

# Data driven characterization of COVID-19

Paolo Di Giamberardino

*Dept. Computer, Control and  
Management Engineering*

*Sapienza University of Rome, Italy*  
paolo.digiamberardino@uniroma1.it

Daniela Iacoviello

*Dept. Computer, Control and  
Management Engineering*

*Sapienza University of Rome, Italy*  
daniela.iacoviello@uniroma1.it

**Abstract**—COVID-19 has caused more than 880.000 victims all over the world (September 2020); despite a large effort of the scientific community and of the governments, it is still a great problem, inducing most of the Nations to adopt restriction to mobility, social relations and economic activities. Since the beginning of the pandemic, COVID-19 appeared a rather mysterious virus, for which neither a vaccination nor specific medications exist. In this paper, COVID-19 is characterized by using the available data of total number of infected, healed and dead people to identify the contagion, the removed and the death rates. These values depend on various aspects related to the population characteristics, the general health conditions, the social and economical situations, as well as to other features not yet identified by the scientific community. The COVID-19 situation in Italy is herein explored, showing the great heterogeneity of the virus spread in different zones.

## I. INTRODUCTION

In the last few months the COVID-19 spread has involved the entire world forcing almost all the humanity to a significant reduction of activities; up to now (September 2020) there has been more than 880.000 victims [1], and a not defined number of infected people. In fact, a large quantity of infected people is asymptomatic not aware of their status being, nevertheless, able to infect. After some months from the official start of the pandemic, declared on March 11 by the World Health Organization (WHO), many characteristics of COVID-19 are still unknown; in particular, it is not clear why this virus spreads more in one region than in another, even if neighbouring. Some main characteristics, as the mean age, the density of the population, and the general health condition are considered important to define the level of susceptibility of a population, [1]. It has been argued also a role of environmental conditions, such as pollution, of vitamin D, of smoking, of gender, but no evidence has been established, [1], [2].

Despite the COVID-19 emergency is relatively recent, literature is vast and covers various aspects, [3], [4], [5], [6], [7], [8]. The basic model used is the SEIR one in which the susceptible (S), exposed (E), infected (I) and removed (R) subjects are considered, [8]; a variation of the classical SIR model is the SIRD one in which the category of dead people  $D$  is introduced, [9]. This class is particularly important for this emergency, since the only real data actually referable to the COVID-19 are those related to the infected (I) and deceased

(D) patients. As said, some important categories, such as the asymptomatic subjects, can only be estimated or deduced by suitable modelling such as in [3], [4]. In the first weeks of the pandemic an estimation of the percentage of asymptomatic individuals was obtained by considering the possible extension of the results of tests done on the entire population of small countries; this idea has been used in [10], where the information on the percentage of asymptomatic subjects in Vo' Euganei (Veneto) helped in identifying parameters of a complex model that considered also quarantined patients.

In this paper the aim is to determine information on the COVID-19 spread in Italy, such as the contact, the recovery and the death rates, considering only the available data from Protezione Civile [11]: in particular, the official numbers of infected patients, of healed individuals and dead people are used. By analysing these data, the not-homogeneous distribution of these quantities is evident, even if, especially from the beginning of March, the containment measures were the same in the entire Nation. A model simple for purpose is analysed by using the above indicated data of representative regions of Italy; some model parameters, the contagion, the recovery and the death rates are identified directly from the data. These parameters are influenced by virus characteristics, at the moment unknown, and by the peculiarities of the population, its social habits, the ambient conditions; these latter data can be deduced from the ISTAT website, [12]. It is still not evident if and how much these characteristics have influenced the COVID-19 spread; moreover, as it will be confirmed in the paper, other elements, not evident up to now, could explain what happened in Italy and probably guide the choices of containment measures.

The paper is organized as follows. Section 2 is divided into four subsections in which a simple but effective model is proposed and analysed; the model parameters regarding the contagion rate, the recovery rate and the death rate are identified by minimizing a suitable cost index. In Section 3, numerical results show the effectiveness of the simplified model proposed. Conclusions and future work are outlined in Section 4.

## II. MATERIALS AND METHODS

In this section a simple model is introduced and motivated; its parameters are deeply discussed, enhancing the possible

influences of social, economic, healthy aspects of the population.

