



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## A SIAT<sup>3</sup>HE model of the COVID-19 pandemic in Bergamo, Italy

Marco Polver\* Fabio Previdi\* Mirko Mazzoleni\*  
Alberto Zucchi\*\*

\* *Università degli Studi di Bergamo.*

\*\* *Agenzia di Tutela della Salute (ATS) della provincia di Bergamo.*

**Abstract:** The aim of this article is to give a better understanding of the dynamics of the SARS-CoV-2 pandemic in the Bergamo province (Italy), one of the most hit areas of the world, between February and April 2020. A new compartmental model, called SIAT<sup>3</sup>HE, was designed and fitted on accurate data about the pandemic provided by ATS Bergamo, the health protection agency of the Bergamo province. Our results show that SARS-CoV-2 reached Bergamo in January and infected 318,000 people, the 28.8% of the province population. The 43.1% of the infected individuals stayed asymptomatic. As 6,028 people died due to COVID-19 till April 30th, the infection fatality ratio of SARS-CoV-2 in the Bergamo province was 1.9%. These results are in very good agreement with available information: the number of infections is consistent with the results of recent serological surveys and the number of deaths due to COVID-19 is close to the excess mortality of the considered period.

Copyright © 2021 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** Healthcare management, Nonlinear system identification, System identification and validation, Compartmental models, COVID-19

### 1. INTRODUCTION

On December 31st 2019, the Wuhan Municipal Health Commission informed the World Health Organization about a cluster of cases of pneumonia of unknown cause in Wuhan, China. A novel coronavirus, then called SARS-CoV-2, was identified in a hospitalized person in Wuhan and rapidly spread outside China, causing a worldwide pandemic.

The first COVID-19 case in Italy was detected in Codogno, on February 20th 2020. However, as explained in [Cereda et al., 2020], it's likely that SARS-CoV-2 reached Italy already in January 2020. The first SARS-CoV-2 infections in the Bergamo province were detected on February 23rd.

The evolution of the pandemic in Bergamo was rapid and caused both the impossibility to effectively track new infections and the saturation of the healthcare system, in particular during the month of March 2020. According to [Perico et al., 2020], around 420,000 citizens of the province, which counts approximately 1.1 million inhabitants, might have been infected, meaning that the 96% of the infections were not tracked, as only 16,000 were reported in the province as of September 25th 2020.

Since only people who died after testing positive to SARS-CoV-2 were considered by Italian authorities, further uncertainties regarded the real number of deaths due to COVID-19 in the Bergamo province. However, data published by the Italian national institute of statistics (ISTAT) about the mortality in Italian municipalities between 2015 and 2020 [ISTAT, 2020] were exploited in [Buonanno et al., 2020] to demonstrate that Bergamo showed a 285% increase in mortality in March 2020 compared to March 2019, implying that the excess mortality in 2020 was twice the number of deaths attributed to

COVID-19. Based on the last finding, it is clear that the real numbers of the SARS-CoV-2 pandemic in Bergamo are not completely known. Thus, a reliable model that helps their estimation is highly envisaged.

Many epidemiological models have been proposed to describe the dynamics of the SARS-CoV-2 pandemic in Italy and most of them are based on the classical SIR and SEIR models [Hethcote, 2000], e.g. [Casella, 2020] and [Calafiore et al., 2020]. In particular, [Giordano et al., 2020] proposes a model, called SIDARTHE, that takes into account the severity of illness (SOI) and distinguishes between detected and undetected infected subjects. However, as the SIDARTHE model aims to show the effectiveness of different control policies, it could not be used directly to carry out a retrospective analysis of the pandemic.

In this paper we introduce a modified version of the SIDARTHE model, called SIAT<sup>3</sup>HE, that was designed to be consistent with the characteristics of the SARS-CoV-2 pandemic in the Bergamo province. The model aims to reliably estimate the real numbers of the SARS-CoV-2 pandemic in Bergamo between February and April 2020. The design and estimation of that model was made possible by the data provided by ATS Bergamo, the health protection agency of the province.

The SIAT<sup>3</sup>HE model takes into account the SOI like the SIDARTHE model, but it focuses in particular on severe cases, distinguishing between subjects that required hospitalization, subjects that required admission to intensive care and individuals that needed either hospitalization or admission to intensive care but had to stay at home due to the saturation of the healthcare system. In order to make model identification possible and effective, the noisiest time-

series were filtered by means of a second order low-pass Butterworth filter. Finally, model identification was carried out by minimization of the mean squared error between available data and the estimated model compartments values.

