run for Infected and Deaths compartments. The parameters used for the SEIIRHD model are obtained from the results of the third wave in Italy (Table 4). To forecast one single day, LSTM looks back at a fixed number of days, which is described by the time window hyper-parameter. It means that predictions for a longer period of time are made looking at previously predicted data, this can lead to errors.

The infected compartments have a similar behaviour, but different results: by June 2022 the SEIIRHD reaches zero infected individuals (Table 7), while LSTM is not able to achieve this result at all. Without any prior assumption, LSTM has probably obtained the more realistic result.

Instead, regarding the deaths compartment, LSTM is not able to reach an asymptote like SEIIRHD. The number of deaths is increasing unconditionally, even if the number of infected individuals is decreasing. This unrealistic behaviour is due to the fact that LSTM makes its predictions without considering other factors (i.e, the number of infected individuals). It only mimics the behaviour of real data, which are cumulative, and so it has no way to understand that, at a certain point, the number of deaths would reach a maximum value.

Furthermore, in these simulations the SEIIRHD model reaches a peak for the infected compartment really fast, as observed in Figure 5. For the Deaths compartment, it reaches an asymptote but it is a quite high value (almost 2M of deaths). This unrealistic behaviour is caused by the strong assumptions made for the mathematical model.

Date	Infected		Deaths	
	SEIIRHD	LSTM	SEIIRHD	LSTM
June 2021	2.82M	172.24K	1.85M	129.99K
September 2021	63.06K	24.00K	1.85M	1.62M
December 2022	1469	3560	1.85M	1.97M
March 2022	35	2589	1.85M	2.24M
June 2022	0	1328	1.85M	2.93M
September 2022	0	873	1.85M	3.57M
December 2022	0	726	1.85M	4.33M

Table 7: **LSTM and SEIIRHD predictions for the end of the pandemic:** number of individuals for each compartment on the first day of the month. The number of infected individuals for the SEIIRHD model is the sum of the symptomatic and the asymptomatic individuals ($I_a + I_s$).

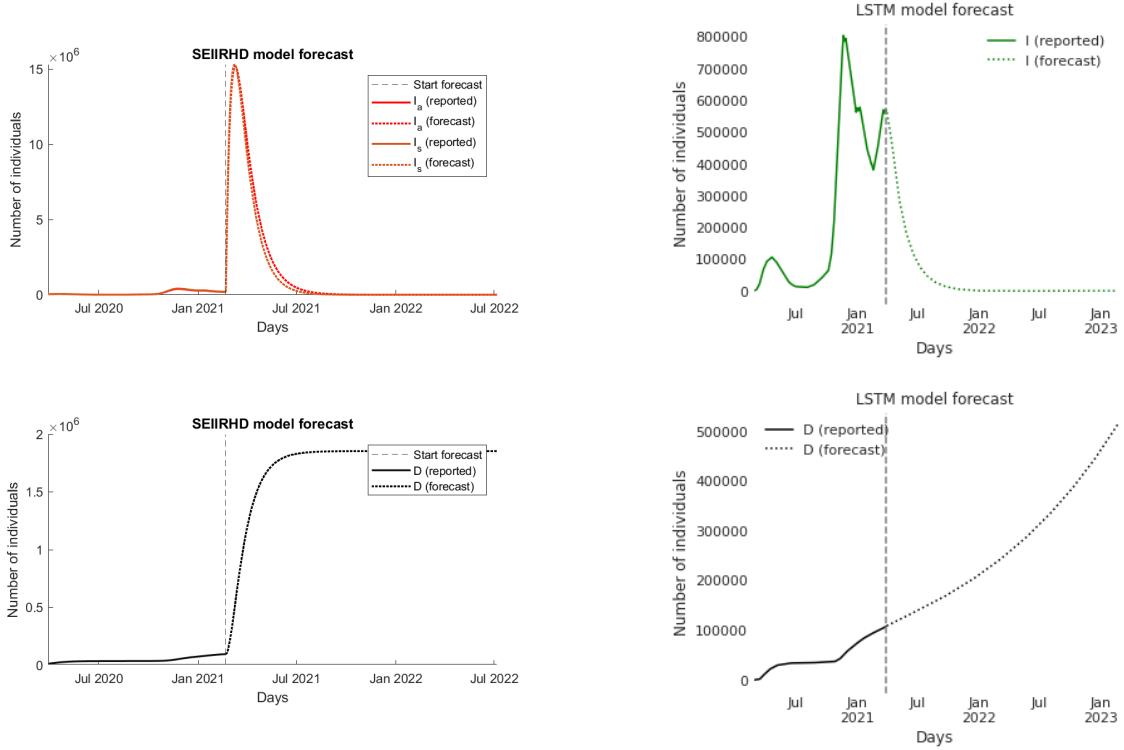


Figure 13: **LSTM and SEIIRHD predictions for the end of the pandemic.** The simulations for the SEIIRHD model are on the left, while the ones for LSTM are on the right. The forecast is done starting from March 31th 2021.