### A. The mathematical model

As discussed in the Introduction, many interesting mathematical models have been proposed to describe specific aspects of the COVID-19 spread. The aim of this paper is to identify some important parameters specific of this virus, by using only directly measured data. For example, it is known the presence of asymptomatic subjects; nevertheless, it is still controversial their infection capability and, most of all, their realistic percentage. Up to now, without a screening on an entire population, this data must be considered unavailable. The unique established data regard the number of infected patients, of individuals recovered from the infection and of dead people. In the model herein proposed and identified, the following classes are considered:

- the class  $S$  of susceptible individuals: healthy people;
- the class  $E$  of exposed subjects: people infected and not infectious;
- the class  $I$  of infected patients: people infected and infectious, requiring medical assistance;
- the class  $R$  of removed individuals: people healed from the COVID-19 infection;
- the class  $D$  of dead patients: people died for COVID-19.

The epidemic spread depends, in general, on the peculiarities of the virus, on the general health, social and economic condition of the population, on the pollution and on the weather conditions, just to mention the most important ones that *probably* strongly influence the spread.

In all the epidemiological models, it is possible to distinguish among parameters typical of the virus evolution, those related to the specific population, and those that can be changed by suitable control actions. For the purposes of the proposed analysis, the following constant parameters are assumed known, because deduced from the characteristics of the specific population or estimated by the physicians:

- $b$ : the rate of new incomers;
- $d$ : the death rate not influenced by the virus presence;
- $\alpha$ : the incubation time.

The following parameters are, instead, strictly related to specific conditions that may strongly affect the evolution of the spread and must be estimated from data:

- $\beta$ : the contact rate;
- $d_I$ : the death rate of infected individuals;
- $\gamma$ : the recovery rate.

Note that they are not fixed, depending also on the varying containment measures. The model used to identify the evolution of these three important parameters is the following

SEIRD description:

$$\dot{S}(t) = b - \beta(t)S(t)I(t) - dS \quad (1)$$

$$\dot{E}(t) = \beta(t)S(t)I(t) - dE - \alpha E(t) \quad (2)$$

$$\dot{I}(t) = \alpha E(t) - \gamma(t)I(t) - d_I(t)I(t) \quad (3)$$

$$\dot{R}(t) = -dR(t) + \gamma(t)I(t) \quad (4)$$

$$\dot{D}(t) = d_I(t)I(t) \quad (5)$$

The model proposed is an extension of the classical SEIR one, with the introduction of the evolution of the number of dead patients.

### B. Discussion on the model parameters

The proposed SEIRD model is a very simplified description of COVID-19 spread, kept simple for purpose, aiming at a reliable model identification based on the available data of the official infected patients, of the healed individuals and of the dead ones. The parameters  $b$  and  $d$  are specific of the population and are related since in absence of the disease the ratio  $\frac{b}{d}$  is equal to the equilibrium value of the susceptible individuals; moreover, the death rate  $d$  is an available official data of each region. The parameter  $\alpha$  is the recovery rate and represents the rate at which an exposed subject becomes infectious; the incubation time is not unique but it has been estimated by the physicians between 3 to 10 days.

The other parameters  $\beta$ ,  $\gamma$  and  $d_I$  are the main objects of this investigation; they do not depend only on the specific virus characteristics, but also on aspects of different nature, such as the general health conditions of the population, on social habits, on the economic situation and can be, therefore, modified by external actions, such as changes in the mobilities, or in resource allocations, and so on. In the following modelling these values will be factorized trying to put in evidence the contribution due to the specificity of the virus, equal in all the populations and contexts, the contributions related to the characteristics of the population/region at hand, such as the mean age, the general health condition (considered constant fixed in the analysed period), and the contribution related to the applied containment measures. The problem, already stated, is that there are still many unknown characteristics of COVID-19. The parameter  $\beta$  in epidemiology is the *transmission rate* and the quantity  $\beta SI$  is the *incidence*, that is the number of individuals who become infected per unit of time [13]. Among the characteristics that can influence  $\beta$  there are:

- $\beta_d$ : the density of the population; it is important for the contact probability among people;
- $\beta_m$ : the percentage of people that everyday take public transportation.

Note, that all these quantities, if increased, contribute to the boost of the incidence. The adopted lock down measures contribute to a decrease of  $\beta$  and therefore, as it will be clear in the next subsection, to the reduction of the reproduction number  $R$ .

As far as the recovery rate parameter  $\gamma$ , it also depends on many factors; among them, there are

- $\gamma_h$ : the hospital facilities and efficiency; an efficient sanitary system could fasten the healing of patients;
- $\gamma_m$ : the availability of general practitioners and specialists.