The identified model was exploited to estimate the evolution of the pandemic. According to the obtained results, SARS-CoV-2 seems to have reached Bergamo in January 2020, in agreement with [Cereda et al., 2020]. Approximately 318,000 people, the 28.8% of the population, got infected. This result is consistent with the findings of recent serological tests, the 29.75% of which gave a positive result. Moreover, this confirms that the 96% of the infections were not tracked.

Among those who were infected, the 43.1% stayed asymptomatic, in agreement with [Lavezzo et al., 2020]. The infection fatality ratio was 1.9%, as 6,028 people died due to COVID-19.

The remainder of the paper is organized as follows. Section II presents the data that were collected, reconstructed and exploited in the model identification stage. Section III introduces and analyzes the SIAT<sup>3</sup>HE model. In Section IV, the results of this work are discussed. Finally, Section V draws the conclusions.

## 2. DATASET RETRIEVEMENT AND DESCRIPTION

The SIAT<sup>3</sup>HE model was designed to be effectively fitted on the available data about the dynamics of the SARS-CoV-2 pandemic in the Bergamo province. In particular, five time-series were provided by ATS Bergamo, while other three were reconstructed.

Name	Description
$\tilde{T}_n$	Daily number of people undergoing oxygen therapy at home.
$\tilde{T}_{d1,in}$	Daily number of new hospitalizations of people who tested positive to SARS-CoV-2.
$\tilde{T}_{d1}$	Daily number of hospitalized people who tested positive to SARS-CoV-2.
$\tilde{T}_{d2,in}$	Daily number of new intensive care admissions of people who tested positive to SARS-CoV-2.
$\tilde{T}_{d2}$	Daily number of intensive care patients who tested positive to SARS-CoV-2.
$\tilde{I}_{in}$	Estimated daily number of people getting infected among those who were then hospitalized.
$\tilde{A}_{in}$	Estimated daily number of people showing first COVID-19 symptoms among those who were then hospitalized.
$\tilde{E}_{in}$	Estimated daily number of deaths due to COVID-19.

Table 1. Available time-series

Table 1 shows the time-series that were exploited to estimate the SIAT<sup>3</sup>HE model parameters. All the time-series are related to the period between February 20th and April 30th 2020 and regard only people from the Bergamo province. The first five time-series were directly provided by ATS Bergamo, while the last three were reconstructed as explained in the following two subsections.

### 2.1 Reconstruction of $\tilde{I}_{in}$ and $\tilde{A}_{in}$

The Papa Giovanni XXIII hospital in Bergamo collected data about 123 COVID-19 patients in intensive care. One of the collected information was the time interval between the onset of symptoms declared by the patients themselves and their hospital admission. A Gamma distribution was fitted to them; the result is shown in Figure 1.

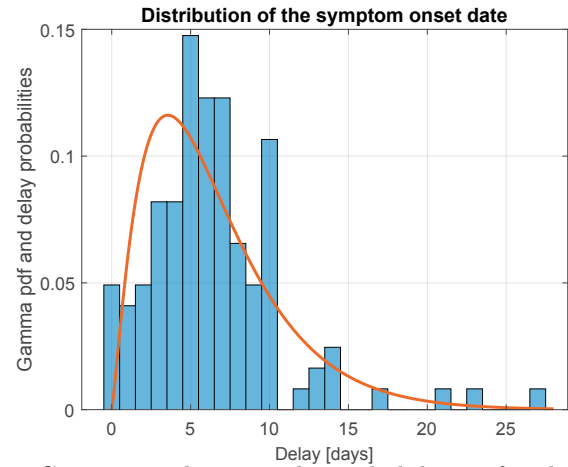


Fig. 1. Comparison between the probabilities of each time interval between symptom onset and hospitalization and the obtained Gamma distribution.

The obtained Gamma distribution was then used to estimate the daily number of people showing symptoms for the first time among those who were hospitalized later. For each subject that was hospitalized due to COVID-19, a random delay between symptom onset and hospital admission was generated from the obtained Gamma distribution and used to obtain an estimation of the symptom onset day. By doing so, the daily number of subjects showing symptoms for the first time among those who were hospitalized later was obtained. The exact reconstruction process is explained in Algorithm 1.