4 Conclusions and further work

In this paper, a novel epidemiological model named SEIIRHD (Section 2.1) is introduced to describe the ongoing pandemic of COVID-19. This model permits the division of the infected compartment into asymptomatic and symptomatic and it introduces the hospitalized and deaths compartments. Their analysis permits to have a clear overview of the pandemic situation in Italy. In Section 2.2 it is shown that this model has one equilibrium, disease-free equilibrium, that is stable with $\mathcal{R}_0 < 1$.

In Section 3.1, some simulations show how little variations of \mathcal{R}_0 value can result in higher resulting values of infected, hospitalized, and death cases. A deeper analysis for \mathcal{R}_0 is done in Section 3.2, where a sensitivity analysis shown that the most significant value for the \mathcal{R}_0 value is the infection rate β_a . It is probably connected to the difficulty in detecting asymptomatic individuals as they do not develop symptoms. A correct swab use and contact tracing are required to check new cases among the population before the transmission rate becomes really high.

In Section 3.3, the numerical simulations are presented for Italy and for two different regions (Molise and Sardegna). During the third wave, these two regions had a really different behaviour because of the novel UK variant of the COVID-19. In Section 3.4, it is introduced a deep learning model called LSTM to see how well it was able to forecast the Italian trend for the two periods used for the SEIIRHD model. In general, neural networks need a great amount of data to be trained, so datasets from other European countries with a similar behaviour were used for the training phase.

A clear comparison between the fittings of LSTM and SEIIRHD is done in Section 3.5.

To conclude this work, an attempt of predicting the end of the ongoing pandemic is done (Section 3.6). Using both LSTM and SEIIRHD, long-term simulations were run for Infected and Deaths compartments. These analysis made clear that neural networks are really powerful, but they need to be helped by other tools. LSTM had overcome the fittings of the SEIIRHD model, but it had obtained illogical results in the long-term prediction for the Deaths compartment. The main drawback of LSTM is that it does not consider the dependencies of the other compartments, which are essential to study the impact of COVID-19.

In the future, the proposed SEIIRHD model can be extended by introducing additional factors like environmental transmission, effect of vaccines, treatment strategies, effect of delay, impact of unlocking, etc. In addition to that, a merge between the two proposed methods can be examined as in [23], using a neural network to help the SEIIRHD model to find more precise parameters according to real-data.

List of Figures

1	Block diagram SIR. Arrows represent infection flow; β is the infection rate and γ is the recovery rate	2
2	Block diagram SEIR. The flow is similar to the SIR one, the only new parameter is α which describes the inverse of the incubation period.	2
3	Block diagram SEIIR. The infected compartment is splitted into asymptomatic I_a and symptomatic I_s , β_a and β_s are the infection rates, f is the probability of being symptomatic.	3
4	Block diagram SEIIRHD. The infected are divided in symptomatic and asymptomatic. The symptomatic individuals can be hospitalized, recover or die; instead the asymptomatic ones can only recover. The parameters γ and α are the same as SEIR model; β_a and β_s are the infection rates; f is the probability of being symptomatic; ν_s is the intervention rate; μ is the mortality rate.	4
5	Compartment evolution of the SEIIRHD model. First quartile (green line), median (blue line) and third quartile (red line) were obtained from 1000 simulations conducted using values of parameters extracted from uniform distributions reported in Table 1.	9
6	Overview of COVID-19 pandemic in Italy until 6th April, 2021. The exposed compartment is defined starting from the infected compartment, for this reason they have a similar behaviour. The three waves are clearly distinguishable.	12
7	Fittings Italy, second and third wave. The plot on the left show the fitting and the forecasting of the second wave, while on the right the third wave. The first forecast starts from November 23 th while the second starts from February 20 st	14
8	Fittings Molise, second and third wave. The plots on the left show the fittings and the forecastings for the second wave, while on the right there are the ones for the third wave.	15
9	Fittings Sardegna, second and third wave. The plot on the left show the fitting and the forecasting of the second wave, on the right the third wave.	15
10	LSTM cell. The figure show the block diagram with all the layers and operations inside a LSTM cell [9].	17
11	LSTM predictions second wave These plots describes a 60 days forecasts starting on November 23th. The dotted lines are the reported data for the forecast period, while the blue line is the forecast provided by the LSTM.	18
12	LSTM predictions third wave. These plots describes a 40 days forecasts starting on February 21th. The dotted lines are the reported data for the forecast period, while the blue line is the forecast provided by the LSTM.	19
13	LSTM and SEIIRHD predictions for the end of the pandemic. The simulations for the SEIIRHD model are on the left, while the ones for LSTM are on the right. The forecast is done starting from March 31 th 2021.	22