In this case the introduction of control actions should increase  $\gamma$ , becoming more effective the healing and therefore the transit from the  $I$  compartment to the  $R$  one. For the death rate  $d_I$  specific of COVID-19 patients the following aspects can be considered:

- $d_{I_s}$ : the percentage of smoking people; in most of the critical cases the respiratory system is compromised by the COVID-19 infection;
- $d_{I_n}$ : the percentage of people over 75 recovered in nursing home; this element is included since many elders died in these places;
- $d_{I_a}$ : the percentage of people with age over 30 years old; in fact it seems that young people are less susceptible to the virus;
- $d_{I_p}$ : the normalized number of cities where the allowed pollution threshold has been exceeded;
- $d_{I_h}$ : the percentage of people with at least two chronic diseases; the COVID-19 is dangerous mostly on people already weakened by other diseases.

As for the  $\beta$  parameters, the control actions aim at reducing the  $d_I$  value. The list of chosen characteristics that influence the parameters  $\beta$ ,  $\gamma$  and  $d_I$  is for sure not exhaustive, but at this very moment and at the best of the current knowledge, it may represent an acceptable attempt to link the models parameters with the specificity of a population. Therefore, it is possible to define:

$$\begin{aligned}\beta(t) &= k_\beta \beta_d \beta_m \beta_V(t)(1 - u_\beta(t)) \\ &= k_\beta \beta_{reg} \beta_V(t)(1 - u_\beta(t))\end{aligned}\quad (6)$$

$$\gamma(t) = k_\gamma \gamma_h \gamma_m \gamma_V u_\gamma = k_\gamma \gamma_{reg} \gamma_V(t) u_\gamma(t) \quad (7)$$

$$\begin{aligned}d_I(t) &= k_{d_I} d_{I_s} d_{I_n} d_{I_a} d_{I_p} d_{I_h} (1 - u_I(t)) \\ &= k_{d_I} d_{reg} d_V(t) (1 - u_I(t))\end{aligned}\quad (8)$$

where

- $\beta_{reg}$ ,  $\gamma_{reg}$ ,  $d_{reg}$  are the contributions, specific of each region, of the  $\beta$ ,  $\gamma$  and  $d_I$  respectively;
- $0 \leq u_\beta \leq 1$ ,  $0 \leq u_\gamma$  and  $0 \leq u_I \leq 1$  are the effects of the containment measures on  $\beta$ ,  $\gamma$  and  $d_{I_s}$  respectively;
- $k_\beta$ ,  $k_\gamma$ ,  $k_{d_I}$  are normalizing coefficients introduced for dimensional purpose.

The values  $\beta_V$ ,  $\gamma_V$  and  $d_V$  are the components of the  $\beta$ ,  $\gamma$ , and  $d_I$  parameters respectively, in which there is both the contribution typical of the virus and other (still) unknown not modelled aspects. In the case in which the considered listed characteristics were the unique significant ones, the quantities  $k_\beta \beta_V$ ,  $k_\gamma \gamma_V$  and  $k_{d_I} d_V$  would be the same in all the regions.

In the next Subsection the model is analysed, determining the reproduction number and enhancing its relation with the model parameters.

### C. Model analysis

In the SEIRD model (1)-(5), assuming constant all the model parameters, it can be easily deduced that a unique equilibrium point exists, and it is the disease free equilibrium one:

$$P_e = (S_e \ E_e \ I_e \ R_e \ D_e)^T = \left(\frac{b}{d} \ 0 \ 0 \ 0 \ 0\right)^T \quad (9)$$

By analysing the Jacobian matrix evaluated in  $P_e$  it can be easily verified the presence of four eigenvalues strictly negative,  $(-d, -d, -d - \alpha, -d - \gamma)$ , and one equal to zero.