---

#### Algorithm 1: Reconstruction of the symptom onset day in hospitalized individuals

---

**Data:**  $\tilde{T}_{d1,in}[20/02/2020, \dots, 30/04/2020]$ , distribution of the delay between symptom onset and hospitalization SymptomOnsetDistribution

**Result:**  $\tilde{A}_{in}[20/01/2020, \dots, 30/04/2020]$  (20/01/2020 was chosen for convenience)

```

 $\tilde{A}_{in} \leftarrow$  array of zeros with index  $\in [20/01/2020, \dots, 30/04/2020]$ ;
foreach  $day\ t \in [20/02/2020, \dots, 30/04/2020]$  do
    admissions  $\leftarrow \tilde{T}_{d1,in}[t]$ ;
    for  $i = 0$  to admissions do
        delay  $\leftarrow$  random sample from SymptomOnsetDistribution;
         $\tilde{A}_{in}[t - \text{delay}] += 1$ ;

```

**return**  $\tilde{A}_{in}$

---

A similar algorithm was used to estimate the day of infection of each hospitalized individual from the Bergamo province ( $\tilde{I}_{in}$ ) from  $\tilde{A}_{in}$ . The delay between infection and symptom onset was randomly generated from a Weibull

distribution proposed in [Backer et al., 2020], which was fitted on data about 88 travelers who got infected in Wuhan.

### 2.2 Reconstruction of $\tilde{E}_{in}$

The problems related to the inaccuracy of the official data about people who died after testing positive to SARS-CoV-2 in Italy were overcome by exploiting the data published by ISTAT about the mortality in each Italian municipality from 2015 to 2020 [ISTAT, 2020], which led to a more accurate estimation of the daily number of deaths due to COVID-19 in the Bergamo province.

The main assumptions behind the calculation of the daily number of deaths due to COVID-19 were the following:

- The average mortality in the Bergamo province was considered constant between 2015 and February 2020.
- The entire unexpected mortality between February and April 2020 was attributed to COVID-19.

The daily unexpected mortality in the Bergamo province from February to April 2020 was calculated as follows:

$$U(t) = M_{2020}(t) - \frac{1}{4} \sum_{i=2015}^{2019} M_i(t), \quad (1)$$

$$t \in [20/02, \dots, 30/04]$$

where:

- $U(t)$  is the unexpected number of deaths on day  $t$  in year 2020.
- $M_i(t)$  is the number of deaths on day  $t$  in year  $i$ .

A comparison between the daily number of deaths of people who tested positive to SARS-CoV-2 and the daily number of excess deaths in the Bergamo province is shown in Figure 2.

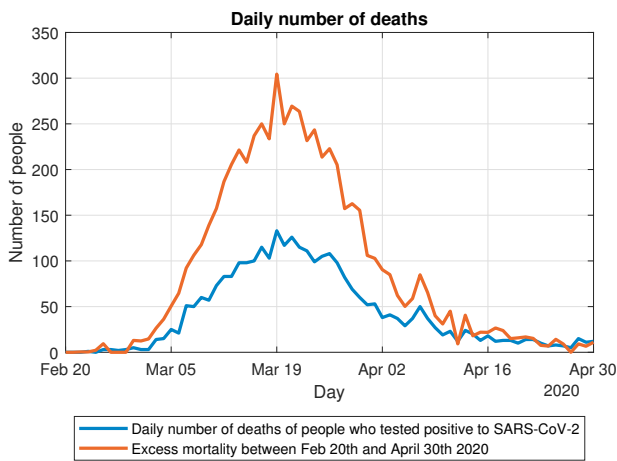


Fig. 2. Comparison between the daily number of deaths of people who tested positive to SARS-CoV-2 (blue) and the daily number of excess deaths (red).

### 3. SIAT<sup>3</sup>HE MODEL

The SIAT<sup>3</sup>HE model is composed of eight compartments:

- (1) *Susceptible (S)*: people who have never been infected by the virus.
- (2) *Infectious (I)*: infectious individuals without symptoms. They can be either asymptomatic, paucisymptomatic or presymptomatic subjects.

- (3) *Ailing (A)*: infectious individuals with mild COVID-19 symptoms (e.g. cough, fever, loss of taste and smell).
- (4) *Threatened non-diagnosed (T<sub>n</sub>)*: infected subjects with severe COVID-19 symptoms that are not hospitalized due to the saturation of the healthcare system.
- (5) *Threatened diagnosed 1 (T<sub>d1</sub>)*: infected subjects with severe COVID-19 symptoms that are hospitalized.
- (6) *Threatened diagnosed 2 (T<sub>d2</sub>)*: infected subjects with extremely severe COVID-19 symptoms and that are hospitalized in intensive care.
- (7) *Healed (H)*: previously infected individuals who got virus-negative.
- (8) *Extinct (E)*: people who died due to COVID-19.