List of Tables

1	Values of the parameters of the SEIIRHD model. For each parameter, it is provided a description along with the interval used to define the uniform distributions for the simulations.	8
2	Simulations' results. Three different values of \mathcal{R}_0 are computed using first quartile, median and third quartile of each parameter distribution. Here, the exact values of these parameters are listed along with the maximum value reached by each of the four compartments in Figure 5.	10
3	Initial conditions for the second and third wave fittings. For Italy, the I.C. of the third wave are obtained from the fitting on the first period in Italy (Table 4); the initial conditions for Sardegna are estimated from the fitting on this second period in Italy (Table 4); for Molise initial conditions are the same for both periods.	13
4	SEIIRHD results. These are the numerical results obtained from fittings described in the previous sections (Figure 7 for Italy, Figure 8 for Molise, Figure 9 for Sardegna).	16
5	LSTM hyper-parameters tuning results. Every compartment was tuned for both the second and third wave. The hyper-parameters tuning was done over number of epochs, batch size, number of neurons of the Dense layer and the time window used by LSTM to forecast.	19
6	LSTM and SEIIRHD performance measures. These results are obtained from 60 days of forecast starting from November 23th for the second wave and from a 30 days of forecast starting from February 21st. The infected compartment considered for the SEIIRHD model is I_s	20
7	LSTM and SEIIRHD predictions for the end of the pandemic: number of individuals for each compartment on the first day of the month. The number of infected individuals for the SEIIRHD model is the sum of the symptomatic and the asymptomatic individuals ($I_a + I_s$).	21
8	Parameters description.	28

References

- [1] World Heath Organization. WHO Timeline - COVID-19. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline>.
- [2] Madani T. A. Ntoumi F. Kock R. Dar O. Ippolito G. Mchugh T. D. Memish Z.A. Drosten C. Zumla A. Hui D. S., Azhar E. I. and Petersen. E. The continuing 2019-ncov epidemic threat of novel coronaviruses to global health- the latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, page 91:264–266, 2020.
- [3] Bernoulli D. Réflexions sur les avantages de l'inoculation. *Mercure de France*, pages 173–190, June 1760.
- [4] Kermack W. O. and McKendrick A. G. Contribution to the Mathematical Theory of Epidemics. *Proc Roy. Soc. Lond. A.A115*, pages 700–721, 1927.
- [5] Faranda D. and Alberti T. Modeling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(11):111101, Nov 2020.
- [6] Cheryl Q. Mentuda Jayrold P. Arcede, Randy L. Caga-anan and Youcef Mammeri. Accounting for Symptomatic and Asymptomatic in a SEIR-type model of COVID-19. <https://arxiv.org/pdf/2004.01805.pdf>, 2020.
- [7] Antonietti P. F. Ardenghi G. Manzoni A. Miglio E. Pugliese A. Verani M. Parolini N., Dede' L. and Quarteroni A. SUIHTER: A new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy. <https://arxiv.org/pdf/2101.03369.pdf>, 2021.
- [8] Singh H.P. Kumari P. and S. Singh. Seiaqrdt model for the spread of novel coronavirus (covid-19): A case study in india. *Appl Intell*, 2020.
- [9] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent System*. Oreilly Associates Inc, 2019.
- [10] Kaczorowska J. Hoste A.C.R. et al. Edridge, A.W.D. Seasonal coronavirus protective immunity is short-lasting. *Nat Med* 26, page 1691–1693, 2020.
- [11] Jung H. Sohraby K. Dadlani A., Afolabi R. O. and Kim K. Deterministic models in epidemiology: From modeling to implementation. <https://arxiv.org/pdf/2004.04675v1.pdf>, 2020.
- [12] Smith RJ Heffernan JM and Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface*, pages 2(4):281–293, 2005.
- [13] Van den Driessche P. and Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1):29–48, 2002.
- [14] European Centre for Disease Prevention and Control. Guidelines for the implementation of non-pharmaceutical interventions against COVID-19 . <https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-guidelines-non-pharmaceutical-interventions-september-2020.pdf>, Sept 2020.
- [15] Resmawan R. and Yahya L. Sensitivity Analysis of Mathematical Model of Coronavirus Disease (COVID-19) Transmission. *Cauchy*, 2020.
- [16] Governo Italiano. Coronavirus, le misure adottate dal governo. <http://www.governo.it/it/coronavirus-misure-del-governo>.
- [17] Sito del Dipartimento della Protezione Civile. Dati COVID-19 Italia. <https://github.com/pcm-dpc/COVID-19>.
- [18] Johns Hopkins University. Mortality Analyses. <https://coronavirus.jhu.edu/data/mortality>, 2021.
- [19] Stephen A Lauren et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine* vol. 172, pages 577–582, Sep 2020.