As far as the evaluation of the reproduction number, it is used the *next generation matrix* method. The two equations (2)–(3), regarding the diffusion of the infection, may be rewritten as follows, once the vector  $Z = (E \ I)^T$  is defined:

$$\dot{Z} = F - V \quad (10)$$

with

$$F = \begin{bmatrix} \beta S I \\ 0 \end{bmatrix} \quad V = \begin{bmatrix} (d + \alpha) E \\ (d + \gamma) I - \alpha E \end{bmatrix} \quad (11)$$

The next generation matrix is defined as:

$$F_{P_e} V_{P_e}^{-1} = \begin{bmatrix} \frac{\beta b}{\alpha d} & \frac{\beta b(d + \alpha)}{d} \\ 0 & 0 \end{bmatrix} \quad (12)$$

and the reproduction number is the eigenvalue with the maximum modulus; therefore, in this case:

$$R_0 = \frac{\beta b}{\alpha d(d + \alpha)(d_I + \gamma)} \quad (13)$$

In the proposed model, the quantities  $\beta$ ,  $\gamma$  and  $d_I$  are assumed time dependent, since they include the effects of the control actions; nevertheless, in the first days of the pandemic the consequences of the containment measures were negligible; therefore in the expression (13) they are assumed constant and  $R_0$  may represent a suitable estimation of the dangerousness of the epidemic at the beginning of the spread. As usual,  $R_0$  shows a direct linear dependence on  $\beta$ ; this is what mainly inspired the containment measures that aimed at increasing the social distancing. An increased value of  $\gamma$ , that is an increment in the hospital efficiency, reduces the reproduction number. The value of  $d_I$  has the same behaviour of  $\gamma$ ; mathematically it means that also an increment of the death rate due to COVID-19 could reduce the reproduction number, which is reasonable since it implies a reduction of the number of infected patients and therefore of the contagions.

### D. Model parameters identification

As already noted, the proposed SEIRD model is kept simple for purpose aiming at the identification of the evolutions of  $\beta$ ,  $\gamma$  and  $d_I$  from the available real data of the total number of infected subjects,  $I_{tot}^R$ , of removed individuals,  $R_{tot}^R$ , and of dead patients,  $D_{tot}^R$ . They are obtained by minimizing the function:

$$J(\beta, \gamma, d_I) = \sqrt{p_1 e_I^2(t) + p_2 e_R^2(t) + p_3 e_D^2(t)} \quad (14)$$

where

$$e_I(t) = I_{tot}^R(t) - I_{tot}(t) \quad (15)$$

$$e_R(t) = R_{tot}^R(t) - R_{tot}(t) \quad (16)$$

$$e_D(t) = D_{tot}^R(t) - D_{tot}(t) \quad (17)$$

with  $I_{tot}$ ,  $R_{tot}$  and  $D_{tot}$  the total number of infected, removed and dead people respectively, obtained from the model (1)–(5); the errors  $e_I$ ,  $e_R$  and  $e_D$  are weighted by  $p_i > 0$ ,  $i = 1, 2, 3$ . A sequential quadratic programming method implemented by the @Matlab *fmincon* function is used.

### III. NUMERICAL RESULTS

In this Section the COVID-19 spread in some Italian regions is considered. As far as the death rate  $d$  (the one not depending on COVID-19), there is not a significant difference from one region to the other; from the ISTAT website [12] it is assumed  $d = 10^{-5}$ . The  $b$  parameter can be chosen as  $S_e d_e$ , as noted in Subsection II-B; therefore, it depends on the number of people in a region. The incubation rate  $\alpha$  is the inverse of the time needed by a subject to transit from the class of exposed subjects to the one of infected patients; from the physician it has been estimated between 3 and 10 days; in all the simulations it has been assumed  $\alpha = \frac{1}{7}$ . From the website of the Italian Civil Protection [11] several data are available starting from February 24 2020, the date in which officially COVID-19 has been recognized as a great sanitary problem in Italy; in particular, in what follows the daily total data of infected patients, of the removed individuals and of people specifically dead for COVID-19 are considered. The Italian regions on which the analysis is applied herein are: Lombardia, Veneto, Marche, Lazio, Calabria, Puglia and Sardegna; they are from the part of Italy most affected by the virus spread (the North), from the Centre, the South and from the most isolated large island of Italy. To put in evidence the differences of the virus spread in Italy, in Fig.1 it is shown the evolution of the number of daily new infections in the chosen regions in the period February 24 to June 14; the strong difference between what happened in Lombardia, where the peak of daily infections occurred in March 21 with 3251 new cases, and the situations in other regions is evident: Veneto with the maximum of 586 reached in the same day, Lazio with 210 cases on March 28 and Marche with a peak of 268 case on March 22. A different situation occurred in Puglia with the peak of 163 new infections reached on March 30, Calabria with 101 new cases on March 27 and Sardegna with 94 maximum cases on March 28. The differences are related, probably, to the fact that the containment measures were promptly applied in the entire Italian nation at the beginning of the spread in the North, even when the cases in the Center and South of Italy were not so numerous. Nevertheless, the reasons for which COVID-19 has been more aggressive in one region with respect to another are still under investigation. In Tables I, II and III the model parameters specific of each region are shown; they are deduced from the ISTAT website, [12]. In Table I the  $\beta_a$  values are normalized with the national density of population