The SIAT<sup>3</sup>HE dynamical system consists of the following ordinary differential equations:

$$\begin{cases} \dot{S}(t) = -S(t)(\beta_1 I(t) + \beta_2 A(t)) \\ \dot{I}(t) = S(t)(\beta_1 I(t) + \beta_2 A(t)) - (\delta + \gamma_1)I(t) \\ \dot{A}(t) = \delta I(t) - (\gamma_2 + \epsilon_1 + \epsilon_2)A(t) \\ \dot{T}_n(t) = \epsilon_1 A(t) - (\gamma_3 + \nu_1)T_n(t) \\ \dot{T}_{d1}(t) = \epsilon_2 A(t) - (\gamma_4 + \epsilon_3 + \nu_2)T_{d1}(t) \\ \dot{T}_{d2}(t) = \epsilon_3 T_{d1}(t) - (\gamma_5 + \nu_3)T_{d2}(t) \\ \dot{H}(t) = \gamma_1 I(t) + \gamma_2 A(t) + \gamma_3 T_n(t) + \gamma_4 T_{d1}(t) + \gamma_5 T_{d2}(t) \\ \dot{E}(t) = \nu_1 T_n(t) + \nu_2 T_{d1}(t) + \nu_3 T_{d2}(t) \end{cases} \quad (2)$$

The uppercase letters in (2) represent the state variables of the system (the eight compartments), while the Greek letters represent the parameters of the system, which are all considered positive. An explanation of the parameters of the system is provided below:

- $\beta_1$  and  $\beta_2$  are the contact rates of the compartments  $I$  and  $A$  respectively. The risk of contagion due to subjects with life-threatening symptoms (compartments  $T_n$ ,  $T_{d1}$  and  $T_{d2}$ ) is considered negligible, as these individuals are assumed to be well isolated. Another assumption that is made is that  $\beta_1 \geq \beta_2$ , as people with COVID-19 symptoms are expected to have fewer contacts compared to people without symptoms.
- $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$ ,  $\gamma_4$  and  $\gamma_5$  denote the recovery rates of the five compartments related to infected subjects. *Recovery rate* is here used to define the time it takes for an infected individual to get virus-negative.
- $\delta$  represents the probability rate at which infectious individuals without symptoms develop mild symptoms.
- $\epsilon_1$  and  $\epsilon_2$  denote the rates at which infectious individuals with mild symptoms develop life-threatening symptoms and respectively stay at home or are hospitalized.
- $\epsilon_3$  represents the rate at which hospitalized individuals are admitted to intensive care.
- $\nu_1$ ,  $\nu_2$  and  $\nu_3$  denote the fatality rates for infected individuals with life-threatening symptoms in compartments  $T_n$ ,  $T_{d1}$  and  $T_{d2}$ . An assumption that is made is that  $\nu_2 \leq \nu_3$ .

#### 3.1 Modeling choices

The choices and the assumptions at the basis of the design of the SIAT<sup>3</sup>HE model are explained below:

- Reinfection of recovered subjects is considered impossible. Since the time period considered while estimating the model parameters (February–April 2020) is rela-

tively short, temporary immunity was likely to be still in place at the end of April in individuals who got infected in that period.

- Temporary immunity is only provided by a previous infection, as no vaccines against SARS-CoV-2 were available between February and April 2020.
- The latency period between exposure to the virus and onset of infectiousness is assumed to be null, as there's evidence that SARS-CoV-2 can be transmitted also during the incubation period [Qian et al., 2020].
- Even if there's the possibility that a certain number of non-hospitalized individuals with life-threatening symptoms were hospitalized later or directly admitted to intensive care, there are no connections between the compartment  $T_n$  and the compartments  $T_{d1}$  and  $T_{d2}$ .
- The SIAT<sup>3</sup>HE model only captures the average effect of the observed phenomenon. The population is considered well-mixed and differences between the evolution of the infection at different ages are ignored.

The SIAT<sup>3</sup>HE model was designed to be consistent with the situation of Bergamo between February and April 2020. For this reason, it is suitable for the representation of the dynamics of the SARS-CoV-2 pandemic in areas characterized by high numbers of infected people and a clear saturation of the healthcare system, such as Milan and Brescia in Italy and Madrid in Spain.

