Article

SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence

Alberto Godio[1*], Francesca Pace[1] and Andrea Vergnano[1]

[1] Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; francesca.pace@polito.it (F.P.); andrea.vergnano@polito.it (A.V.)

* Correspondence: alberto.godio@polito.it (A.G.); Tel. +39 011 0907656

Abstract: We applied a generalized SEIR epidemiological model to the recent SARS-CoV-2 outbreak in the world, with a focus on Italy and its Lombardia, Piemonte, and Veneto regions. We focus on the application of a stochastic approach in fitting the model numerous parameters using a Particle Swarm Optimization (PSO) solver, to improve the reliability of predictions in the medium term (30 days). We analyze the official data and the predicted evolution of the epidemic in the Italian regions, and we compare the results with data and predictions of Spain and South Korea. We link the model equations to the changes in people's mobility, with reference to Google's COVID-19 Community Mobility Reports. We discuss the effectiveness of policies taken by different regions and countries and how they have an impact on past and future infection scenarios.

Keywords: SARS-CoV-2; COVID-19; SEIR modeling; Italy; stochastic modeling; swarm intelligence. Google COVID 19 Community Mobility Reports.

1. Introduction

We intend to present an updated version of the predictive model of epidemic phenomena based on the approach called SEIR (Susceptible-Exposed-Infective-Recovered), widely used to analyze infection data during the different stages of an epidemic outbreak; the model is also suitable to predict possible contagion development scenarios. It is based on a series of dynamic ordinary differential equations that consider the amount of the population subject to contagion, the trend over time of individuals who recover after infection, and the individuals who unfortunately die [1].

The main objective is to furnish a reliable tool for modeling and predicting possible epidemic evolution, based on the data given in official repositories about infective individuals, recovered, and deaths. We also explore the main drawbacks of the suggested method, mostly related to the uncertainty of the input data and the complexity in the inclusion into the model of all the external factors (population density and ages, previous diseases, efficiency of the public health system…) which have impacts on the virus diffusion, the recovery rate and the number of deceases.

The SEIR model is not a novelty in the modeling and forecasting of epidemic phenomena. The main innovation we have introduced concerns the stochastic approach to solve the model and to assess the propagation of the uncertainties of the model solution. We applied it to the Italian situation at a national level, and at a regional scale, focusing on the most impacted regions in Northern Italy. We also tested the approach to other situations such as Spain, which offers many similarities in the virus diffusion with the Italian scenario, and South Korea, that is instead characterized by a completely different scenario.

This work is carried out during the crucial development phase of the epidemic in Italy (mid-April 2020), with the operational difficulties linked to the impossibility of verifying and validating the databases, and with the difficulty of comparing and calibrating the results with other studies. The purpose, however, is to provide an easy to read and useful tool that can help the policymakers, responsible for strategic choices, in assessing the social and economic scenarios related to the

development of the epidemic. We are conscious that it is a predictive model which, although based on a scientific approach, is conditioned by a series of intrinsic and endogenous factors that can affect its medium-term reliability, but we aware that any political decision not based on the rational and critical evaluation of all available data risks being based on mere sensations, often dictated by sentimental suggestions [2].

The generalized SEIR model we used is based on the system of differential equations found by Peng et al. (2020) [3] in the analysis of the SARS-CoV-2 outbreak in China. We chose this model, that adds complexity to the classical SIR or SEIR models, for being able to represent the various conditions of susceptible and infected individuals during an epidemic outbreak (especially quarantined people, who are not able to infect other people during their quarantine). In the SEIR methods, the coefficients of the equations represent the different ratios of variation over time of the different categories of individuals, that is, infected, dead, and recovered [4]. These coefficients have been often considered constant [5] so that they are unable to take due account of external influences, such as the actions containing the spread of the infection, that may occur at different times during the development of the infection itself, or the possible change in health conditions of infected individuals due to pharmacological development. Like all models, the quality of the observed data and their validation is a crucial node in assessing the reliability of the model results.

The approach herein discussed introduces time-dependent model parameters. In particular, we made the infection ratio time-dependent, assuming that the number of contacts between people, during the lockdown, decreases proportionally to the decrease of their overall mobility, calculated using a big-data repository made available by Google [6].

Besides, we introduce a stochastic approach to solve the generalized SEIR model. This approach is based on a metaheuristic method, the Particle Swarm Optimization (PSO) algorithm, belonging to the family of computational swarm intelligence [7]. In fact, citing the overview made by Parham (2012) [1], the analysis of temporally-forced non-linear epidemic models within stochastic frameworks has received little attention to date due to the complexity of the problem, despite representing the most realistic framework for capturing the behavior of many intrinsically or extrinsically forced infectious diseases. With respect to the standard deterministic approach in solving the SEIR model, the advantage of the PSO approach is that the adaptive exploration of the space domain of the solutions decreases the risk of being trapped into a local minimum and iteratively searches for the global minimum as the final solution. Moreover, the PSO method could provide a set of model solutions as probable scenarios calculated by means of a-posteriori probability density distribution.

2. Materials and Methods

2.1. Database

The analysis is based on the data collected and made available via a dashboard by the John Hopkins University in the USA. They represent an official database as they collect the data from different official organizations such as World Health Organization (WHO), European Centre for Disease Prevention and Control (ECDC), USA Centers for Disease Control and Prevention, and other organizations. The Italian data are entirely collected through the bulletin of the *Protezione Civile Italiana*. The Italian National Institute of Statistics (ISTAT) was the source for the number of national and regional populations. The number of people living in the studied regions is reported in Table 1.

Table 1: population (approximated) for Italy and Italian regions and for other countries included in the following analysis

Countries / Regions	Overall Population	Database (Year)
Italy	60,359,546	Istituto Nazionale di Statistica - ISTAT (2019)
Lombardia	10,060,574	Istituto Nazionale di Statistica - ISTAT (2019)
Veneto	4,905,854	Istituto Nazionale di Statistica - ISTAT (2019)

Peer-reviewed version available at International Journal of Environmental Research and Public Health 2020, 17; doi:10.3390/ijerph17103535

Piemonte	4,356,406	Istituto Nazionale di Statistica - ISTAT (2019)
Spain	47,100,396	Istituto Nacional de Estadìstìca – INE (2019)
South Korea	51,629,512	Korean Statistical Information Service – KOSIS (Nov. 2018)

2.2. Overview of the generalized SEIR model

The SEIR model simulates the time-histories of an epidemic phenomenon. In its classical form, it models the mutual and dynamic interaction of people between four different conditions, the susceptible (S), exposed (E), infective (I), and recovered (R).

The classical SEIR model can be described by a series of ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta I(t) \cdot \frac{S(t)}{N} \quad 1)$$
$$\frac{dE(t)}{dt} = \beta I(t) \cdot \frac{S(t)}{N} - \gamma E(t)$$
$$\frac{dI(t)}{dt} = \gamma E(t) - (\lambda + \kappa) I(t)$$
$$\frac{dR(t)}{dt} = (\lambda + \kappa) I(t)$$

The susceptible (S) is the part of the population that could be potentially subjected to the infection: at the initial time, without further information, it is represented by the whole population. The exposed (E) is the fraction of the population that have been infected but does not show symptoms yet: it can be called a latent phase, and at this stage, a disease can be infectious, partially infectious or not infectious [8]. The infective (I) represents the infective population after the latent period. The removed (R) are the people after healing, and they are generally not reintroduced into the susceptible category if it is supposed that they became immune to the disease. This aspect is strongly debated in this epidemic, as in some countries the second infection of recovered people has been recorded. At this stage of our study, we don't have enough data to include this effect on the model; this would require the introduction of another term in the previous system of equation, including another coefficient that takes into account the re-population of the Susceptible compartment. The R category comprehends also the individuals who died of the disease. A characteristic of this model is that the sum of the four categories is equal to the total population (N) at any time:

$$S(t) + E(t) + I(t) + R(t) = N \quad 2)$$

As can be seen, it does not consider the natural births and deaths of the population during the time span of the disease.