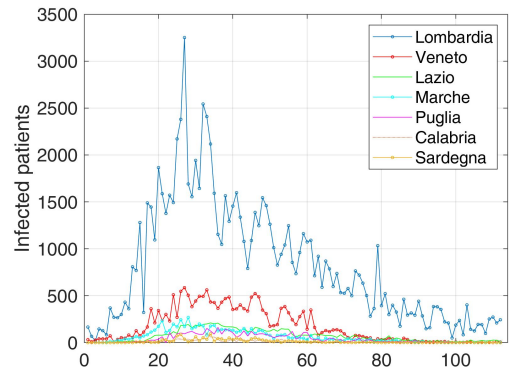


Fig. 1. Evolution of the number of daily new infections in Lombardia, Veneto, Lazio, Marche, Puglia, Calabria, Sardegna.

in Italy, currently estimated in 200 habitants for  $Km^2$ . For the mobility coefficient  $\beta_m$  it is considered the percentage of people (students and workers) that usually use public transportation. For the  $\gamma$  parameter representing the recovery rate influenced by the efficiency of the sanitary system, in Table II  $\gamma_h$  represents the number of hospital beds, whereas  $\gamma_m$  is percentage of general practitioners and specialists. From these data some elements stand out, in particular, in Table I the high mobility and density coefficients of Lombardia; up to now it has been argued that they have been among the most important cause of infection. From Table II it can be noted that the parameters values are almost equivalent in all the regions, whereas in Table III the most evident difference among the introduced parameters regards the possible pollution influence on mortality. By applying the minimization of the cost index (14), for each region, with  $p_1 = 0.4$ ,  $p_2 = p_3 = 0.3$ , it is determined the evolution of  $\beta$ ,  $\gamma$  and  $d_I$ . In Fig. 2 the fitting between real data of the total number of infected patients  $I_{tot}^R(t)$ , the removed individuals  $R_{tot}^R(t)$ , the dead subjects  $D_{tot}^R(t)$ , and the corresponding simulated quantities  $I_{tot}(t)$ ,  $R_{tot}(t)$ , and  $D_{tot}(t)$  is shown for Lombardia, Marche, Calabria and Sardegna, representative of all the parts of Italy. The period considered for each region starts on February 24 and ends when the curve of the total number of infected patients stops increasing; this is the reason for the difference abscissas. The evolutions of  $\beta$ ,  $\gamma$  and  $d_I$  in the first days of the diffusion may be considered still not influenced by the containment measures applied; the corresponding mean values  $\hat{\beta}$ ,  $\hat{\gamma}$ ,  $\hat{d}_I$  evaluated over the first week (February 24 – March 2) are reported in the last but one column of Table I- II and last column of Table III respectively. The estimation of  $\beta$  in the first week is almost the same in the regions considered; this suggests that an important element in the COVID-19 spread in Italy (and probably also in other countries) has been the number of people infected before the adoption of the containment measures; in Lombardia there were 166 patients on February 24, whereas in some regions of Center and South of Italy there wasn't any infected patient, probably only some

exposed. The delay (more or less one week) with respect to North and the contemporary adoption of control actions have probably saved many lives. The estimation  $\hat{\gamma}$  shows a significantly higher value in Lombardia, meaning that in the first week, due to the worrying number of infected people, there has been applied a great effort from a sanitary point of view with an increased number of physicians coming from the rest of Italy and other countries. As far as the death rate estimation  $\hat{d}_I$  the highest value is still obtained in Lombardia, as expected.

