Differential evolution to estimate the parameters of a SEIAR model with dynamic social distancing: the case of COVID-19 in Italy

I. De Falco[1], A. Della Cioppa[1,2], U. Scafuri[1], E. Tarantino[1]

[1]INSTITUTE FOR HIGH−PERFORMANCE COMPUTING AND NETWORKING (ICAR), NATIONAL RESEARCH COUNCIL OF ITALY (CNR), NAPLES, ITALY; [2]NCLAB, DIEM, UNIVERSITY OF SALERNO, FISCIANO (SA), ITALY

1. Introduction

Several compartmental models have been designed in epidemiology to simplify the mathematical modeling of infectious diseases, so as to describe their spreading in a population of individuals. Among them, we will make use here of the SEIAR [1−3] that expands the basic SIR model [4] and the SEIR one [5].

The choice of SEIAR model is due to the fact that we wish to estimate here the spreading of the coronavirus COVID-19 in Italy. In fact, several papers, as, e.g., Ref. [6], have stressed the issue that for this pandemic, the number of asymptomatic infectious subjects is very high. Given that asymptomatic subjects are obviously not contained in official figures, their presence causes a much wider and longer spread of this disease, with more infectious people. Moreover, an important remark on the use of the SEIAR model [6] is that the basic reproduction number R_0 it computes is much higher than that provided by the use of SIR and SEIR models. In many cases, this allows better explaining the fast increases experienced in COVID-19 outbreaks in many countries worldwide. For example, in the Diamond Princesss cruise ship outbreak, R_0 has been recently estimated as 14.8 [7], which is much higher than the value proposed by the World Health Organization for COVID-19. In that paper, the use of R_0 values typically acknowledged for COVID-19 as 3.7 does now allow obtaining a good modeling of what happened aboard that ship.

As a consequence, unlike most papers dealing with COVID-19, we feel that a model as SEIR, which does not take these asymptomatic subjects into account, cannot suitably model COVID-19. SEIAR, instead, explicitly deals with asymptomatic infectious subjects.

In these days, an extension to SEIR has been proposed by Hubbs that is called SEIR with social distancing (SEIR-SD) [8]. As the name suggests, this latter model also accounts for issues related to keeping individuals as distanced as possible one from another.

In this chapter, we introduce a new model by further extending SEIAR through the utilization of a user-defined dynamic social distancing (DSD) by means of the definition of a time-dependent social distancing function. As a result, we obtain the SEIAR−DSD model.

To properly describe a real-world situation, the exact values of the model parameters that control the rate of spread from different compartments must be found. In fact, even slight differences in these values could lead to completely different spread of the diverse disease phases described by the model equations. Finding this set of values is a far-from-easy task, as a huge number of possible such combinations exists. Technically speaking, this is an NP −hard problem. This is the main reason for the practical limitation on the use of the above-described models. Nevertheless, considered the advances of the research, this task can be suitably tackled through the use of heuristic optimization techniques. From among the many available such techniques, we have decided to investigate the differential evolution (DE). Its description is beyond the scope of this chapter, and the interested reader can find more detailed information in Ref. [9]. Suffice it to say here that DE is a heuristic population-based optimization technique, within the machine learning technology, widely used in artificial intelligence. The scientific literature has proved that DE is extremely able in finding optimal or sub-optimal solutions to multidimensional real-valued problems, even multiobjective and constrained, taken from many different fields.

By exploiting the ability of this evolutionary technique in finding the most suitable set of parameters of the SEIAR−DSD model, we aim at applying it to Italy and to some of its most important regions, and at approximately evaluating for each of these test cases the daily number of infectious individuals, the day(s) in which this number will reach its highest peak, the corresponding value of the peak, and the future evolution of the spread.

This study could serve as a useful guideline to Italian Government as well to the Government of any other State in which COVID-19 spreading is occurring.

Section 2 describes the state of the art and helps in understanding the novelty of our approach. Section 3 presents the SEIAR model and introduces our SEIAR− DSD approach. Section 4 contains our simulations, and provides details on the methods, the experimental results, and a discussion. Our conclusions are given in Section 5.

2. Related works

As a consequence of the emerging in Wuhan, China, in December 2019 of a novel coronavirus (SARS.Cov-2), several studies relying on predictive mathematical models [5,10,11] have been employed to estimate the progression of the COVID-19 pandemic

with the goal to provide some insights to plan effective control strategies able to limit the worldwide infection. A short summary of these works is reported in the following.

Li et al. [12] exploit statistical methods on the novel case of COVID-19 by collecting demographic information, exposure history, and the progress of laboratory-confirmed cases in Wuhan to determine the spread of the disease.

Batista [13] tries to estimate the final size of the pandemic for the whole World using the logistic grow model [14−16] and the SIR model [17]. The model parameters are estimated exploiting the time series of available data and minimizing the difference between the actual and predicted number of total number of cases (susceptible and infected), and the Shanks transformation to evaluate the series limit for the prediction of the final number of recovered persons.

Chen et al. [18] propose the employment of a time-dependent SIR model, where both the parameters, i.e., the transmission rate and the recovering rate, are functions of time. They use machine learning methods to estimate the parameters and then exploit such parameters to predict the number of the infected persons and the recovered persons at a certain time in the future. The time-dependent SIR model is able to dynamically adjust its parameters.