- [20] Sito del Istituto Superiore di Sanità. Monitoraggi settimanali COVID-19. <https://www.iss.it/monitoraggio-settimanale>.
- [21] Jarvis C.I. CMMID COVID-19 Working Group. et al. Davies, N.G. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*, 2021.
- [22] Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 Data Repository. <https://github.com/CSSEGISandData/COVID-19>.
- [23] Carmichael Z.-Daram A. Dimpy P. S. Clark K. Potter L. Soures N., Chambers D. and Kudithipudi D. Sirnet: Understanding social distancing measures with hybrid neural network model for covid-19 infectious spread, 2020.

5 Appendix

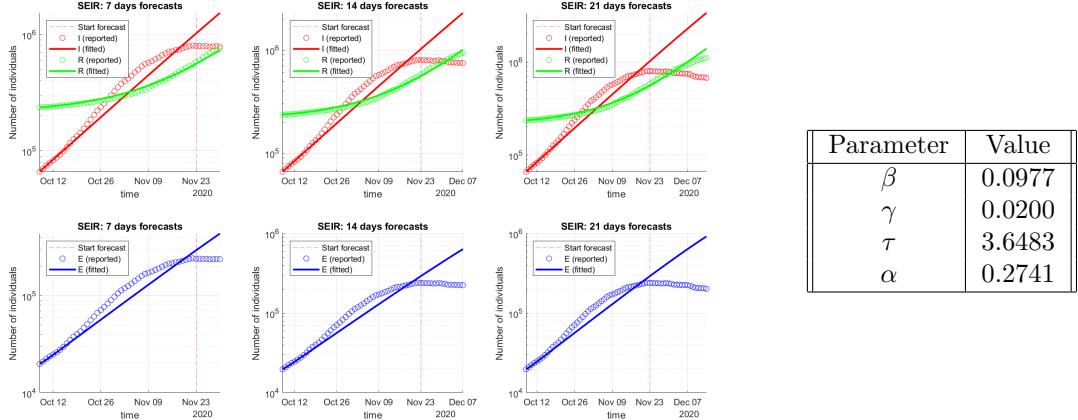
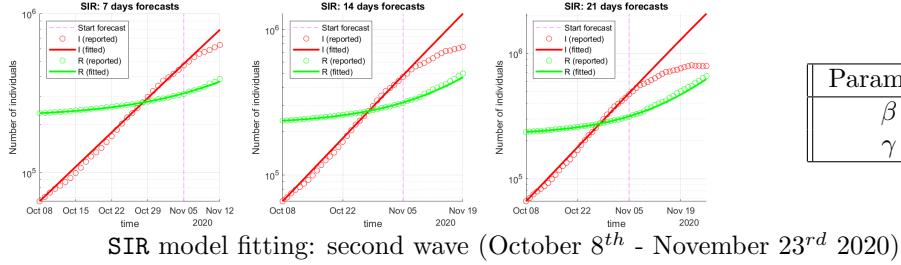
5.1 SIR, SEIR and SEIIR models fitting