As an example of the evolution of  $\beta(t)$ ,  $\gamma(t)$  and  $d_I(t)$ , the corresponding results of the minimization of (14) for Lombardia, Veneto and Lazio are shown in Fig. 3; they have been chosen since they have the higher mobility coefficient  $\beta_m$  and the restriction of social contacts and displacements have been the most important and effective containment measures. The real data are rather noisy, as seen also in Fig. 1; therefore the output of the minimization process will be shown by using a moving average over 7 days, with an overlapping of 4 days. These evolutions include, as said in Subsection II-B, the effects of the control actions, such as the restrictions of social contacts, the increase in the healthcare effort, as well as the increase in the medical knowledge on the virus that helps in reducing mortality. For the  $\beta$  evolutions there is a sort of delay in their behaviours in Veneto and Lazio with respect to what happened in Lombardia. This is reasonable, since, despite the containment measures have been adopted everywhere in Italy in the same way since the first days of March, in Lombardia the situation appeared critical already in February and some restrictions were implemented efficiently since then. Also for the  $\gamma$  a similar delay is evident as well as a general strong effort from the middle of March; for the death rate  $d_I$  it can be appreciated how the control reduced the death rate due to the virus. As a final remark, consider the equations (6)- (8); the values of the  $\beta$ ,  $\gamma$  and  $d_I$  may be assumed constant in the first week of analysis and given by:

$$\beta = k_\beta \beta_{reg} \beta_V \quad \gamma = k_\gamma \gamma_{reg} \gamma_V \quad d_I = k_{d_I} d_{reg} d_V \quad (18)$$

For the left hand sides of equations in (18) the corresponding values  $\hat{\beta}$ ,  $\hat{\gamma}$ ,  $\hat{d}_I$  are used; this allows the estimation of  $k_\beta \beta_V$ ,  $k_\gamma \gamma_V$  and  $k_{d_I} d_V$  for each considered region. All these values should be equal if in  $\beta_{reg}$ ,  $\gamma_{reg}$  and  $d_{reg}$  the most significant characteristics of the virus are included. The mobility as well as the density of the population, while yielding useful information, are not enough to describe the modalities of infection of a susceptible individual in  $S$  class by an infected patient in  $I$ , see the last column of Table I. This consideration should help in determining the containment measures. Generally, the main control action applied has been the social distancing and the decrease of mobility (the lock down), thus reducing the dangerous contacts among people; the results of Table I suggest that probably some other important elements, not taken into account up to now, have influenced the contagion. These characteristics would also justify the extreme different behaviour of the virus in neighbouring regions. In Table II

a similar situation is obtained in the last column; this is reasonable, since the healing capability of an infected subject depends on reasons that are not easily identifiable. The ones considered here (the hospital capability in receiving patients and the number of physician) are only partially informative, especially in a nation, like Italy, in which there is a substantially homogeneous distribution of resources. The values of  $k_{d_I} d_V$  that can be deduced from Table III are of different order of magnitude; in particular, for Lombardia it is equal to 20.82, Veneto  $4.5 \cdot 10^{-5}$ , Lazio  $4.5 \cdot 10^{-3}$ , Marche 7.29, Puglia 0.0047, Calabria 14.79 and Sardegna 11.39. This implies that other features influence the death parameter  $d_I$ ; recently, the gender and the quantity of vitamin D seem to have a role in the susceptibility to the virus. The identification of the most informative features should help in determining effective medications, waiting for the vaccine.

TABLE I  
EFFECTS OF THE REGIONAL CHARACTERISTICS ON  $\beta$

	$\beta$ components			
	$\beta_d$	$\beta_m\%$	$\hat{\beta}$	$k_\beta \beta_V$
Lombardia	2.10	0.63	$1.4 \cdot 10^{-7}$	$1.09 \cdot 10^{-7}$
Veneto	1.33	0.21	$10^{-7}$	$3.58 \cdot 10^{-7}$
Lazio	1.70	0.43	$1.3 \cdot 10^{-7}$	$1.77 \cdot 10^{-7}$
Marche	0.81	0.07	$8.1 \cdot 10^{-7}$	$1.49 \cdot 10^{-5}$
Puglia	1.03	0.17	$3 \cdot 10^{-7}$	$1.7 \cdot 10^{-6}$
Calabria	0.64	0.08	$7.1 \cdot 10^{-7}$	$1.3 \cdot 10^{-5}$
Sardegna	0.34	0.06	$7.4 \cdot 10^{-7}$	$3.65 \cdot 10^{-5}$