Li et al. [19] develop a mathematical model to simulate the spatio-temporal spread of infections among 375 Chinese cities. Spatial spread of SARS-CoV2 is acquired by considering the daily number of people traveling between the different cities and a multiplicative factor. Four model state variables, namely the susceptible, exposed, documented infected, and undocumented infected (SEII) sub-populations in a specific city are estimated. The model parameters are inferred by using an iterated filter-ensemble adjustment Kalman filter [20,21]. Framework Prem et al. [22] examine how unprecedented measures taken in Wuhan in response to the adopted restrictions, including the closures of school and workplaces, have affected the virus progression. The authors use an age-structured SEIR model under different scenarios of restrictions of social contacts. To simulate the pandemic, they exploit parameter values taken from literature.

López and Rodó [23], by following the approach of Peng et al. [24], apply a modified SEIR compartmental model accounting for the spread of infection for different levels of population isolation from undiagnosed individuals to estimate the effects of the reduction in personal contacts in the pandemic evolving. Public data are used to estimate the designed model parameters. Specifically, an optimization algorithm capable of evaluating the normalized least-squares error of the model approximation and the infected reported cases is considered.

Lin et al. [25] adopt a SEIR model to estimate the progression of COVID-19 spread in Wuhan after the control measures taken by the Chinese government including setting up special hospitals, travel restrictions, quarantine of patients to mitigate the spread. The model parameters are considered either constants or stepwise functions based on assumptions or on precedent pandemic studies.

Yang et al. [26] exploit a modified SEIR epidemiological model that considers the domestic migration data before and after a starting date together with recent COVID-19 epidemiological data to predict the outbreak progression. The model prediction is corroborated by using a machine-learning artificial intelligence approach trained on the data related to 2003 SARS coronavirus epidemic. The parameters are derived by performing multiple fitting on the available data related to the province of Hubei.

Hubbs [8] introduces the concept of social distancing in the SEIR model. The aim is to incorporate in such a model the ensemble of strategies devised to reduce person-to-person contact, as, for example, city lock down, school and universities closures, smart working and quarantine of people, with the aim to slow the disease spreading.

Mwalili et al. [27] apply the epidemiological SEIR model for the evaluation of COVID-19 pandemic dynamics by incorporating pathogen in the environment and intervention of social distancing. The next generation matrix approach is employed to find the basic reproduction factor R_0.

Lyra et al. [28] develop a modified SEIR model incorporating the concept of social distancing which includes movement restrictions, asymptomatic transmission, quarantine, and hospitalization. The rate of transmission is dynamic and deduced from the observed delayed fatality rate, while the parameters of the epidemic are evaluated by a Markov chain Monte Carlo algorithm.

Giordano et al. [29] propose an extension of the classical SIR model, named SIDARTHE, for the COVID-19 pandemic in Italy. In such a model, the total population is subdivided into eight interacting stages of disease. The official data of Civil Protection and Ministry of Health are exploited to infer the model parameters. These parameters are updated over time to take into account the adoption of progressive restriction measures. A best-fit approach is used to find the parameters by reiterating a local minimization of the sum of the error squares.

Summarizing, in these days, wide use is being made in the scientific literature of the SEIR model. As we have seen, this model seems not perfectly suited to study this COVID-19 pandemic because several studies have found that a high number of asymptomatic infectious subjects exist, who cannot be modeled by SEIR.

Hence, differently from all the papers above, we consider a SEIAR−DSD framework, which can explicitly account for asymptomatic infectious subjects, and we exploit an evolutionary algorithm to optimize the model parameters.

As far as we know, the only paper making use of SEIAR model to deal with COVID-19 is that by Pribylova and Hajnova [6]. They introduce and derive the basic reproduction number as the weighted arithmetic mean of the basic reproduction numbers of the symptomatic and asymptomatic cohorts. They show that European pandemic outbreaks in various European countries correspond to the simulations with commonly used parameters based on clinical characteristics of the disease COVID-19, but R_0 is around three times bigger if the asymptomatic cohort is taken into account, with values up to around nine. They investigate on the introduction of quarantine and social measures and on their effects by setting variations in the β parameter on given days.

Yet, in their model, there is no explicit introduction of a continuous time-varying social distancing function, nor do they find the values of the model parameters through an optimization phase based on machine learning algorithms, so our SEIAR−DSD model seems to be original with respect to the existing literature.

3. The SEIAR model

SEIAR model [1−3] represents an extension to both the SIR [4] and the SEIR [5] models.

In fact, SIR model is composed by three compartments, and accounts at any given time t for the number of Susceptible (S), Infectious (I), and Recovered (R) subjects. Hence it is based on three time functions $S(t)$, $I(t)$, and $R(t)$. These functions are normalized so that at any given time t the following holds: $S(t) + I(t) + R(t) = 1$.

With respect to it, both SEIR and SEIAR take into account the fact that there exists an incubation period in which an individual has been infected but is yet infectious. This is represented by a further compartment containing such exposed (E) individuals, and by the time function $E(t)$.

Moreover, with respect to SEIR, in SEIAR also asymptomatic infectious (A) subjects are explicitly considered, leading to an explicit further compartment for them, and to the time function $A(t)$.

As its predecessors, also SEIAR model accounts for vital dynamics in terms of birth rate and death rate, so that the population size can vary. Actually, given the short lifetime considered here, we are not interested in this feature, so that $S + E + I + A + R = N$ and the equations of this model reduce to:

$$\frac{dS}{dt} = -\beta \cdot \frac{S(t) \cdot [I(t) + q \cdot A(t)]}{N} \tag{5.1}$$

$$\frac{dE}{dt} = \beta \cdot \frac{S(t) \cdot [I(t) + q \cdot A(t)]}{N} - \alpha \cdot E(t) \tag{5.2}$$

$$\frac{dI}{dt} = p \cdot \alpha \cdot E(t) - \gamma \cdot I(t) \tag{5.3}$$

$$\frac{dA}{dt} = (1 - p) \cdot \alpha \cdot E(t) - \gamma \cdot A(t) \tag{5.4}$$

$$\frac{dR}{dt} = \gamma \cdot [I(t) + A(t)] \tag{5.5}$$

where:

− N is the sum of S, E, I, A, and R;
− α represents the inverse of the incubation period ($1/t_{incubation}$);
− β represents the average contact rate in the population;
− γ represents the inverse of the mean infectious period ($1/t_{infectious}$);

− q represents the probability that an asymptomatic individual transmits the disease. It is a fraction of the probability that an infectious individual transmits the disease;
− p represents the fraction of the exposed individuals who become infectious.