### 3.2 Mathematical analysis

The SIAT<sup>3</sup>HE model (2) is a bilinear system described by eight differential equations. If all the state variables are initialized with non-negative values at time 0, their values remain non-negative for all  $t > 0$ , thus making the system positive. The system demonstrates the mass conservation property  $\dot{S}(t) + \dot{I}(t) + \dot{A}(t) + \dot{T}_n(t) + \dot{T}_{d1}(t) + \dot{T}_{d2}(t) + \dot{H}(t) + \dot{E}(t) = 0$ , which makes the sum of the states constant:

$$S(t) + I(t) + A(t) + T_n(t) + T_{d1}(t) + T_{d2}(t) + H(t) + E(t) = 1 \quad (3)$$

The values of the compartments sum up to 1 because 1 denotes the total population and each compartment contains a fraction of the entire population.

The overall system can be divided into three subsystems:

- (1) a subsystem containing the non-linear part of (2), which will be called *non-linear subsystem* from now on;
- (2) a linear subsystem containing the equations related to the compartments  $A$ ,  $T_n$ ,  $T_{d1}$  and  $T_{d2}$ , which will be called *AT<sup>3</sup> subsystem* from now on;
- (3) a linear subsystem that describes the evolution of the compartments  $H$  and  $E$ , two cumulative variables that depend only on the other variables.

The newly defined subsystems are characterized by the following three state vectors:

$$\begin{aligned} x_1 &= [S \ I]^\top \\ x_2 &= [A \ T_n \ T_{d1} \ T_{d2}]^\top \\ x_3 &= [H \ E]^\top \end{aligned} \quad (4)$$

Defining the input of the non-linear subsystem as  $u_1(t) = x_{2,1}(t)$ , the non-linear subsystem can be rewritten as follows:

$$\begin{cases} \dot{x}_{1,1}(t) = -\beta_1 x_{1,1}(t)x_{1,2}(t) - \beta_2 x_{1,1}(t)u_1(t) \\ \dot{x}_{1,2}(t) = \beta_1 x_{1,1}(t)x_{1,2}(t) + \beta_2 x_{1,1}(t)u_1(t) \\ \quad - (\gamma_1 + \delta)x_{1,2}(t) \\ y_1(t) = x_{1,2}(t) \end{cases} \quad (5)$$

The AT<sup>3</sup> subsystem can be defined as follows:

$$u_2(t) = y_1(t) = x_{1,2}(t) \quad (6)$$

$$\begin{aligned} \dot{x}_2(t) &= Fx_2(t) + B_2u_2(t) \\ &= \begin{bmatrix} -r_1 & 0 & 0 & 0 \\ \epsilon_1 & -r_2 & 0 & 0 \\ \epsilon_2 & 0 & -r_3 & 0 \\ 0 & 0 & \epsilon_3 & -r_4 \end{bmatrix} x_2(t) + \begin{bmatrix} \delta \\ 0 \\ 0 \\ 0 \end{bmatrix} u_2(t) \end{aligned} \quad (7)$$

$$y_2(t) = C_2x_2(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_2(t) = \begin{bmatrix} x_{2,1}(t) \\ x_{2,2}(t) \\ x_{2,3}(t) \\ x_{2,4}(t) \end{bmatrix} \quad (8)$$

where  $r_1 = \gamma_2 + \epsilon_1 + \epsilon_2$ ,  $r_2 = \gamma_3 + \nu_1$ ,  $r_3 = \gamma_4 + \epsilon_3 + \nu_2$  and  $r_4 = \gamma_5 + \nu_3$ .

The last subsystem is instead defined by the following equations:

$$u_3(t) = y_2(t) = x_2(t) \quad (9)$$

$$\begin{aligned} \dot{x}_3(t) &= b_{3,1}u_2(t) + B_{3,2}u_3(t) \\ &= \gamma_1 u_2(t) + [\gamma_2 \ \gamma_3 \ \gamma_4 \ \gamma_5] u_3(t) \end{aligned} \quad (10)$$

$$y_3(t) = C_3x_3(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x_3(t) = \begin{bmatrix} x_{3,1}(t) \\ x_{3,2}(t) \end{bmatrix} \quad (11)$$

where  $u_2$  is defined in (6).

The block diagram of the SIAT<sup>3</sup>HE model is shown in Figure 3.