The equations of the classical SEIR model are governed by the parameters β, γ, λ and κ. We adopt the symbology used in Peng et al. (2020) [3]. As usual in this field, the following parameters have day^{-1} as a unit of measurement.

- β is called infection ratio. It is the number of people that an infective person infects each day. It is equal to $b \cdot p$, where b, or contact rate, is the number of people an average person enters in contact with each day, and p is the probability that a contact provokes the transmission of the disease. In the SEIR model, β is the vector which transports people from the S category to the E category. It is multiplied by the ratio S/N to avoid counting contacts between two people who cannot infect each other (e.g. because one of them has already recovered, or because both are infective)
- γ is the inverse of the average latent time and governs the lag between having undergone an infectious contact and showing symptoms: in the equations, it brings people from the E category to the I category.
- λ and κ are the recovery rate and the death rate, respectively, and they are united together in a single parameter in the classical SEIR model. They give information about how fast the people may recover from the disease (1/λ is the average recovery time), and how many of them, unfortunately, die.

Peer-reviewed version available at International Journal of Environmental Research and Public Health 2020, 17; doi:10.3390/ijerph17103535

Given the complexity of the disease, many authors have implemented different variations of the classical SEIR model, both in the equations and in the parameters, and managing different fitting techniques to make the model represent the reality the closest possible.

We used a generalized SEIR model following the idea of a recent publication by Peng et al. (2020) [3], who studied the COVID-19 infection in some China provinces, and we applied it to the Italian situation. It is described by the following system of equations:

$$\frac{dS(t)}{dt} = -\beta(t)I(t)\cdot\frac{S(t)}{N} - \alpha S(t) \quad (3)$$

$$\frac{dP(t)}{dt} = \alpha S(t)$$

$$\frac{dE(t)}{dt} = \beta(t)I(t)\cdot\frac{S(t)}{N} - \gamma E(t)$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t)$$

$$\frac{dQ(t)}{dt} = \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t)$$

$$\frac{dR(t)}{dt} = \lambda(t)Q(t)$$

$$\frac{dD(t)}{dt} = \kappa(t)Q(t)$$

It adds these features in addition to those, previously described, of the classical SEIR model:

- It supposes that the susceptible population decreases thanks to lockdown policies and improvements in public health behaviors, as wearing face masks. Each day, a number of individuals equal to $S\cdot\alpha$ pass from the susceptible category to the protected category (P), where α is the protection rate.
- It adds the category of quarantined people (Q). The passage from the infective to the quarantined category is done through the parameter δ, which is the inverse of the average time required to quarantine a person with symptoms: this happens usually after the person has been tested positive. The quarantined people are so removed from the infective category (I) because they are supposed to not have any contact with others. The quarantined category matches with the "active confirmed cases" in Italian official datasheets, and, according to the common habit of quarantining positive people, this is true also for data from most developed countries. This is a critical point of the system of equations, that should be better defined according to us. In fact, not every single person infected is automatically quarantined, because often the authorities were not able to test enough people to keep pace with the spread. This is especially difficult because many people do not develop symptoms at all, but can transmit the infection to others. So, we think that δ contains also some information about the percentage of infectives detected. A study from Calafiore et al. (2020) [5] proposed the introduction of an additional parameter to better understand this issue.
- It separates the categories of recovered (R) and dead (D) people, linked to the quarantined category through the λ and κ parameters, the cure rate and the mortality rate respectively. λ and κ are set to be time-dependent, because the health system can improve its capability to treat people over time, e.g. with the introduction of new drugs. On the basis of the data collected from Chinese reports by Peng et al. (2020) [3], which suggested an exponential evolution of the two parameters, we constrained them to fit an exponential trend, similarly to Cheynet (2020) [9], who affirms that the idea behind these functions is that the death rate should become closer to zero as time increases while the recovery rate converges toward a constant value:

  $\lambda(t) = \lambda_0[1 - \exp(-\lambda_1 t)]$ (4)

  $\kappa(t) = \kappa_0 \exp(-\kappa_1 t)$ (5)
- The λ_0 represents the final asymptotic value of the cure rate. It is related to the health system ability to tackle the infection after adapting to the new outbreak and depends also on other factors like the good health of the citizens. λ_1 is related to how fast the adaptation to the

Peer-reviewed version available at International Journal of Environmental Research and Public Health 2020, 17; doi:10.3390/ijerph17103535

emergency was. The equation of λ was not able to fit well the data of South Korea, which at the beginning of April was already in a post-peak phase of the disease spread, probably because of the more complicated trend, compared to other countries. Therefore, only for the Korean model, we decided to allow the λ to not follow the constrain of an exponential law. We increased the degrees of freedom of its trend by imposing a sinusoidal law:

$$a \cdot \sin(bt + c) + d \cdot \sin(et + f) \quad (6)$$

fitting the six parameters to the equation $\lambda = dR/(dt \cdot Q)$, taken rearranging one of the model equations, where R and Q are the data series of recovered and quarantined people from official data. This approach proved successful to provide an overall better fit of the model to the data.

- The parameter κ_0 represents the initial value of the mortality rate. κ_0 is related to the initial health system ability to tackle the infection and depends on the good health of the citizens. κ_1 measures how the rate has changed with time. The mortality rate is supposed to decrease over time, and the higher κ_1 is, the faster this decrease.
- We introduced an important change to the β parameter, compared to Peng's model, that is its time dependency. Because the infection rate is proportional to the contact rate β, as stated before, we supposed that the use of big data to estimate the variation in the contact rate was an interesting idea. We took advantage of the recent publication of Google's COVID-19 Community Mobility Reports [6], a database built on GPS data collected from mobile devices with the "Location History" option turned on. It provides data about the reduction in the mobility of people in recent months. For each investigated region, we calculated the average mobility decrease over time, and we fitted the curve with a second-order polynomial trend line. Then, we constrained β to be proportional to that specific trend line. In a preliminary test on a simple SEIR model, we noticed how the introduction of this constrain allowed the model to obtain clearly better fitting results. Therefore, we applied this idea to our SEIR model to ensure good data fitting for all the regions and countries investigated. Interestingly, we noticed that the mobility data of South Korea do not show a significant decrease in people's mobility, because the country adopted a strongly different approach in lockdown policies compared to most European countries. Strict lockdowns were not imposed, but efforts were addressed to track the infection spread at the early stages, with tight controls and strict quarantine protocols for infected individuals. For the 30 days of model prediction, the time-dependent β parameter had the same value of the last observed day (i.e., mid-April) since we do not have reliable predictions about future mobility. This means that the lockdown policies will continue or the reopening of business will be made with particular attention to health protection procedures.

Summarizing, the expected evolution of the equations is the following: the susceptible category decreases over time, feeding the Exposed (E) category through the beta parameter β and the Protected (P) category through the α parameter; the latter represents the part of the population which for various reasons becomes insusceptible to the disease. The Exposed (E) category is only a temporary category: its individuals pass into the infective (I) category after a latent time ($1/\gamma$), on average. The infective category generates newly infected people over time, removing them from the Susceptible category. The detected infected individuals are quarantined (Q) to avoid spread. Then, they evolve into recovered (R) or death (D) cases, according to various causes, like health care system effectiveness, age, co-presence of other diseases. It is important to keep in mind that the most reliable data series provided by national agencies are Q(t), R(t), and D(t). The fitting of these data in the structure of the generalized SEIR model allows us to estimate the trend of other categories with some degrees of uncertainty, as well as the prediction for the subsequent 30 days.