If we impose that the sum of the five equations above must be equal to 0, we obtain that the population size N remains constant. Hence, the dynamics of this model is completely determined by setting the values for the parameters discussed above. All of them may take on nonnegative real values.

It should be remarked that in this model, as well as in SIR and SEIR, R represents the sum of both the individuals who actually return healthy after being infected and of those who die due to the epidemic.

A very important parameter in SIR model and in its derivations is the ratio, $R_0 = \dfrac{\beta}{\gamma}$ called basic reproduction ratio. It represents the expected number of currently susceptible individuals that will be infected by an infectious one, so it accounts for the degree of infectiousness of the specific epidemic being examined. In SEIAR model, it is more precisely computed as:

$$R_0 = \beta \cdot \left[p \cdot \frac{1}{\gamma} + (1-p) \cdot \frac{q}{\gamma} \right] \tag{5.6}$$

3.1   The SEIAR model with social distancing

With the aim at reducing the spreading of COVID-19, as in many other countries in which COVID-19 virus is present, also in Italy the Government has decided to enforce *social distancing*. Basically, many shops considered unnecessary have been closed, and so have been schools and universities. Travel restrictions have been established, smart working has been allowed wherever possible, and quarantine of persons has been imposed. Furthermore, in all public spaces left open to people, as food shops and supermarkets and pharmacies, and in the streets as well, a distance of at least 1 m should be kept between people.

The way this influences the spreading of the virus has been modeled in several ways. We will make reference here to the approach recently proposed by Hubbs [8] for SEIR, and will translate it to the needs of the SEIAR model. We will shortly refer to it as SEIAR with Social Distancing (SEIAR−SD). This is based on the introduction of a new real-valued parameter ρ, ranging within 0.0 and 1.0. The value 0.0 represents the ideal case in which everyone is locked down in quarantine, whereas the value 1.0 reduces this model to the SEIAR model in which no social distancing is considered. The introduction of ρ implies the modification in the SEIAR model of the equations related to S and E, so that the SEIAR−SD model is the following:

$$\frac{dS}{dt} = -\rho \cdot \beta \cdot \frac{S(t) \cdot [I(t) + q \cdot A(t)]}{N} \tag{5.7}$$

$$\frac{dE}{dt} = \rho \cdot \beta \cdot \frac{S(t) \cdot [I(t) + q \cdot A(t)]}{N} - \alpha \cdot E(t) \tag{5.8}$$

$$\frac{dI}{dt} = p \cdot \alpha \cdot E(t) - \gamma \cdot I(t) \tag{5.9}$$

$$\frac{dA}{dt} = (1 - p) \cdot \alpha \cdot E(t) - \gamma \cdot A(t) \tag{5.10}$$

$$\frac{dR}{dt} = \gamma \cdot [I(t) + A(t)] \tag{5.11}$$

Strictly speaking, ρ is another parameter of the model, so that this depends on six parameters, i.e., α, β, γ, p, q, and ρ. Their values should be suitably found, so that the model can closely represent the situation being studied.

3.2   The SEIAR–SD model with dynamic social distancing

We have noticed that keeping a value of ρ constant during the whole evolution of the pandemic over time is unrealistic. Rather, what happened in the Italian case is that in the first days, in which we are aware of just few cases, no social distancing was enforced. Then, after a few days, some general rules of thumb were suggested, as avoiding unnecessary travels and exits from homes and, in such cases, keeping at a distance of at least 1 m from other people. As the situation started to get worse, some parts of Italy were isolated, and later on movements within Italy were more and more limited. With time, more and more activities and shops were closed by Government. Finally, even going out from home was very strictly limited by law. Of course, this cannot be represented by a single value of ρ kept constant over time.

Hence, we have designed a time-varying (or dynamic) social distancing function by considering the different decrees issued on different dated by Italian First Minister, which have led over time to lower and lower freedom of movement for Italian citizens. We represent it as $\rho(t)$ by using the logistic function:

$$\rho(t) = \frac{\rho_i - \rho_f}{1 - e^{k(t - t_0)}} + \rho_f \tag{5.12}$$

where ρ_i and ρ_f are the values of ρ on the first and the last day, t_0 is the time of the inflection point, i.e., the date of the steepest decline in ρ (the main lockdown date), and k lets us vary how quickly ρ declines. To choose the values for ρ_i, ρ_f, t_0, and k, so that Eq. (5.12) can describe at its best the situation in Italy, we have downloaded the mobility data for Italy. This latter is publicly available thanks to mobility trends reports which are published daily and reflect requests for directions in Apple Maps (https://www.apple.com/covid19/mobility). These data take into account both driving, walking, and public transportation mobility, so we feel it can well represent what has happened in Italy. Based on such data, we have set $\rho_i = 1$, $\rho_f = 0.05$, $t_0 = 10$ and $k = 0.1$.

As a consequence of the above, in Eqs. (5.7) and (5.8), we use $\rho(t)$ rather than ρ, and we name such model as SEIAR–DSD.