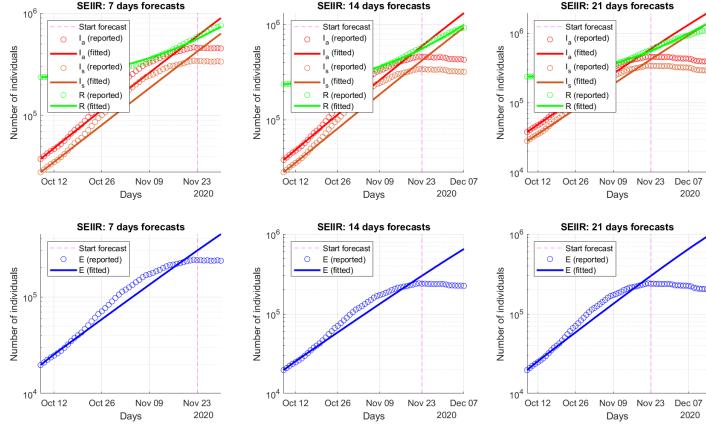
The results of the SIR, SEIR and SEIIR models are reported. As before, these models are tested on Italy, Molise, and Sardegna data.

Parameter	Description
β	Infection rate
γ	Recovery rate
τ	Incubation period
α	Inverse of the incubation period
f	Probability of being symptomatic
β_a	Transmission rate from S to E from contact with I_a
β_s	Transmission rate from S to E from contact with I_s

Table 8: Parameters description.

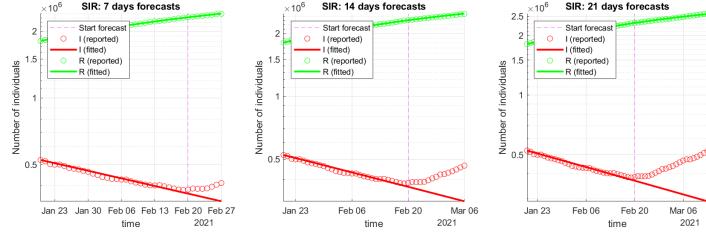
5.1.1 Italy





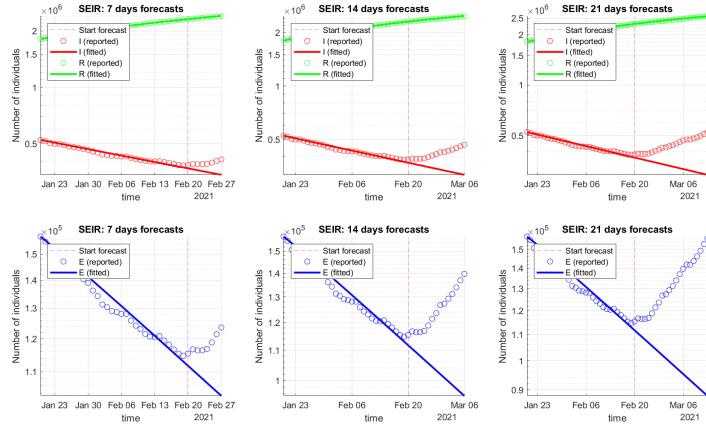
Parameter	Value
f	0.4134
β_a	0.0144
β_s	0.2146
γ	0.0200
τ	3.7864
α	0.2641

SEIIR model fitting: second wave (October 8th - November 23rd 2020)



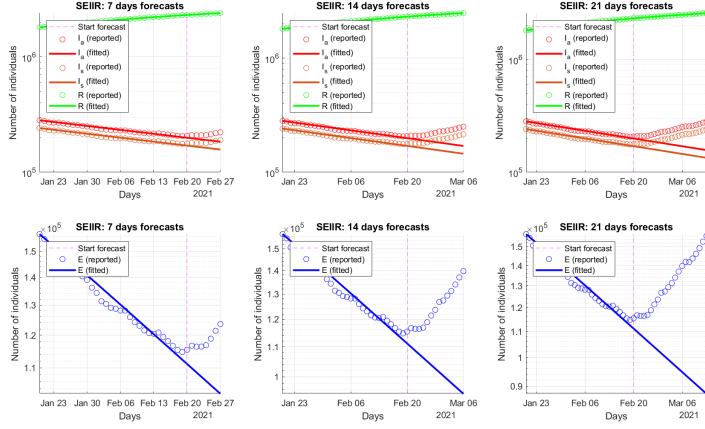
Parameter	Value
β	0.0267
γ	0.0369

SIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
β	0.0233
γ	0.0369
τ	11.7924
α	0.0848