The main outputs of the model are the following data series:

- S, target time-histories of the susceptible cases,
- E, the target time-histories of the exposed cases,
- I the target time-histories of the infective cases,

- Q, the target time-histories of the quarantined cases,
- R, the target time-histories of the recovered cases,
- D, the target time-histories of the death cases,
- P, the target time-histories of the insusceptible cases.

The α, β, γ, δ, λ and κ parameters can be considered also an important output of the model. In particular, the evolution over time of λ and κ could provide information about the changes in the health system response to the contagion. β is also time-dependent, and it is constrained to be proportional to the people's mobility trend extrapolated from Google's big data. α is the other parameter related to policies, although compared to β is not closely related to a precise aspect of the government strategies. α was not forced to follow any particular law during the modeling since we tested no significant improvement by modifying it from constant to a time-dependent parameter. The generalized SEIR-model scheme is graphically described in Figure 1.

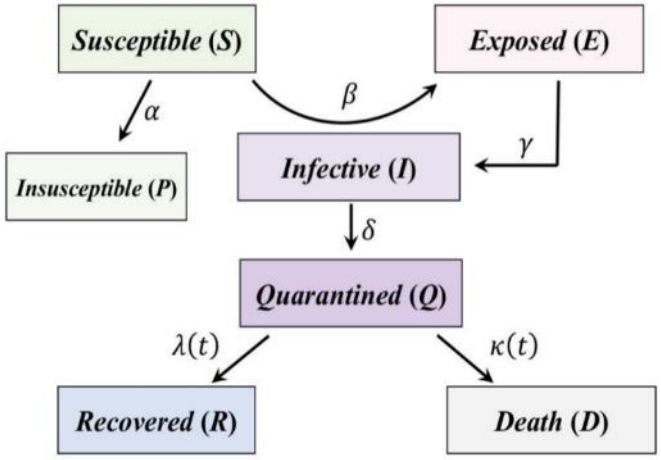


Figure 1. Generalized SEIR model scheme (modified from [5]).

The various researches about SEIR-like models applied to the current SARS_CoV-2 epidemic make minor or major changes to the classical SEIR model. It can be useful to overview the values of the main parameters, to have an estimate of an acceptable range of values, and to better understand their meaning. We used these values to set the lower and upper boundaries of the parameters, to constrain the PSO solution space within a realistic range. It must be remembered that the different values of these parameters are due also to different methods, so the following bibliographic research, summarized in table 2, should be considered qualitatively. We added some notes to highlight some peculiarities of the various studies.

Table 2: values of α, β, γ, δ, λ, and κ describing SARS-CoV-2 outbreak in the recent literature.

Authors	Country/region	Date	α	β	γ	δ	λ	κ	Notes
Peng et al. (2020) []	China without Hubei province	20 Jan-9 Feb	0.172	1	0.5	0.15	0.005-0.04	0.005-0.015	
Peng et al. (2020) [3]	Hubei province without Wuhan city	20 Jan-9 Feb	0.133	1	0.5	0.139	0.005-0.015	0.005-0.02	
Peng et al. (2020)	Wuhan	20 Jan-9 Feb	0.085	1	0.5	0.135	0.005-0.015	0.005-0.03	

Peng et al. (2020)	Beijing	20 Jan-9 Feb	0.175	0.99	0.5	0.175	0.005-0.04	0.002	
Peng et al. (2020)	Shanghai	20 Jan-9 Feb	0.183	1	0.5	0.179	0.005-0.04	0	
Calafiore et al. (2020) [5]	Italy	23 Feb-30 Mar		0.22			0.017	0.012	1)
WHO report [10]	China	12 Feb		≈0.1-0.2					2)
Dandekar et al. (2020) [10]	Wuhan	- 1 Apr		1			0.023		3)
Dandekar et al. (2020)	Italy	- 1 Apr		0.74			0.032		
Dandekar et al. (2020)	South Korea	- 1 Apr		0.68			0.004		
Dandekar et al. (2020)	US	- 1 Apr		0.69			0.008		
Shaikh et al. (2020) [11]	India	14-26 Mar		0.59			0.1		4)
Lin et al. (2020) [12]	Wuhan	15 Jan-24 Feb		0.59-1.68	0.33		0.2		5)
Iwata et al. (2020) [13]	-			0.1-1	0.07-0.5		0.1-1		6)

1. This study presents the values of the parameters also for each Italian region. The peculiar SEIR model used here introduces two new parameters: α (with a meaning different from ours) and ω. α represents how many times the real number of infected people is higher than the number of detected infected people, and it is estimated to be about 63. ω is a fixed percentage of total people susceptible to the disease, and according to this model, it is about 0.124 (12%).
2. This report, on page 52, shows a table with a list of estimates of Serial Interval, which is the average time between infection and subsequent transmission. We reported the latter since it is closely related to the inverse of β. It ranges from 5 to 9 days, according to the various studies considered.
3. Interestingly, this paper calculates the change over time of a term, called quarantine strength (Q(t)), fitted thanks to a neural network-based approach.
4. This study considers separately the Asymptomatic and Symptomatic categories.
5. The most interesting feature of this model is the time-dependent β, implemented with a different design compared to our model: $\beta(t) = \beta_0(1-\alpha)\left(1-\frac{D}{N}\right)^k$ where α is a stepwise function that represents the governmental action, estimated to range from 0 to 0.85 according to the strength of lockdown policies. while k is the citizen response, estimated to be about 1100.
6. The authors propose a model which does not fit real data but investigates about possible scenarios deriving from different combination of parameters. Their range is reported as a credible range reference.

2.3. Implementation of the SEIR model with a stochastic approach

The model equations were implemented using the MATLAB code provided by [9] available in Matlab File Exchange (https://it.mathworks.com/matlabcentral/fileexchange/74545-generalized-seir-

epidemic-model-fitting-and-computation). The data are extracted from the official repository and are composed of: confirmed, recovered and death cases (Q, R, and D, respectively). These values represent the initial assumptions. The parameters α, β, γ, δ, λ, and κ are the problem unknowns. The differential equations are numerically solved by means of the Runge-Kutta method.
The standard approach of the source code used as default a least-square fitting solver to match observed data and calculated response (of Q, R, and D). At the beginning of modeling, the initial values of the six parameters are given as first esteem. Then, their values are calculated following a least-square solver that considers the observed data (Q, R, D) with time.

We modified the standard release of the code by introducing a new solver, the PSO algorithm, belonging to the family of computational swarm intelligence (population-based nature-inspired metaheuristics) [14,15]. This optimization solver minimizes an objective function, which is set to decrease the misfit between observed data and calculated responses of Q, R, and D by varying the six parameters, i.e., the problem unknowns. The stochastic approach has a number of advantages over the deterministic method to solve the SEIR model. The adaptive exploration and exploitation of the search space of the model solutions avoid the risk of being trapped in some local-minima solutions and also enhances the independence from the initial assumption of the six parameters which could bias the final solution. The solution search-space is sampled by a set of 200 particles, representing the possible solutions, which are randomly initialized. The adaptive behavior and the convergence and stability of the final solution are ensured by using a PSO variant, the hierarchical PSO with time-varying acceleration coefficients (HPSO-TVAC) [16]. Convergence was achieved in 150 iterations. Each run of 150 iterations was repeated for 50 trials in order to test the variability of the solutions due to the random initialization of the parameters. Finally, the trial showing the minimum normalized root mean square error (NRMSE) was selected as the best solution. The solutions from the remaining trials were a-posteriori evaluated with their probability density distribution. The solutions within 5% of the minimum NRMSE were chosen as representative of other probable scenarios.