TABLE II  
EFFECTS OF THE REGIONAL CHARACTERISTICS ON  $\gamma$

	$\gamma$ components			
	$\gamma_h\%$	$\gamma_m\%$	$\hat{\gamma}$	$k_\gamma \gamma_V$
Lombardia	0.28	0.37	0.156	$1.38 \cdot 10^{-6}$
Veneto	0.27	0.34	$1.1 \cdot 10^{-7}$	$1.2 \cdot 10^{-6}$
Lazio	0.25	0.46	$3.9 \cdot 10^{-6}$	$3.39 \cdot 10^{-5}$
Marche	0.25	0.38	$2.9 \cdot 10^{-5}$	$3.11 \cdot 10^{-4}$
Puglia	0.25	0.37	$2.8 \cdot 10^{-6}$	$3.2 \cdot 10^{-5}$
Calabria	0.19	0.38	$3.9 \cdot 10^{-4}$	0.0054
Sardegna	0.30	0.48	$1.3 \cdot 10^{-4}$	$9.3 \cdot 10^{-4}$

TABLE III  
EFFECTS OF THE REGIONAL CHARACTERISTICS ON  $d_I$

	$d_I$ components					
	$d_{I_s}\%$	$d_{I_n}\%$	$d_{I_a}\%$	$d_{I_p}$	$d_{I_h}\%$	$\hat{d}_I$
Lombardia	0.19	0.12	0.66	0.76	0.21	0.05
Veneto	0.16	0.12	0.67	0.95	0.19	$1.1 \cdot 10^{-7}$
Lazio	0.23	0.11	0.65	0.38	0.21	$6.3 \cdot 10^{-6}$
Marche	0.20	0.13	0.68	0.38	0.017	$8 \cdot 10^{-4}$
Puglia	0.17	0.11	0.65	0.15	1.9	$10^{-6}$
Calabria	0.16	0.11	0.64	0.15	0.24	$6 \cdot 10^{-3}$
Sardegna	0.19	0.12	0.69	0.14	0.26	$2.7 \cdot 10^{-4}$

#### IV. CONCLUSIONS

A data driven characterization of COVID-19 spread is proposed. A simple SEIRD model is proposed and discussed: the

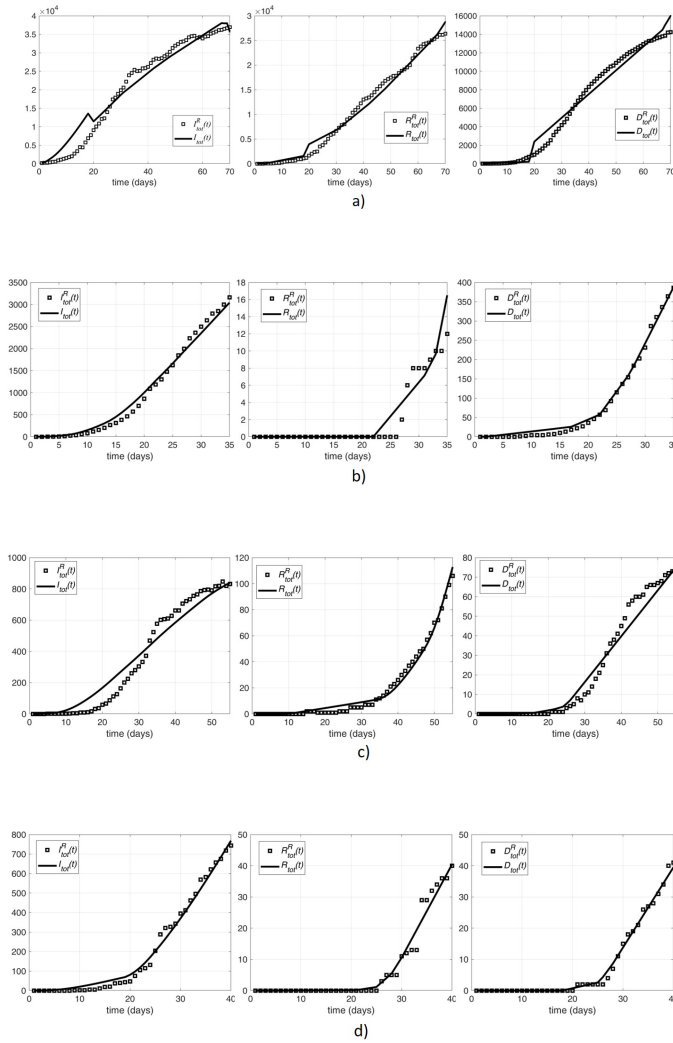


Fig. 2. Situation in Lombardia (row a), Marche (row b), Calabria (row c) and Sardegna (row d) in the first period on the epidemic spread; evolution of the total number of infected patients (left panel), removed individuals (centre panel), dead subjects (right panel).