In the following of our simulations, we will always make use of this time varying function.

4. Simulations

4.1   Methods

The simulations are carried out by using the SEIAR−DSD model on the data set reporting the numbers of infectious cases taking place in each of these days in Italy and in each of its regions, as they can be downloaded from the Ref. [30] free repository, thanks to Italian Ministry for Health.

Our goal is to be able to evaluate the approximate daily numbers of infectious individuals until the end of virus spreading, the approximate day(s) in which this number will be at its highest peak, and the approximate day in which the infected cases will become very close to zero. We will make reference both to Italy as a whole, and to some of its most important regions starting from February 24th, 2020. This study could serve as a useful guideline to Italian Government as well as to the Government of any other State in which COVID-19 spreading is occurring.

From an optimization point of view, we have to find the best possible set of values for the SEIAR−DSD model parameters that allows us to follow as closely as possible the series of the real infection cases taking place in Italy in these days. As already said above, we make use here of DE.

Given a real-valued optimization problem with size L, differential evolution is an optimization method that starts by randomly creating an initial set of possible solutions to the problem, each represented as a vector of L real-valued numbers. The cardinality of the population is termed *population size* (P_s) and is kept constant during the evolution. Each time a solution is created, its quality at solving the problem is evaluated by means of a suitable fitness function Φ that must be optimized. Then starting from the current population, a new one is created for the next generation thanks to the use of a suitable mutation mechanism. Many different mechanisms exist, basically each of them starts from the generic i−th individual, receives in input a set of other randomly-chosen current individuals (two, three, or four) and, based on the values of two parameters, named *mutation factor* (F) and *crossover ratio* (C_r), mixes their features to those of the i−th individual so as to obtain a new trial individual. This latter is compared to the current i−th one, and the better enters the new population under construction. This mechanism is repeated for P_s times at each generation, so that at the end we will obtain the new P_s−sized new population for next generation. This is repeated for a number of generations represented by G. The individual with the best fitness value in the final population is the solution found by the algorithm.

For the problem at hand, differential evolution should find the most suitable values for the parameters of the SEIAR−DSD model. We find these values by taking as the goal of the optimization process the minimization of the root mean square error (RMSE) between the number of infectious cases estimated by the SEIAR−DSD model I_e and the actual number of infectious cases in Italy I_a, where this RMSE is computed over the

number of days N_d starting from the official onset of COVID-19 in Italy and the last day for which we have actual data available. In formulas:

$$\text{RMSE} = \sqrt{\frac{1}{N_d} \cdot \sum_{i=1}^{N_d} (I_a - I_e)^2}$$ (5.13)

Actually, to avoid overfitting, we do not provide all the available data series to the SEIAR–DSD model. Rather, we divide the series into two sections: we perform training onto the first, and test the model onto the second. During the evolution of the DE run, we only consider the RMSE over the train set. At the end of the run, we consider the solution with the lowest such RMSE, and we take into account its RMSE value over the test set.

Given the length of the data series, we use the first 62 data for train and the last three for test. Only for Campania region, we use 59 data for train and 3 for test. This is because the related data series is shorter by three data than those for Italy and the other regions, as COVID-19 in Campania appeared on February 27, 3 days later than in the rest of Italy (February 24).

Once obtained in this way, the set of parameter values minimizing the RMSE, by running the SEIAR–DSD with those values we will attain the evolution of Covid-19 in Italy. Consequently, we will be able to approximately estimate the daily number of infectious cases. Interestingly, we will be able to approximately estimate the day(s) in which this number will be at its highest peak, and the future evolution of the spread.

Based on our experience in the use of differential evolution, and after a short preliminary phase of parameter tuning, we have decided to use the JADE [31] version that is able to auto-adapt the F and C_r parameters, and to make use of the following values for the remaining parameters: population size 40, number of generations 500, *rand/1/bin* as the mutation mechanism.

4.2   Simulation results

We have taken into account a set of test cases, namely Italy as a whole, and then some of its most important regions, i.e., Lombardy, Veneto, Emilia–Romagna, and Campania. We will report on each of these cases in next subsections.

4.2.1   Modeling Italy

Italy was probably the first country in Europe to suffer from a strong onset of COVID-19, and it precedes in time countries as Germany, France, Spain, United Kingdom, and USA as well, by 1 week or two. Hence, studying the evolution over time of this pandemic in Italy is of high interest to all European countries, both to understand diffusion mechanisms and to foresee the future evolution. Italian population as of end 2019 is equal to 60,359,546 people. The execution of DE has allowed us to determine for Italy the most suitable values for the SEIAR–DSD model as follows: $\alpha = 0.08$, $\beta = 1.91$, and $\gamma = 0.13$. These values correspond to a value of the RMSE equal to $2.49 \cdot 10^{-5}$.
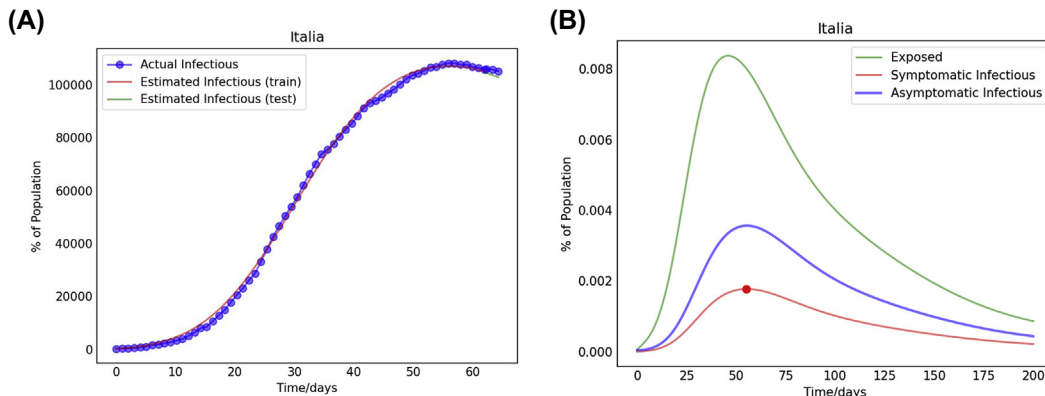
(A)

(B)

FIGURE 5.1 Daily infectious cases: the data computed by our approach and the data for Italy (A). The evolution over time of COVID-19 pandemic in Italy based on the parameter values found for the SEIAR−DSD model by differential evolution (B).