Deploying a stochastic approach increased the computational cost of the modeling. Therefore, the code was parallelized to be run on multiple cores. The simulations ran on the academic High Performance Computing (HPC) cluster of Politecnico di Torino. The sustained performance of the cluster is globally 20.13 TFLOPS and the CPU model of one node is 2x Intel Xeon E5-2680 v3 2.50 GHz 12 cores. We adopted 24 cores of a single node.

3. Results

Here we present the time series obtained by the standard deterministic approach, compared with the data series that highlight the best prediction solutions obtained by the stochastic approach, based on the Particle Swarm Optimization (PSO) algorithm. First, we analyze the Italian framework at a national and regional scale. Then, we provide the results of SEIR modeling for two other countries: Spain and South Korea. While Spain was chosen because the epidemic spread has been similar to the Italian one, South Korea represented a testing data set as the epidemic peak has been already overcome. The final NMRSE of the modeling and the values of the SEIR coefficients are supplied in Table 3.

The prediction of the Italian situation according to deterministic solver shows the trends given in Figure 2. We can point out that the results are biased by the selection of the starting point of model parameters, according to the main limits of the deterministic methods. The result of the PSO approach is shown in Figure 3. Observed data of quarantined, recovered and death cases are marked in red, green, and black circles, respectively. Data in red shows the number of individuals tested positive and placed in quarantine (at home, or hospitalized, or in intensive care). It should be pointed out that the sum of Quarantined, Recovered and Deaths, at a certain date, represents the total confirmed cases at that moment. The Italian data set starts from 1st March because we chose to begin the modeling from the day when the confirmed cases are 1% of the maximum counted cases. The predicted curves are plotted with solid lines in Figures 2 and 3. The set of most probable PSO solutions (within 5% of

the minimum NMRSE) is plotted with dashed lines in Figure 3. The predicted peak in the red curve represents the status in which the rate of recoveries becomes greater than the rate of infection. It is important because reflects also one of the moments of greatest stress on the health system, because the quarantined people, both at home and in the hospitals, are at their maximum. For the Italian situation, the maximum number of the predicted quarantined cases is expected after 27th April according to the deterministic approach and before that day according to the best solution of the PSO approach. The curves of recovered and deceased cases in Figures 2 and 3 are fairly similar. The final NRMSE was 0.035 and 0.043 for PSO and deterministic modeling, respectively (Table 3).

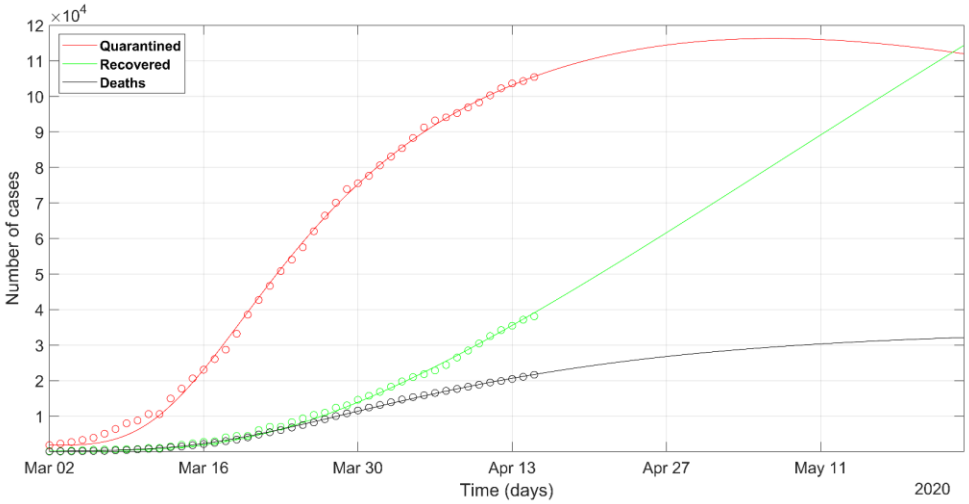


Figure 2. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for Italy, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with the deterministic approach.

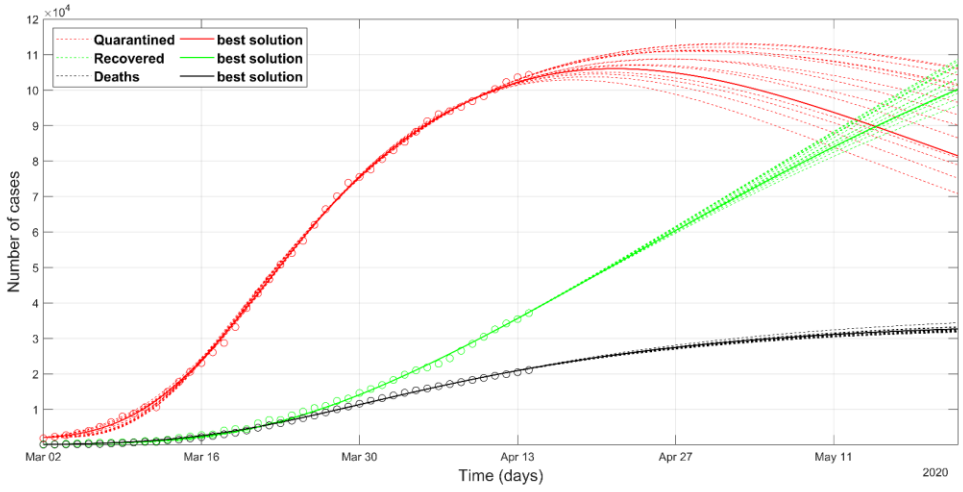
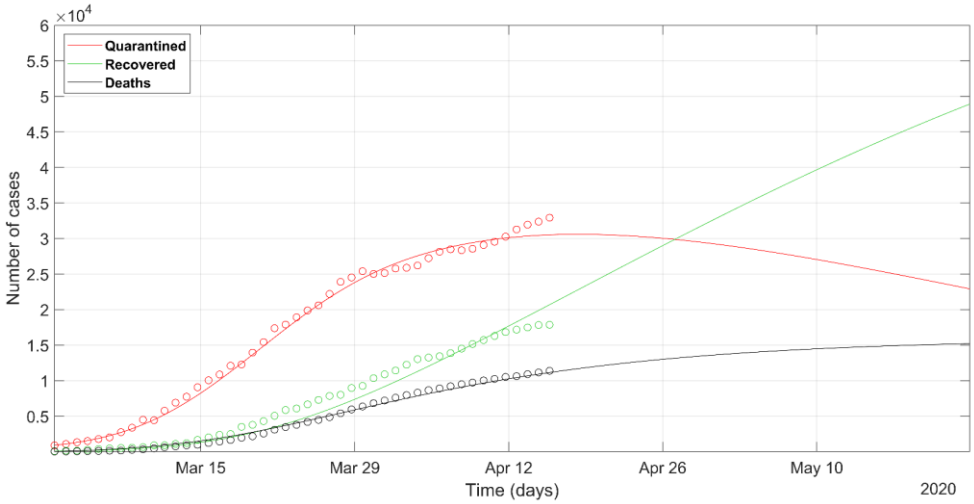