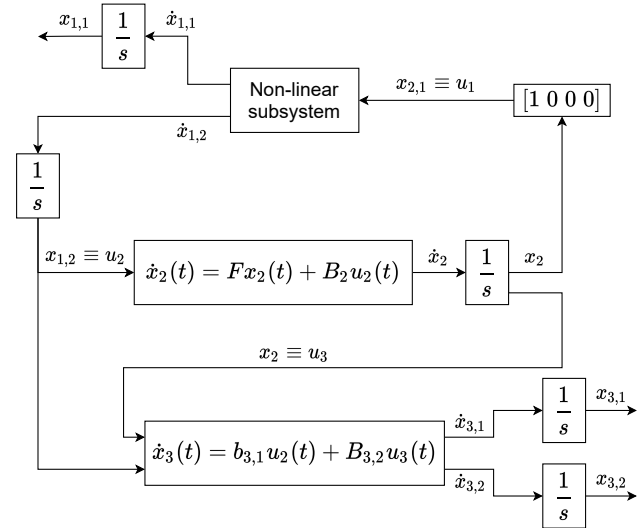


Fig. 3. Block diagram of the SIAT<sup>3</sup>HE model.

### 3.3 System equilibria

We now demonstrate that all the equilibria that can be reached by the system are of the type:

$$\bar{S} \geq 0, \bar{I} = 0, \bar{A} = 0, \bar{T}_n = 0, \bar{T}_{d1} = 0, \bar{T}_{d2} = 0, \bar{H} \geq 0, \bar{E} \geq 0 \quad (12)$$

given non-negative initial values  $S(0)$ ,  $I(0)$ ,  $A(0)$ ,  $T_n(0)$ ,  $T_{d1}(0)$ ,  $T_{d2}(0)$ ,  $H(0)$ ,  $E(0)$  that are consistent with (3).

When an equilibrium is reached, the left-hand side of all the ODEs in system (2) become 0. By summing the first two equations of the newly obtained system, we get

$-(\delta + \gamma_1)\bar{I} = 0$ . Since all the model parameters are positive, this leads to:

$$\bar{I} = 0 \tag{13}$$

Taking (13) into account, the third equation becomes  $-(\gamma_2 + \epsilon_1 + \epsilon_2)\bar{A} = 0$ , which leads to  $\bar{A} = 0$ . The consequence of this is that also  $\bar{T}_n = \bar{T}_{d1} = \bar{T}_{d2} = 0$ .

Different considerations are made about  $\bar{S}$ ,  $\bar{H}$  and  $\bar{E}$ . Considering a generic non-trivial initial condition that includes infectious individuals  $S(0) > 0, I(0) > 0, A(0) = T_n(0) = T_{d1}(0) = H(0) = E(0) = 0$ ,  $H(t)$  and  $E(t)$  increase over time, while  $S(t)$  decreases without reaching negative values. This leads to:  $\bar{S} \geq 0, \bar{H} \geq 0, \bar{E} \geq 0$ , with  $\bar{S} + \bar{H} + \bar{E} = 1$  to demonstrate the mass conservation property (3).

The meaning of the last finding is that the SIAT<sup>3</sup>HE model reaches an equilibrium only when there are no more infected and infectious individuals.

### 3.4 Parameters estimation

Model identification was carried out by bounded and constrained minimization of the MSE between available data and compartments estimations.

Table 1 shows the exploited time-series and names them consistently with the compartment each time-series is related to. In particular:

- $\tilde{I}_{in}$  is proportional to the daily number of people entering compartment  $I$  (from now on called  $I_{in}$ ).
- $\tilde{A}_{in}$  is proportional to the daily number of people entering compartment  $A$  (from now on called  $A_{in}$ ).
- $\tilde{T}_n$  is proportional to the daily number of people in compartment  $T_n$ .
- $\tilde{T}_{d1,in}$  is an accurate measure of the daily number of people entering compartment  $T_{d1}$  (from now on called  $T_{d1,in}$ ).
- $\tilde{T}_{d1}$  is an accurate measure of the daily number of people in compartment  $T_{d1}$ .
- $\tilde{T}_{d2,in}$  is an accurate measure of the daily number of people entering compartment  $T_{d2}$  (from now on called  $T_{d2,in}$ ).
- $\tilde{T}_{d2}$  is an accurate measure of the daily number of people in compartment  $T_{d2}$ .
- $\tilde{E}_{in}$  is a measure of the daily number of people entering compartment  $E$  (from now on called  $E_{in}$ ).

As the available time-series had different accuracies, they were given different weights during the calculation of the MSE:

- $I_{in}$  and  $A_{in}$  were given a weight of 0.25 as  $\tilde{I}_{in}$  and  $\tilde{A}_{in}$  were reconstructed from other data.
- $T_n$  was given a weight of 0.5 because  $\tilde{T}_n$  was the noisiest time-series in the dataset.
- The other variables were given a weight of 1 as the related time-series were considered sufficiently accurate.