As it can be seen in Fig. 5.1A, the data provided by the model are in excellent accordance with the real-world data.

The use of the above parameter values leads to model the diffusion of coronavirus COVID-19 in whole Italy as it is shown in Fig. 5.1B.

From the figure, we can appreciate that the peak of the infection is expected to take place around April 23, with a number of infectious subjects equal to about 107,596.

4.2.2   Modeling Lombardy

We have run the same experiments by taking into account Lombardy region. This is important because it is the most populated Italian region with about 10 million inhabitants (10,060,574 is the official data for 2019). Moreover, Lombardy registered the onset in Italy, has currently the highest number of infected subjects, and registers the highest number of deaths due to this pandemic. After the first infection area (Lodi province), other highly populated areas of this region experienced COVID-19 (Bergamo, Brescia, and Milan provinces).

Following the general procedure shown above, we have run the DE algorithm to find the best parameter values for the SEIAR−DSD model. For region Lombardy, the set of parameters has resulted to be: $\alpha = 0.08$, $\beta = 2.34$ and $\gamma = 0.08$, $q = 0.41$, $p = 0.19$, with an RMSE value equal to $1.2 \cdot 10^{-4}$.

As it can be seen in Fig. 5.2A, the accordance of the estimated data and the actual ones is worse than that for the Italy case. In our opinion, this is probably due to some anomalies in the collected data, as it is evident in the slope of the actual curve.

The use of the above parameter values leads to model the diffusion of coronavirus COVID-19 in Lombardy as it is shown in Fig. 5.2B.

From the figure, we can appreciate that the peak of the infection is expected to take place in Lombardy around April 19, with a number of infectious subjects equal to about 33,493.
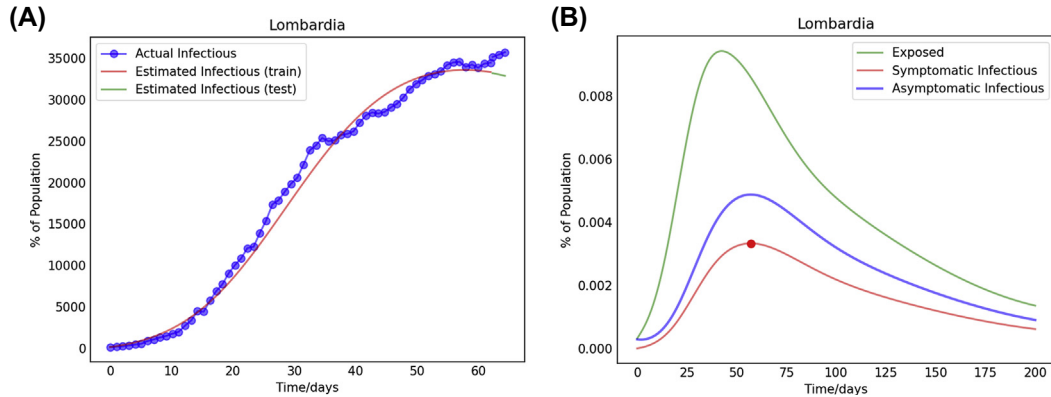
FIGURE 5.2 Daily infectious cases: the data computed by our approach and the data for Lombardy (A). The evolution over time of COVID-19 pandemic in Lombardy based on the parameter values found for the SEIAR−DSD model by differential evolution (B).

4.2.3 Modeling Veneto

We have also considered Veneto region. This is important because it is the other region in Italy in which COVID-19 virus started spreading. Veneto had about 4906 million inhabitants in 2019. Yet, the spread evolution has been different, with better results in terms of protection of its inhabitants. This has been due to the immediate enforcement of a restricted area where the onset was found (Vò town and surroundings), and to an immediate and high number of tests being performed.

Following the general procedure shown above, we have run the Differential Evolution algorithm to find the best parameter values for the SEIAR−DSD model. For region Veneto, the set of parameters has resulted to be: $\alpha = 0.09$, $\beta = 3.16$, and $\gamma = 0.21$, $q = 0.14$, $p = 0.96$, with an RMSE value equal to $3.5 \cdot 10^{-5}$.