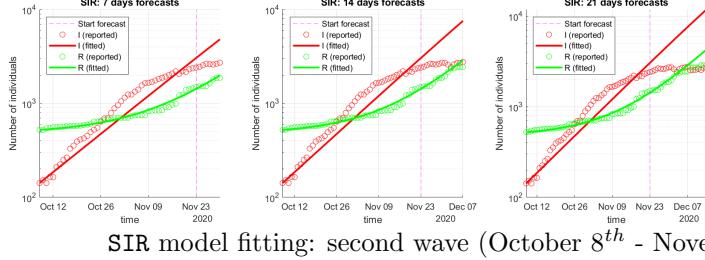
SEIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
f	0.4608
β_a	0.0209
β_s	0.0242
γ	0.0369
τ	11.72
α	0.0853

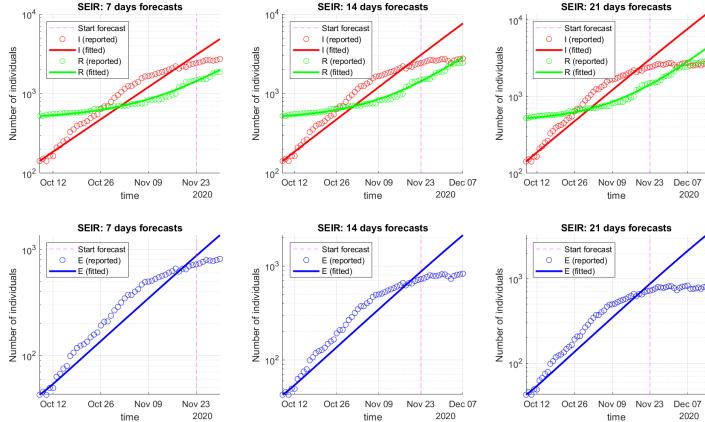
SEIIR model fitting: third wave (January 21st - February 11th 2021)

5.1.2 Molise



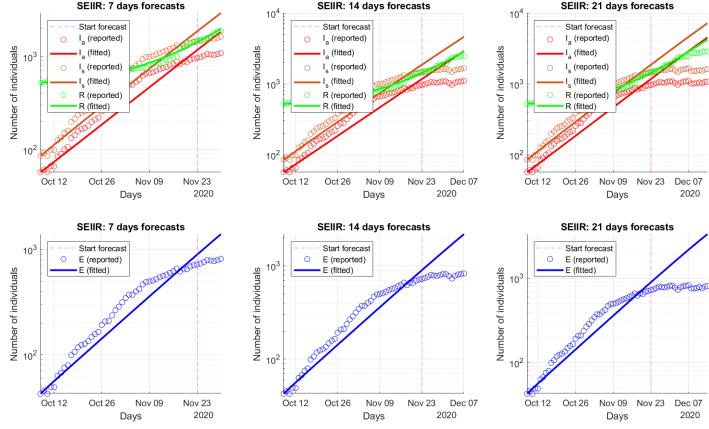
Parameter	Value
β	0.0877
γ	0.0207

SIR model fitting: second wave (October 8th - November 23rd 2020)



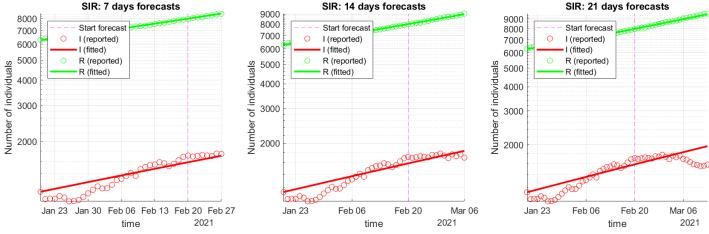
Parameter	Value
β	0.1067
γ	3.2949
τ	3.6483
α	0.3035

SEIR model fitting: second wave (October 8th - November 23rd 2020)