Together with the model parameters defined before, also the day of the first infection in the Bergamo province, called  $t_0$ , was estimated. In particular,  $t_0$  was rigorously defined as the day when the SIAT<sup>3</sup>HE model's state vector was  $\left[1 - \frac{1}{N} \quad \frac{1}{N} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\right]^T$ , where  $N$  is the total population of the province.

Since the dynamics of the pandemic in Bergamo changed over time due to the different measures taken by the Italian government, the model parameters were estimated along three different time intervals:

- 24/01/2020 - 09/03/2020: no lockdown was imposed.
- 10/03/2020 - 25/03/2020: a lockdown was imposed, but many work activities were still open.
- 26/03/2020 - 30/04/2020: hard lockdown; only strategic and necessary work activities were open.

January 24th is the first day included in the first time interval as it's the estimated value of  $t_0$ .

## 4. RESULTS AND DISCUSSION

The parameter sets that were found in model identification were used to estimate the dynamics of the SARS-CoV-2 pandemic in Bergamo between January 24th and April 30th 2020. The fit between the estimated model and the available data is good, as revealed by the NRMSE values in Table 2. The quality of the fit can also be seen in Figure 4, where the evolution of each compartment is shown.

Compartment	NRMSE
$T_n$	62.884
$T_{d1}$	6.018
$T_{d2}$	0.436
$E$	35.159

Table 2. NRMSE of each compartment. Its value was calculated as  $NRMSE = \frac{RMSE}{\bar{y}}$ , where  $\bar{y}$  is the average value of the observed variable  $y$ .

As the time-series  $\tilde{T}_n$  was the noisiest in the dataset, the compartment  $T_n$  was expected to have the worst fit, but the obtained result is still acceptable.

Data provided by ATS Bergamo show that 2,973 people from the Bergamo province died after testing positive to SARS-CoV-2 between February 20th and April 30th. The calculated excess deaths related to the same period are instead 6,175, which is justified by the inability to test all people showing COVID-19 symptoms in that period. The estimated final value of the compartment  $E$  is 6,028, which is close to the unexpected mortality.

The estimated day of the first infections in the Bergamo province is January 24th 2020, in agreement with [Cereda et al., 2020], which proves that SARS-CoV-2 was already spreading in Lombardy in January 2020.

The estimated total number of infections till April 30th is approximately 318,000, the 28.8% of the population of the province. This finding is consistent with the results of recent serological surveys. In particular, ATS Bergamo collected data about all the serological tests that were executed in the Bergamo province: the 29.75% of the 130,704 tests that were executed till August 12th 2020 gave a positive result. The total number of infections also allows to estimate the real infection fatality ratio (IFR) of SARS-CoV-2 in the Bergamo province, which was 1.9%.

As of April 30th 2020, only 11,313 SARS-CoV-2 infections were detected in the Bergamo province [Dipartimento della Protezione Civile, 2020]. This means that the 96% of the infections were not tracked, in agreement with the results obtained by [Perico et al., 2020].