Figure 3. Observed data (circles) of quarantined (red), recovered (green) and deaths (black) for Italy, updated on April 15th, 2020 The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with the stochastic approach. The solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

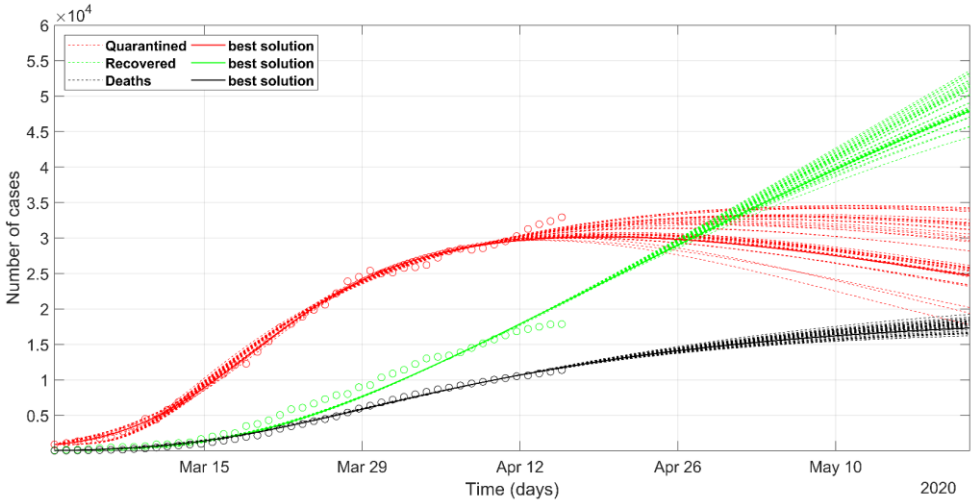
The analysis of the situation of Lombardia, Veneto, and Piemonte regions is depicted in Figures 4, 5, and 6, respectively. Lombardia has been strongly impacted by SARS-CoV-2, as at the end of March, nearly 40,000 novel infected cases and more than 5,000 deaths have been recorded in a

population of 10 million. On the contrary, the Veneto region evidences 7,000 cases and about 300 deaths in a population of 5 million people.

The prediction of the situation in Lombardia, according to the deterministic approach (Figure 4a), appears rather optimistic, as the trend of the quarantined should start to decrease in few days (red solid line). This probably does not reflect the evolution of the true situation in that region, even if the rate the recovered suggests positive feelings. If we look at the most probable scenarios predicted according to the PSO analysis (dashed lines in Figure 4b), the wide spreading of the trend of the quarantined indicates how any eventual less-restrictive policy must be evaluated with great care in the next days. The set of most probable solutions from PSO presents a wide range of solutions, wider than that for Italy (Figure 3). The final NRMSE was 0.062 and 0.061 for PSO and deterministic modeling, respectively (Table 3).



(a)



(b)

Figure 4. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for the Lombardia region, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with **(a)** deterministic approach, **(b)** stochastic approach. In **(b)** the solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

Figure 5a and b show the SEIR model prediction for the Veneto region according to the deterministic and PSO approaches, respectively. While the predicted recovered and death cases are in accordance, the curves of quarantined cases present a slightly different location for the predicted

peak, which is comprised between 11th and 25th April. The final NRMSE was 0.035 and 0.04 for PSO and deterministic modeling, respectively (Table 3).

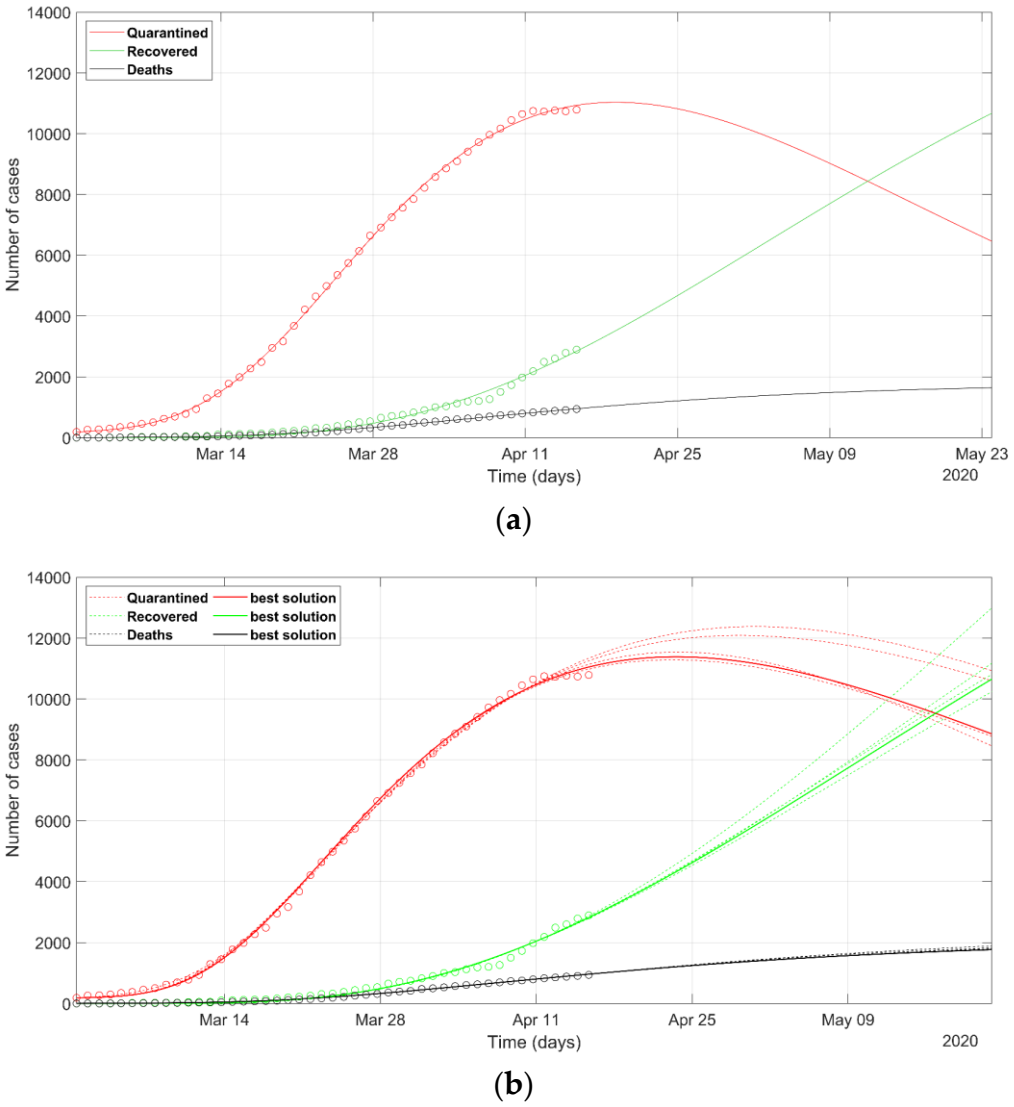


Figure 5. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for Veneto region, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with **(a)** deterministic approach, **(b)** stochastic approach. In **(b)** the solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

The SEIR modeling for the Piemonte region is shown in Figure 6a and b after deterministic and PSO predictions, respectively. The scenarios predicted from PSO are a little worse than those of the deterministic solutions. However, the observed data of Piemonte yield a wide range of probable solutions (dashed lines), which can be overlapped to the determinist solution in some cases. The final NRMSE was 0.056 and 0.05 for PSO and deterministic modeling, respectively (Table 3).