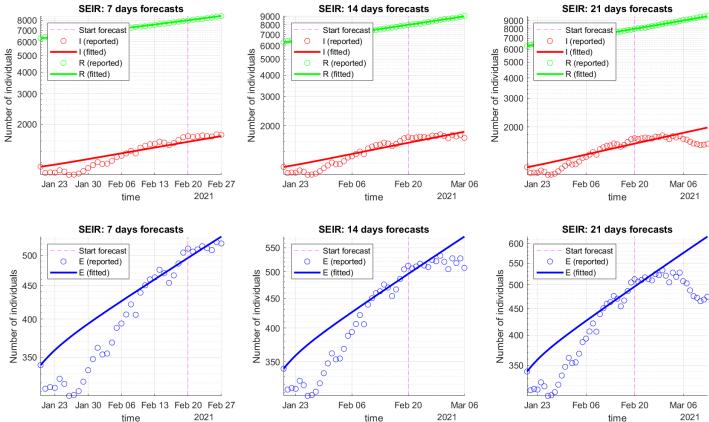
Parameter	Value
f	0.6141
β_a	0.2152
β_s	0.0384
γ	0.0206
τ	3.4300
α	0.2907

SEIIR model fitting: second wave (October 8th - November 23rd 2020)



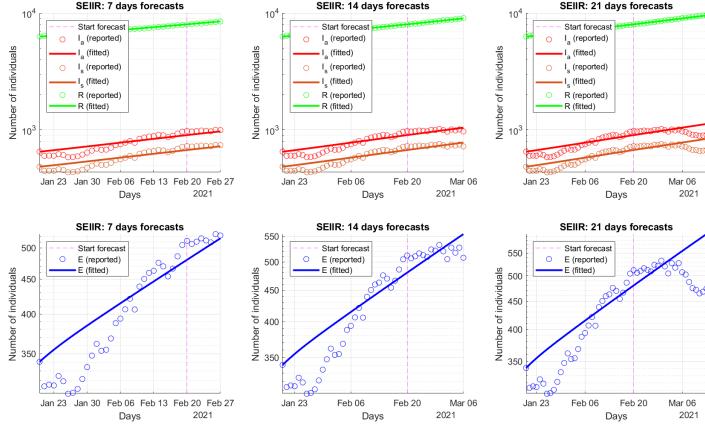
Parameter	Value
β	0.0533
γ	0.0411

SIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
β	0.0576
γ	0.0413
τ	5.9916
α	0.1669

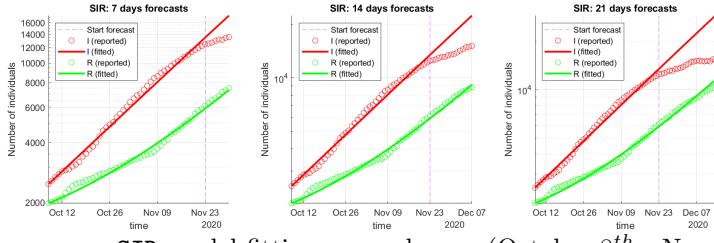
SEIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
f	0.4272
β_a	0.0489
β_s	0.0643
γ	0.0413
τ	5.7803
α	0.1703

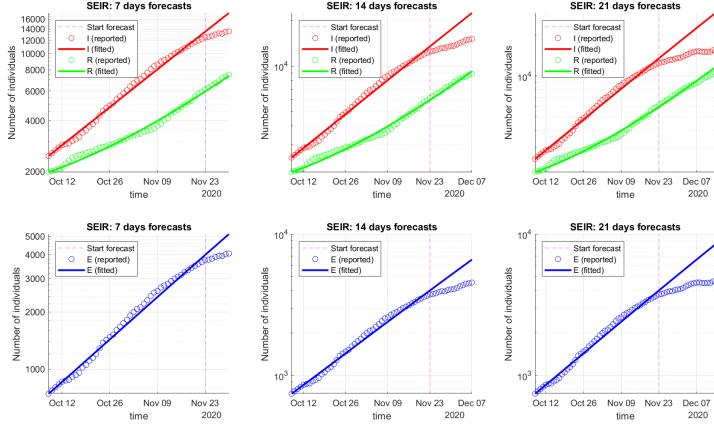
SEIIR model fitting: third wave (January 21st - February 11th 2021)