classical SEIR model has been completed by the compartment of the patients dead by the COVID-19; the reason is that the unique official data regards the number of infected, removed and dead subjects. By using these data it has been possible the identification of the contagion rate, the recovery and the death rates evolutions; these parameters are strongly related to characteristics of the population, of the society and the region. From the numerical analysis it is evident that there are still elements not yet identified that influence these parameters. It is worth to be stressed the importance of the availability of data consistently collected: especially in the first days of pandemic, it has been difficult to manage with different modalities and timing of data collection. Future work will be devoted in data analysis regarding the identification of possible features characterizing regions and populations with respect to COVID-19 susceptibility; this could guide the choice of suitable and

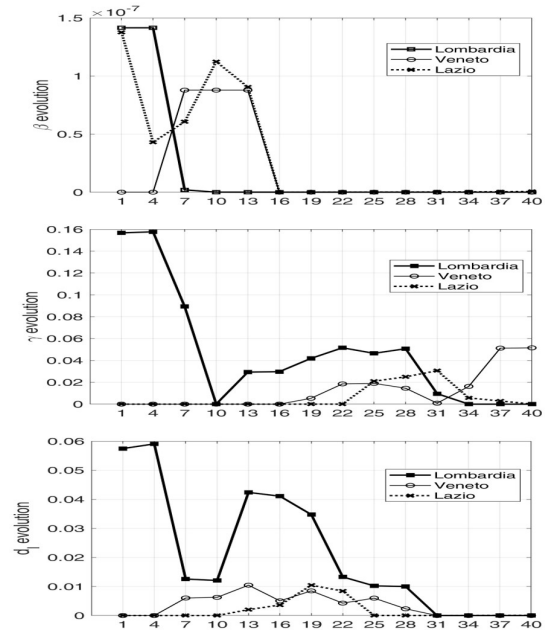


Fig. 3. Evolution of  $\beta$  (upper panel),  $\gamma$  (centre panel) and  $d_I$  (lower panel) in Lombardia, Veneto and Lazio respectively evaluated on a moving average: starting day February 24.

effective control actions.

## REFERENCES

- [1] "World Health Organization," <https://www.who.int/>.
- [2] "Istituto superiore di sanità," <https://www.iss.it/>, 2020.
- [3] P. Di Giamberardino and D. Iacoviello, "A control based mathematical model for the evaluation of intervention line in epidemic spreads without vaccination: the covid-19 case study," *Submitted to IEEE Journal of Biomedical and Health Informatics*, 2020.
- [4] G.Giordano, F.Blanchini, R.Bruno, P.Colaneri, A.DeFilippo, and M.Colaneri, "Modeling the covid-19 epidemic an implementation of population wide interventions in italy," *Nature Medicine*, vol. 26, 2020.
- [5] C.C.Lai, T.P.Shih, W.C.Ko, H.J.Tang, and P.H.Hsueh, "Severe acute respiratory syndrome coronavirus 2 (sars-cov2) and coronavirus disease - 2019 (covid-19): the epidemic and challenges," *International Journal of Antimicrobial Agents*, in press.
- [6] J. Wu and K.Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study," *Lancet*, pp. 1–9, 2020.
- [7] S.Zhao, Q.Lin, J.Ran, S.S.Musa, G.Yang, W.Wang, Y.Lou, D.Gao, L.Yang, D.He, and M.H.Wang, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases*, vol. 92, pp. 214–217, 2020.
- [8] Z.Yang and et al., "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of thoracic disease*, vol. 12, no. 3, pp. 165–174, 2020.
- [9] S.Chatterjee, A.Sarkar, S.Chatterjee, K.Mintu, and P.Raja, "Studying the progress of covid-19 outbreak in india using sird model," *Indian Journal of Physics*, 2020.
- [10] P. Di Giamberardino, D. Iacoviello, F.Papa, and C.Sinisgalli, "Dynamical evolution of covid-19 in italy with an evaluation of the size of the asymptomatic infective population," in press, *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [11] "Protezione civile," <http://www.protezionecivile.gov.it/>, 2020.
- [12] "Istituto nazionale di statistica," <https://www.istat.it/en/>, 2020.
- [13] M. Martcheva, "An introduction to mathematical epidemiology," *Text in Applied Mathematics 61*, Springer, 2015.