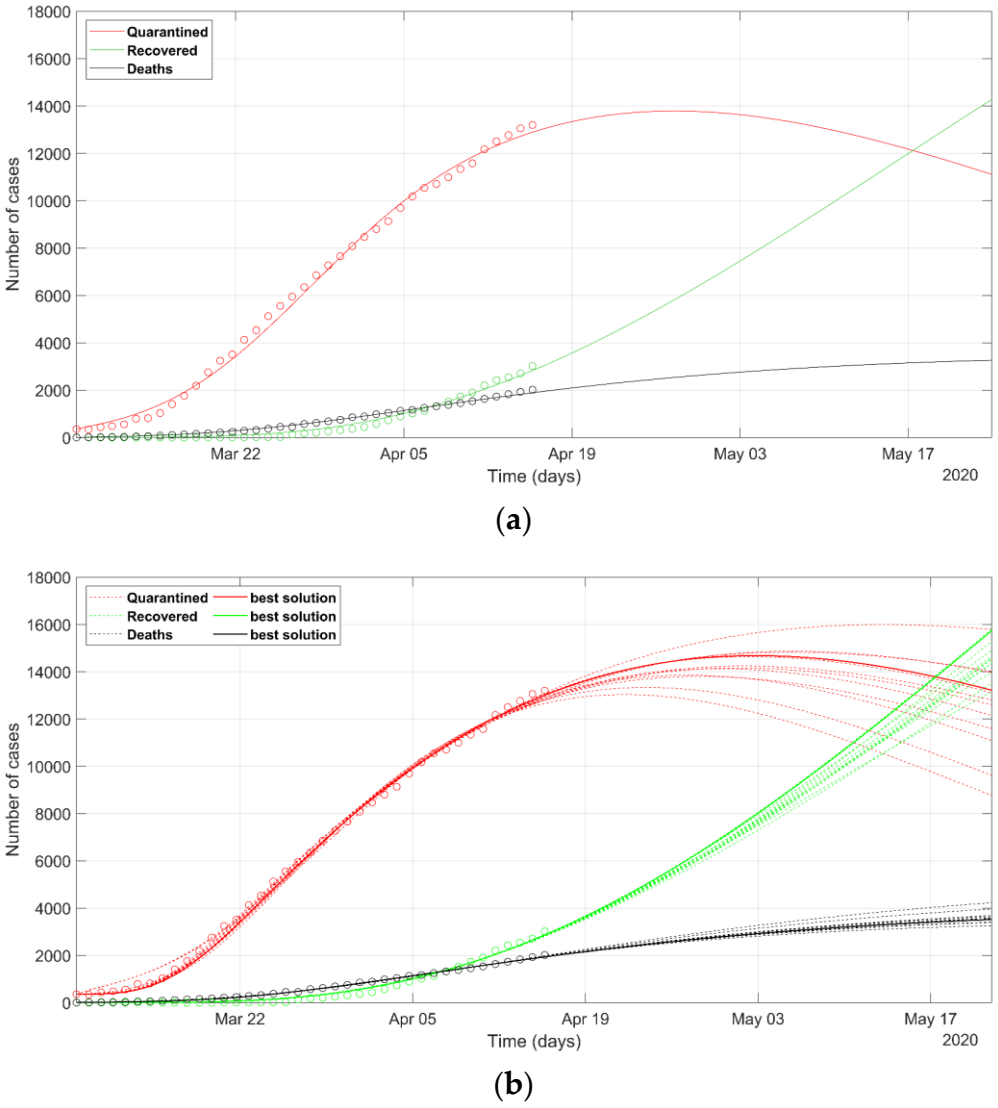


Figure 6. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for the Piemonte region, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with **(a)** deterministic approach, **(b)** stochastic approach. In **(b)** the solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

The picture of the situation in Spain is shown in Figure 7. The crisis exploded in a few days after the Italian collapse, as the direct consequence of the delay in undertaking restrictions in the business and social activities to limit the spreading of the infection. At this stage of the evolution of the phenomenon in Spain, after one month, the results obtained by the deterministic approach forecast a trend of the recovered that seems very optimistic if compared with respect to the Italian situation. We can assume that the Spanish health system will react promptly to the last evolution of the infectious. The result of PSO modeling is shown in Figure 7b. The final NRMSE was 0.046 and 0.052 for PSO and deterministic modeling, respectively (Table 3).

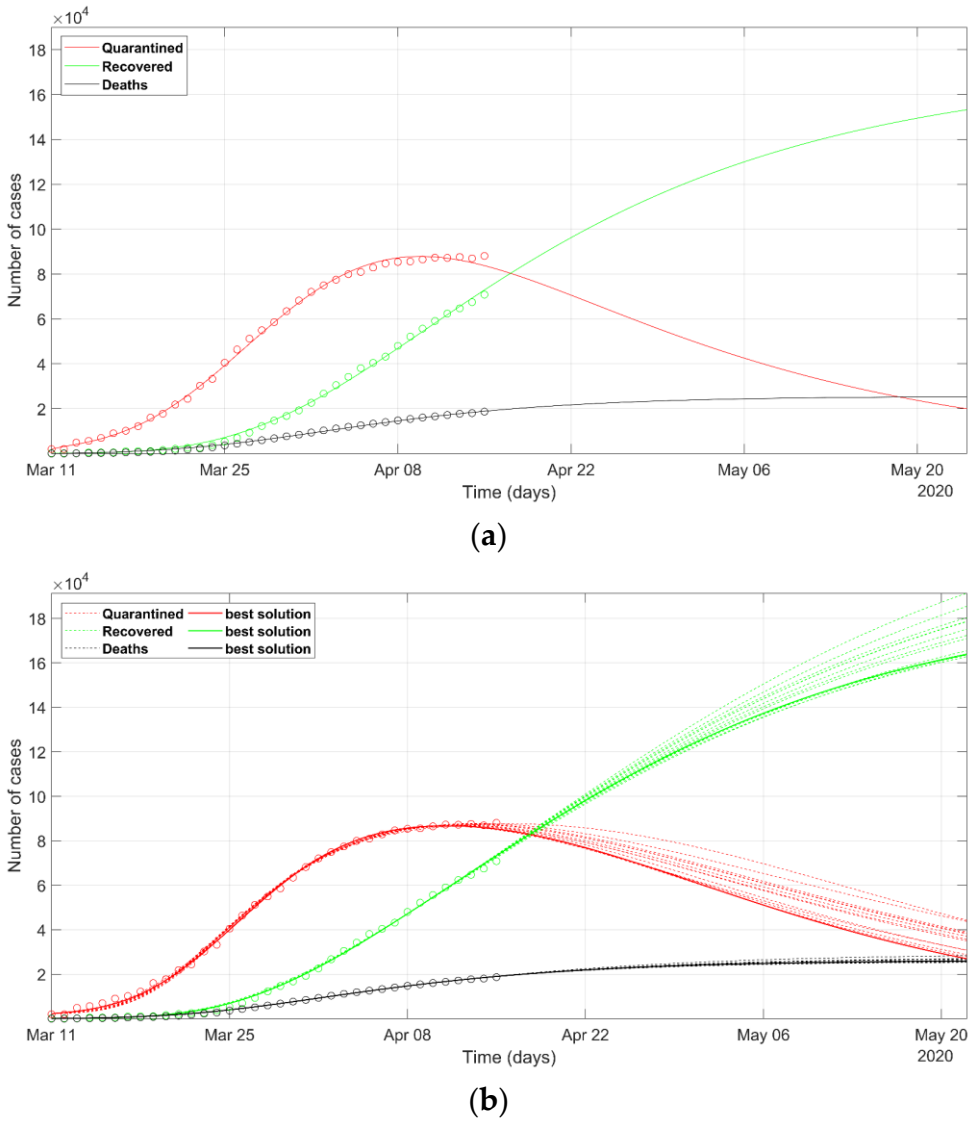


Figure 7. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for Spain, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with **(a)** deterministic approach, **(b)** stochastic approach. In **(b)** the solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