5.1.3 Sardegna



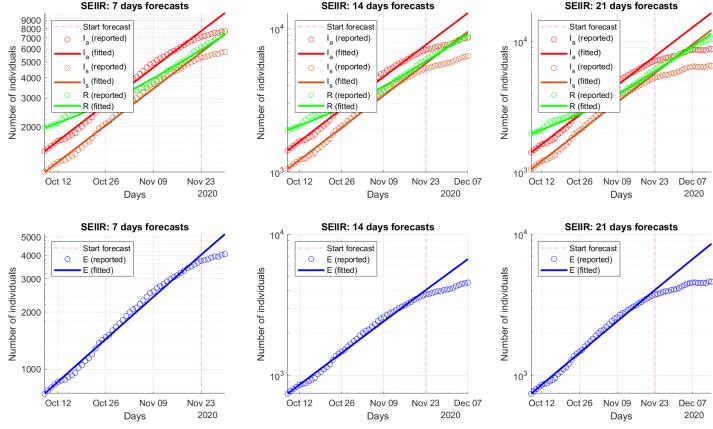
Parameter	Value
β	0.0503
γ	0.0132

SIR model fitting: second wave (October 8th - November 23rd 2020)



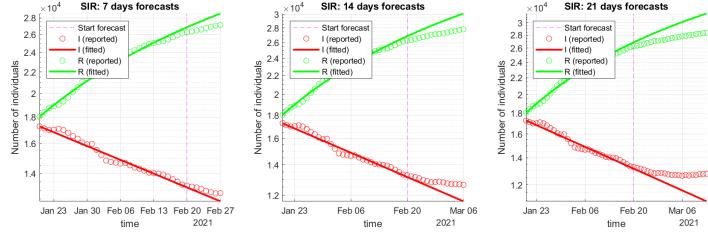
Parameter	Value
β	0.0614
γ	0.0133
τ	5.9453
α	0.1682

SEIR model fitting: second wave (October 8th - November 23rd 2020)



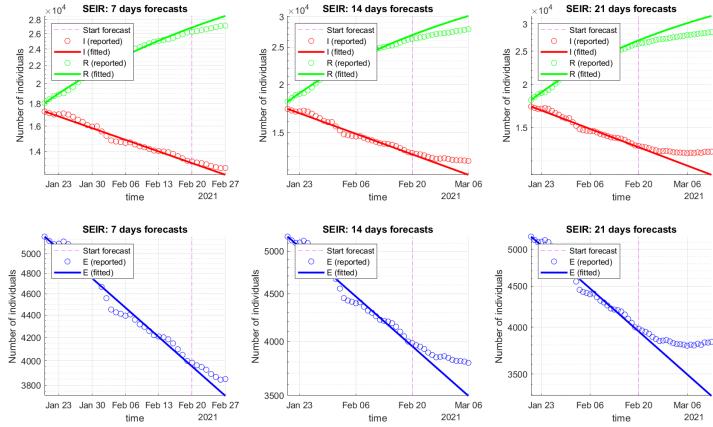
Parameter	Value
f	0.4255
β_a	0.0536
β_s	0.0715
γ	0.0132
τ	5.9952
α	0.1668

SEIIR model fitting: second wave (October 8th - November 23rd 2020)



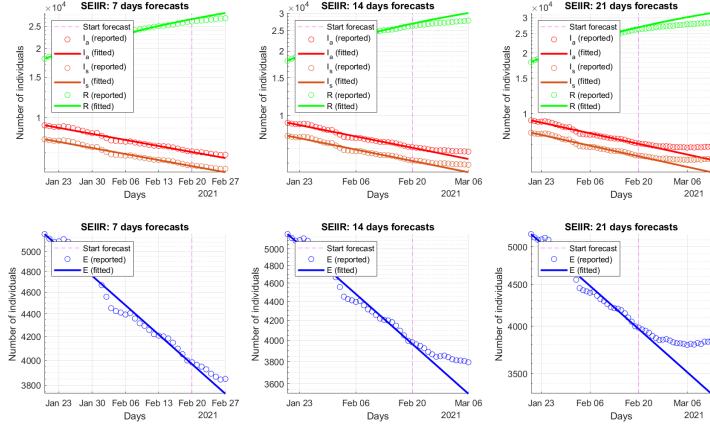
Parameter	Value
β	0.0103
γ	0.0188

SIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
β	0.0076
γ	0.0188
τ	29.85
α	0.0335

SEIR model fitting: third wave (January 21st - February 11th 2021)



Parameter	Value
f	0.4637
β_a	0.0070
β_s	0.0081
γ	0.0188
τ	29.5857
α	0.0338

SEIIR model fitting: third wave (January 21st - February 11th 2021)