Fig. 5.3A shows that the data estimated by the model closely fit the real-world data.
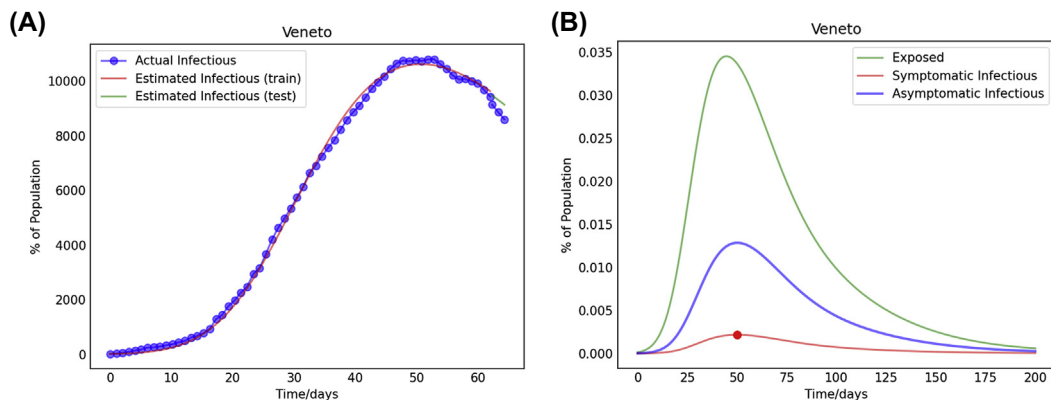

FIGURE 5.3 Daily infectious cases: the data computed by our approach and the data for Veneto (A). The evolution over time of COVID-19 pandemic in Veneto based on the parameter values found for the SEIAR−DSD model by differential evolution (B).

The use of the above parameter values leads to model the diffusion of coronavirus COVID-19 in Veneto as it is shown in Fig. 5.3B.

From the figure, we can appreciate that the peak of the infection is expected to take place in Veneto around April 14, with a number of infectious subjects equal to about 10,621.

4.2.4   Modeling Emilia–Romagna

We have also examined Emilia–Romagna region. This is an important region in North Italy, with about 4.459 million inhabitants (10,060,574 is the official data for 2019). Moreover, Emilia–Romagna was the third region in Italy to experience COVID-19 onset, probably due to people living there and traveling to the neighboring Lombardy. Measures taken in this region are somehow intermediate between those of Lombardy and Veneto, so that, although experiencing COVID-19 later than Veneto, with time the figures for this region have become higher than those of Veneto.

Following the general procedure shown above, we have run the Differential Evolution algorithm to find the best parameter values for the SEIAR–DSD model. For region Emilia–Romagna, the set of parameters has resulted to be: $\alpha = 0.07$, $\beta = 6.52$, and $\gamma = 0.30$, $q = 0.44$, $p = 0.39$, with an RMSE value equal to $3.5 \cdot 10^{-5}$.

In Fig. 5.4A, it can be observed that the estimated data are in excellent agreement with the real-world ones.

The use of the above parameter values leads to model the diffusion of coronavirus COVID-19 in Emilia–Romagna as it is shown in Fig. 5.2B.

From the figure, we can appreciate that the peak of the infection is expected to take place in Emilia–Romagna around April 13, with a number of infectious subjects equal to about 13,709.
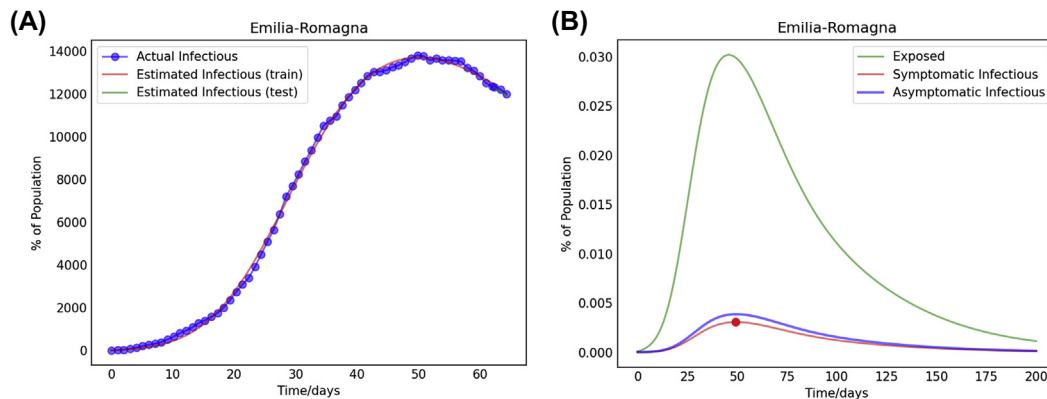

FIGURE 5.4 Daily infectious cases: the data computed by our approach and the data for Emilia–Romagna (A). The evolution over time of COVID-19 pandemic in Emilia–Romagna based on the parameter values found for the SEIAR–DSD model by differential evolution (B).

4.2.5   Modeling Campania

The investigation of Campania is important because Campania is the third most populated region in Italy, with almost six million inhabitants (5,861,529 is the official data for 2019). Moreover, the interest lies in the fact that Campania has been infected much later than Lombardy, Veneto, and Emilia−Romagna, is distant from these latter, and, as of today, shows a much lower number of both infectious cases and deaths. Therefore, we wonder whether the evolution over time of this pandemic in Campania can be different from that in the other three regions.

Also in this case we have run the DE algorithm to find the best parameter values for the SEIAR−DSD model by using the data for the infectious subjects in Campania. By running our experiments on the data for Campania region, instead, the set of parameters has resulted to be: $\alpha = 0.14$, $\beta = 5.19$, and $\gamma = 0.21$, with an RMSE value equal to $9.07 \cdot 10^{-6}$.

Also for the Campania case, the data provided for the infectious individuals are in excellent accordance with the real-world data as shown in Fig. 5.5A.

The use of the above parameter values leads to model the diffusion of coronavirus COVID-19 in Campania as it is shown in Fig. 5.5B.

From the figure, we can appreciate that the peak of the infection in Campania is expected to take place around April 13, with a number of infectious subjects equal to about 3088.