The trend of the cases and the predicted response of South Korea's situation is presented in Figure 8. The analysis of the data about South Korea could be useful in order to look at the Italian situation with respect to a country where, for many reasons, the contagious has been limited, even if the crisis seemed very dramatic at the early stage. The abrupt changes of the recovered and quarantined trend required a careful analysis of the SEIR coefficients and their temporal variation. The final data fitting was indeed not ideal because of the marked oscillations in both the time-series of quarantined and recovered cases. The final NRMSE was 0.074 and 0.078 for PSO and deterministic modeling, respectively (Table 3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 May 2020 doi:10.20944/preprints202004.0073.v2

Peer-reviewed version available at International Journal of Environmental Research and Public Health 2020, 17; doi:10.3390/ijerph17103535

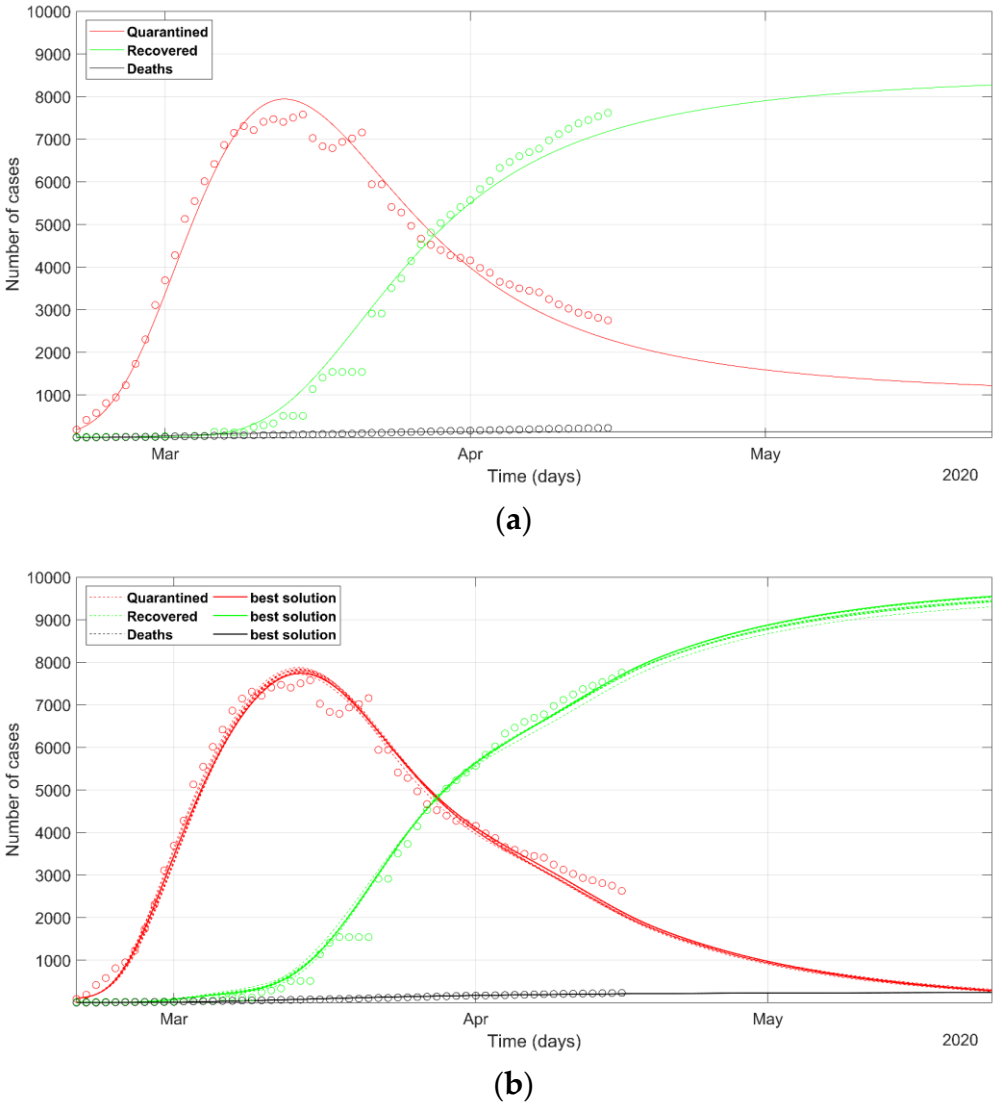


Figure 8. Observed data (circles) of quarantined (red), recovered (green), and deaths (black) for South Korea, updated on April 15th, 2020. The continuous lines refer to the predicted evolution in 30 days according to the SEIR model solved with **(a)** deterministic approach, **(b)** stochastic approach. In **(b)** the solid line refers to the best PSO solution, the dashed lines refer to the most probable solutions (i.e., the solutions within 5% of the minimum NRMSE).

4. Discussion

We adopted a generalized SEIR model to offer a quantitative overview of the complex analysis of the SARS-CoV-2 epidemic meanwhile the disease is still running. The parameters were fitted in a least-square sense with a deterministic approach, and then with a stochastic approach, using a Particle Swarm Optimization (PSO) algorithm, a novelty in the field of epidemiological studies.

The analysis of the results given by the simulation using the stochastic approach gives an overview of the most probable scenarios selected among the solutions within 5% of the normalized root mean square (NRMSE) of the best solution. For each investigated area, we performed 50 trials of PSO simulations and from 5 to 15 trials belonged to the most probable set. It can be well noticed that the different predicted model responses were obtained with approximately equivalent goodness of the data fitting (normalized with respect to the mean value within an L_2-norm < 0.05). The probable scenarios sometimes presented a wide range of possible solutions because of the intrinsic setting of the stochastic approach. The different scenarios are achieved by a deeper investigation of the model-

space domains where the solutions are not driven and influenced by the initial guess of the SEIR model coefficients, as happens in the deterministic approach.

The data seem to confirm that while Lombardia and Piemonte applied similar approaches to social distancing and retail closures, Veneto's strategy applied a much more proactive effort to limit the contagious, by means of extensive testing of symptomatic and asymptomatic cases early on, jointly with an effective tracing of potential positives. The different actions undertaken by the Regions are well depicted in the future trend of the model, with evident advantages in an earlier end of the infection spreading in Veneto (compare Figures 4 and 6 with 5). In fact, the peak of quarantined in Veneto lies before than those of Lombardia and Piemonte. The descending curve in Veneto has a sharper trend than that of the other two regions. Moreover, in Veneto, the predicted fatalities are ten times lower and the recovered are 5 times lower than those in Lombardia.

The behavior of Piemonte (Figure 6) deals with a peculiar trend, introducing a time-delay of the recovered (green curve) with respect to the death cases (black curve), since the number of the recovered in the month of March is always lower than the deaths. This involved that the intersection of the trends of recovered and death cases was reached after the other regions herein analyzed, i.e., 7th April. This probably resulted from the regional testing policy that tested (and counted as confirmed) only patients with severe symptoms or at high risk. The high rate of death cases in March was also caused by the unexpected stress on the health system and the scarcity of intensive-care units. Differently, Lombardia and Veneto experienced a higher rate of recovered patients at the early stage of the epidemic outbreak.