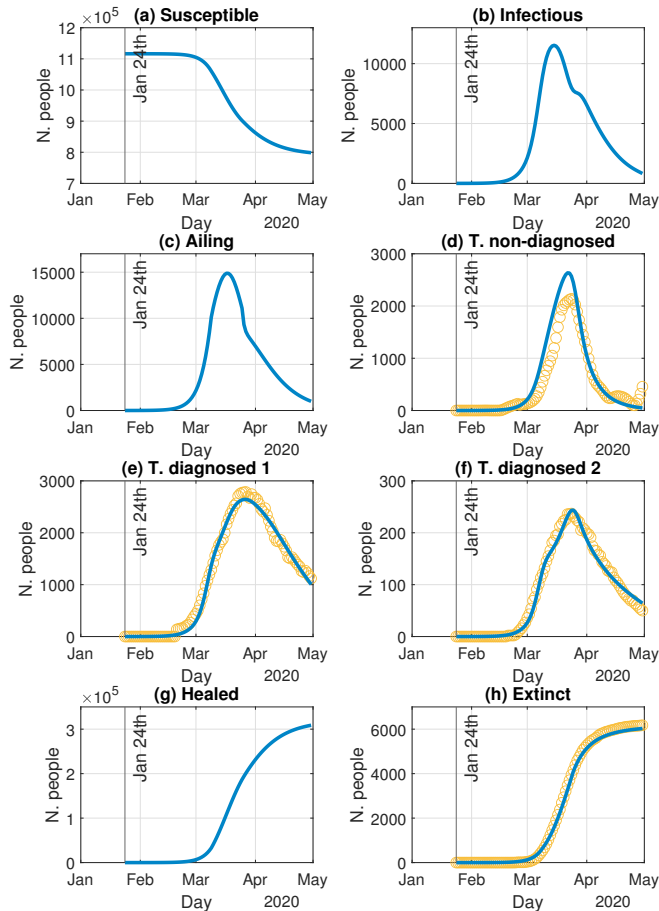


Fig. 4. Evolution of the model compartments: (a) Susceptible — (b) Infectious — (c) Ailing — (d) Threatened non-diagnosed: comparison between  $\kappa T_n$  (yellow) and estimated  $T_n$  values (blue) — (e) Threatened diagnosed 1: comparison between real values (yellow) and estimated values (blue) — (f) Threatened diagnosed 2: comparison between real values (yellow) and estimated values (blue) — (g) Healed — (h) Extinct: comparison between excess deaths (yellow) and estimated deaths due to COVID-19 (blue).

Also the number of symptomatic and asymptomatic cases was estimated by means of the identified SIAT<sup>3</sup>HE model. Approximately 181,000 people showed COVID-19 symptoms. This means that the 56.9% of the subjects who got infected also developed symptoms, while the 43.1% stayed asymptomatic. This estimation is consistent with the percentage of asymptomatic cases that was detected during the testing of the whole municipality of Vo' (Italy) [Lavezzo et al., 2020], which was the 42.5%.

During the second half of March, when around 2,600 subjects were hospitalized with COVID-19 symptoms and 240 patients were undergoing intensive care, the number of infected people with severe symptoms at home reached a maximum of 2,635. This means that almost a half of the individuals showing life-threatening COVID-19 symptoms couldn't be hospitalized during the worst period of the pandemic in Bergamo due to the saturation of the healthcare system.

## 5. CONCLUSIONS

While serological surveys give important information about the total number of past infections in a certain area, the

value of the SIAT<sup>3</sup>HE model lies in its ability to explicit how the numbers of the SARS-CoV-2 pandemic evolved over time.

Our estimations outline a profile of the SARS-CoV-2 pandemic in the Bergamo province that is totally different from the profile that could be outlined in March-April 2020. Considering only official data, SARS-CoV-2 reached Bergamo in February 2020 and rapidly led to the saturation of the province hospitals, with an IFR equal to the 26.2%. According to the SIAT<sup>3</sup>HE model here introduced, SARS-CoV-2 reached Bergamo in January, meaning that the increase in contagions was not as steep as imagined. Hospitals and intensive care departments started to suffer from the effects of the pandemic more than one month later. This allows to conclude that the number of hospitalizations and deaths can be limited by effectively detecting as many infected people as possible, even when they don't show symptoms of any kind.

The IFR was the 1.9%, much lower than the previously found 26.2%. This finding is crucial, because it allows to understand that the high number of deaths in the Bergamo province was mainly due to the high number of infections, most of which were not detected. This means that, even though the entire world would benefit from the availability of both an effective cure to COVID-19 and a vaccine, the latter can be considered more urgent than the former.

## REFERENCES

- Backer, J. et al. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*, 25.
- Buonanno, P., Galletta, S., and Puca, M. (2020). Estimating the severity of COVID-19: Evidence from the Italian epicenter. *Plos one*, 15(10), e0239569.
- Calafiore, G.C. et al. (2020). A modified SIR model for the COVID-19 contagion in Italy. *arXiv preprint arXiv:2003.14391*.
- Casella, F. (2020). Can the COVID-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5(3), 1079–1084.
- Cereda, D. et al. (2020). The early phase of the COVID-19 outbreak in Lombardy, Italy.
- Dipartimento della Protezione Civile (2020). Dati COVID-19 Italia.
- Giordano, G. et al. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6.
- Hethcote, H.W. (2000). The mathematics of infectious diseases. *SIAM review*, 42(4), 599–653.
- ISTAT (2020). Decessi anni 2015-2020.
- Lavezzo, E. et al. (2020). Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*, 584(7821), 425–429.
- Perico, L. et al. (2020). COVID-19 and Lombardy: Testing the impact of the first wave of the pandemic. *EBioMedicine*, 61, 103069.
- Qian, G. et al. (2020). COVID-19 Transmission Within a Family Cluster by Presymptomatic Carriers in China. *Clinical Infectious Diseases*, 71(15), 861–862.