4.3   Discussion

A first important issue that should be underlined is that in all of the examined test cases the curve of the asymptomatic infectious subjects is always higher than that of the infectious ones. This means that, at least in Italy, asymptomatic subjects are a numerically very important vehicle of transmission for COVID-19. This issue is very evident
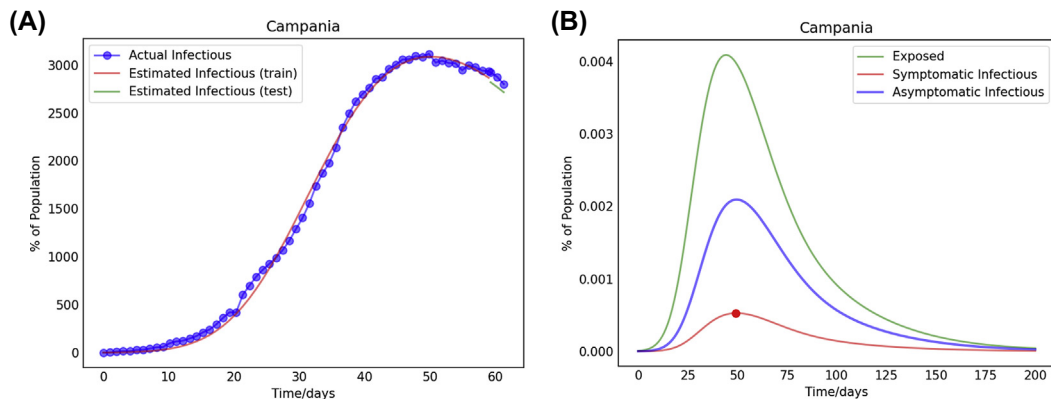

FIGURE 5.5 Daily infectious cases: the data computed by our approach and the data for Campania (A). The evolution over time of COVID-19 pandemic in Campania based on the parameter values found for the SEIAR−DSD model by Differential evolution (B).

especially when we look at the curves for Italy, Veneto, and Campania, whereas this can be appreciated at a smaller extent for Lombardy and Emilia−Romagna. This seems to reflect the different containment policies decided by the different regions: where drastic policies have been chosen and high numbers of tests have been immediately made, the numbers of Asymptomatic subjects are lower both in absolute terms and with respect to Infectious.

A second issue to be discussed is related to the values for the parameters that have been obtained by differential evolution for the different test cases.

Table 5.1 summarizes for each of the test cases the obtained values, as well as the approximate date of the peak, the approximate height of the peak, and the RMSE obtained in the run.

Values for α and γ are not too different from a test case to another. This is to be expected, as they are respectively related to incubation period and mean infectious period, which are both features specific to this virus itself.

It is worth paying attention to the different values for q and p obtained in the different test cases. For example, the value of q for Veneto, 0.14, is much lower than that for Lombardy, 0.41. At a first sight, this difference may appear odd, especially if we consider that COVID-19 outbreak in Italy started in these two regions, so the starting situation was about the same for them. Actually, this difference is due to the different policies decided by the local governments of these two regions. In fact, on the one hand, Veneto immediately started a wide campaign of tests through swabs that as of today (April 30), consists of a total of 337,656 subjects being tested, corresponding to 6.88% of the population. Lombardy, on the other hand, was less responsive in this, in the first days tested less people, and as of today has tested 365,895 subjects, corresponding to 3.64%, so about half the percentage of Veneto. Because of this, on the one hand, in Veneto many subjects that would have been asymptomatic have been correctly recognized as infectious, hence its lower value for q and, consequently, its very high value for p: summarizing, few asymptomatic, many Infectious. For Lombardy, instead, the opposite is true: initially a lower number of tests was performed, so many asymptomatic subjects remained undiscovered, and their contribution to the spread of the virus in that region is noticeably higher than in Veneto, which results in $q = 0.41$, and, conversely, the

Table 5.1 The parameter values and other information for the different test cases investigated.

Case	α	β	γ	Q	p	Peak date	Peak value	RMSE
Italy	0.08	1.91	0.13	0.33	0.97	April 23	107,596	$2.4 \cdot 10^{-5}$
Lombardy	0.08	2.34	0.08	0.41	0.19	April 19	33,493	$1.2 \cdot 10^{-4}$
Veneto	0.09	3.16	0.21	0.14	0.96	April 14	10,621	$3.5 \cdot 10^{-5}$
Emilia Romagna	0.07	6.52	0.30	0.44	0.39	April 13	13,709	$3.5 \cdot 10^{-5}$
Campania	0.14	5.19	0.21	0.20	0.43	April 13	3088	$9.7 \cdot 10^{-6}$

infectious contributed percentually less to COVID-19, as shown by $p = 0.19$. As regards both q and p parameters, Emilia− Romagna test case resembles more Lombardy than Veneto, which should be expected due to the local policy adopted. Finally, Campania, although only relatively hit by the spread, has a value for q similar to Veneto, whereas its value for p is more similar to those of Lombardy and Emilia−Romagna.

5. Conclusions

This chapter has introduced a variant of the SEIAR model for epidemics added with DSD in which the social distancing value is represented as a function varying over time. Moreover, it has described how this model can be coupled with DE for the individuation of its most suitable parameter values. The resulting mechanism has been applied to model the spreading of the coronavirus COVID19 in Italy and in some of its most important regions.

Consequently, for these scenarios, we have been able to approximately evaluate the evolution over time of the daily number of infectious cases. This has allowed us to approximately estimate the day(s) in which this number will be at its acme, the corresponding value of the peak, and the future evolution of the spread.

We hope this study can serve as a useful guideline to Italian Government as well to the Government of any other State in which COVID-19 spreading is occurring.

References

[1] G. Chowell, C.E. Ammon, N.W. Hengartner, J.M. Hyman, Estimating the re- production number from the initial phase of the Spanish flu pandemic waves in Geneva, Switzerland, Math. Biosci. Eng. 4 (3) (2007) 457.

[2] G. Chowell, H. Nishiura, L.M. Bettencourt, Comparative estimation of the repro- duction number for pandemic influenza from daily case notification data, J. R. Soc. Interface 4 (12) (2007) 155−166.

[3] K. Sohn, B.L. Boulier, Estimating parameters of the 1918-19 influenza epidemic on US military bases, J. Appl. Bus. Econ. 13 (4) (2012) 30−42.

[4] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. 115 (772) (1927) 700−721. Series A, containing papers of a mathematical and physical characterontaining papers of a mathematical and physical character.

[5] O. Diekmann, H. Heesterbeek, T. Britton, Mathematical Tools for Understanding Infectious Disease Dynamics, vol. 7, Princeton University Press, 2012.

[6] L. Pribylova, V. Hajnova, SEIAR Model with Asymptomatic Cohort and Con- Sequences to Efficiency of Quarantine Government Measures in Covid-19 Epidemic, 2020 arXiv preprint arXiv:2004.02601.

[7] J. Rocklov, H. Sjodin, A. Wilder-Smith, COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures, J. Trav. Med. 27 (3, taaa030) (February 2020) 1−7.

[8] C. Hubbs, Social Distancing to Slow the Coronavirus. https://towardsdatascience.com/social-distancing-to-slow-the-coronavirus- 768292f04296. (Accessed: 29 March 2020).

[9] K. Price, R.M. Storn, J.A. Lampinen, Differential Evolution: A Practical Approach to Global Optimization, Springer Science & Business Media, 2006.

[10] R.M. Anderson, R.M. May, Infectious Diseases of Humans, Oxford Univ. Press, 1991.

[11] F. Brauer, Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, second ed., Springer, 2012.

[12] Q. Li, X. Guan, P. Wu, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus−infected pneumonia, N. Engl. J. Med. 382 (2020) 1199−1207.

[13] M. Batista, Estimation of the Final Size of the Covid-19 Epidemic, 2020. https://doi.org/10.1101/2020.02.16.20023606 medRxiv.

[14] X.S. Wang, J.H. Wu, Y. Yang, Richards model revisited: validation by and application to infection dynamics, J. Theor. Biol. 313 (2012) 12−19.

[15] F. Brauer, Mathematical epidemiology: past, present, and future, Infect. Dis. Modell. 2 (2) (2017) 113−127.

[16] B. Pell, Y. Kuang, C. Viboud, G. Chowell, Using phenomenological models for forecasting the 2015 ebola challenge, Epidemics 22 (2018) 62−70.

[17] H. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599−653.

[18] Y.C. Chen, P.E. Lu, C.S. Chang, T.H. Liu, A Time-dependent SIR Model for Covid-19 with Undetectable Infected Persons, 2020 arXiv:2003.00122.

[19] R. Li, S. Pei, B. Chen, et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2), Science (2020). https://doi.org/10.1126/sci-ence.abb3221.

[20] E.L. Ionides, C. Breto, A.A. King, Inference for nonlinear dynamical systems, Proc. Natl. Acad. Sci. U. S.A. 103 (49) (2006) 18438−18443. https://doi.org/10.1073/pnas.0603181103.

[21] A.A. King, E.L. Ionides, M. Pascal, M.J. Bouma, Inapparent infections and cholera dynamics, Nature 454 (2008) 877−880.

[22] K. Prem, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, M. Jit, P. Klepac, The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in Wuhan, China: a modelling study, Lancet (2020). https://doi.org/10.1016/S2468-2667(20)30073-6.

[23] L. Lòpez, X. Rodò, A Modified SEIR Model to Predict the Covid-19 Outbreak in Spain and Italy: Simulating Control Scenarios and Multi-Scale Epidemics, 2020. https://doi.org/10.1101/2020.03.27.20045005 medRxiv.

[24] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic Analysis of Covid-19 in China by Dynamical Modeling, 2020 arXiv preprint arXiv:2002.06563.

[25] Q. Lin, S. Zhao, D. Gao, et al., A conceptual model for the coronavirus disease 2019 (covid-19) outbreak in Wuhan, China with individual reaction and govern- mental action, Int. J. Infect. Dis. 93 (2020) 211−216.

[26] Z. Yang, Z. Zeng, K. Wang, et al., A modified SEIR and AI prediction of the epidemics trend of covid-19 in China under public health interventions, J. Thorac. Dis. 12 (3) (2020).

[27] S. Mwalili, M. Kimanthi, V. Ojiambo, et al., SEIR Model for Covid-19 Dynam- Ics Incorporating the Environment and Social Distancing, 2020. https://doi.org/10.21203/rs.3.rs-19249/v1.

[28] W. Lyra, J.D. do Nascimento, J. Belkhiria, et al., Covid-19 Pan- Demics Modeling with SEIR(+CAQH), Social Distancing, and Age Stratifi- Cation. The Effect of Vertical Confinement and Release in Brazil, 2020. https://doi.org/10.1101/2020.04.09.20060053.

[29] G. Giordano, F. Blanchini, R. Bruno, et al., Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy, Nat. Med. (2020). https://doi.org/10.1038/s41591-020-0883-7.

[30] Italian Ministry for Health: Covid−19 - Italia - monitoraggio - situ- azione. https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc- covid19-ita-regioni.csv. (Accessed March 29 2020).

[31] J. Zhang, A.C. Sanderson, JADE: adaptive differential evolution with optional external archive, IEEE Trans. Evol. Comput. 13 (5) (2009) 945−958.