A recent analysis [17] pointed out how, according to the guidance from public health authorities in the central government, Lombardia's actions involved a more conservative approach mainly focusing on the symptomatic cases. They also suppose that the set of policies enacted in Veneto has minimized the burden on hospitals and minimized the risk of spreading in medical facilities: they tried to prevent the diffusion by capillary actions at the local scale, to limit the contagious with additional measures in the hot spots of the infection at the early stage of the epidemic.

The expected trend of these regions is controlled by many factors outside the control of policymakers, including Lombardia's greater population density and a higher number of cases at the explosion of the crisis. Nevertheless, the different public health policies at the early stage of the epidemic phenomena also had an impact, and it seems that tailored capillary actions, as in the example of Veneto, obtained better results than applying only a regional lockdown. The difference in the approaches can be underlined by observing that many municipalities or provinces have declared "red zones", where, due to high transmission of the infection, additional restrictive measures have been introduced, compared to the rest of the regional territory. In the red zones, the different policies act in response to local epidemiological situations. Instead, in Piedmont and Lombardy, no red zone has been established, but restrictive individual distancing measures are regulated on a regional scale. According to the evidenced results of different policies, in the next phase of governmental policies, the reopening of business and activities should be tailored to the local situations, focusing on the organization and integration of all figures of the health system. In particular, the central government should require from the regions an effort to provide local epidemiological data in real-time, to lockdown only limited areas, while the reopening of regional-scale business can be eased.

The estimated parameters that regulate the equations of the SEIR model are reported in Table 3. For each parameter obtained with stochastic approach models, the best-solution value is shown in bold, with the mean and the variance of the solutions within 5% of the minimum NRMSE in brackets. The parameter calculated from the deterministic approach is written in italic-bold. In Table 3 we can appreciate the comparison of the parameters between different regions, and between the stochastic and deterministic approach. Both approaches provided models that fitted the observed data with good accuracy, although stochastic one has, in general, a slightly lower NRMSE.

Table 3. Coefficients: **PSO best-solution** (mean, variance), ***deterministic solution***

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 May 2020 doi:10.20944/preprints202004.0073.v2

Peer-reviewed version available at International Journal of Environmental Research and Public Health 2020, 17; doi:10.3390/ijerph17103535

Country	α	β^*	γ	δ	λ_0	λ_1	κ_0	κ_1	*NRMSE* ***PSO*** *(determ inistic)*
Italy	**0.021** (0.086, 0.004) ***0.012***	**0.510** (1.058, 0.200) ***1.170***	**0.265** (0.859, 0.226) ***1.065***	**0.103** (0.095, 0.01) ***0.020***	**0.017** (0.017, $6.3{\cdot}10^{-9}$) ***0.017***	**2** (1.696, 0.180) ***1.983***	**0.029** (0.030, $2.7{\cdot}10^{-6}$) ***0.033***	**0.038** (0.040, $3.8{\cdot}10^{-6}$) ***0.043***	**0.035** (0.043)
Spain	**0.037** (0.087, 0.002) ***0.026***	**1.777** (1.376, 0.082) ***2***	**0.946** (0.954, 0.198) ***0.154***	**0.238** (0.095, 0.004) ***0.614***	**0.044** (0.044, $6.8{\cdot}10^{-7}$) ***0.043***	**0.156** (0.159, 0.0004) ***0.160***	**0.030** (0.030, $1.6{\cdot}10^{-6}$) ***0.028***	**0.046** (0.047, $3.7{\cdot}10^{-6}$) ***0.044***	**0.046** (0.052)
South Korea	**0.292** (0.270 0.0004) ***0.1***	**2** (1.915, 0.009) ***0.974***	**2** (1.846, 0.046) ***1.902***	**0.123** (0.136, $7.8{\cdot}10^{-5}$) ***0.313***	**≈0.05****	**-**	**8.3·10-4** ($8.3{\cdot}10^{-4}$, $3.5{\cdot}10^{-11}$) ***0.007***	**7·10-6** ($2.1{\cdot}10^{-6}$, $2.1{\cdot}10^{-11}$) ***0.134***	**0.074** (0.078)
Lombardia	**0** (0.132, 0.012) ***$8.9{\cdot}10^{-4}$***	**0.460** (1.658, 0.188) ***0.81***	**0.295** (1.093, 0.531) ***0.302***	**0.145** (0.198, 0.065) ***0.253***	**0.027** (0.026, $3{\cdot}10^{-8}$) ***0.027***	**0.981** (1.576, 0.247) ***1.925***	**0.036** (0.036, $6.9{\cdot}10^{-6}$) ***0.045***	**0.031** (0.031, $6.9{\cdot}10^{-6}$) ***0.0405***	**0.062** (0.061)
Veneto	**0.133** (0.102, 0.002) ***0.049***	**1.704** (1.175, 0.190) ***0.97***	**0.920** (0.698, 0.144) ***0.246***	**0.032** (0.034, 0.0004) ***0.09***	**0.049** (0.182, 0.093) ***0.099***	**0.009** (0.008, 0.000) ***0.004***	**0.008** (0.008, $6.4{\cdot}10^{-8}$) ***0.009***	**0.0215** (0.021, $1.3{\cdot}10^{-6}$) ***0.024***	**0.035** (0.040)
Piemonte	**0.240** (0.163, 0.009) ***0***	**1.990** (1.518, 0.232) ***0.994***	**0.265** (1.191, 0.374) ***0.195***	**0.012** (0.117, 0.052) ***0.344***	**0.386** (0.309, 0.104) ***0.069***	**0.001** (0.005, $2.5{\cdot}10^{-5}$) ***0.007***	**0.019** (0.018, $3.7{\cdot}10^{-6}$) ***0.019***	**0.034** (0.031, $1.9{\cdot}10^{-5}$) ***0.035***	**0.056** (0.050)

*The β parameter is time-dependent, as explained in the Methods section. The reported value is the initial value. It decreases by about 70% after the mid of March, which is after 11th March national lockdown, with slight differences between Italian regions. This decrease is much less accentuated in the South Korea equation (approximately 10%).

**For South Korea, λ's six model coefficients are not shown since they represent a different mathematical law from other countries, as explained in the Methods section. However, based on our fitting to the real data, it can be observed that λ gradually increases up to 0.5 at the mid of March.

The parameters calculated with PSO are reported with the best-solution value, mean, and variance. We can observe that α, β, γ and δ had a high variance due to the intrinsic variability due to the stochastic approach. Sometimes the best solution was not aligned with the mean value. λ and κ values, instead, are strictly gathered around the mean in almost all PSO solutions, hence the low variance. This is explained by considering that, since the number of parameters was higher than the available data series (Q, R, and D), the problem was underdetermined, so that the stochastic approach can find more than one series of parameters which fits the data within an acceptable misfit. Then, λ and κ did not show large variability among the most-probable scenarios because they govern the equations that correlate Q with R and D, that is the official data series. Therefore, the estimated λ and κ were always found in the same region of the search space of solutions. South Korea and Veneto

show the highest recovery ratios (λ_1) with values around 0.05, followed by Spain (0.044). This confirms the reports which praise the Veneto model, because its administration had the capabilities of testing more quickly than other Italian regions, and the family doctors worked in a stronger synergy with the health structures. It evidences also a lower death ratio (κ_0), probably due to the better health system efficiency to treat patients, but also to the greater number of tests. In both data and policies, the Veneto region is more similar to South Korea than other parts of Italy. These aspects had an impact on the outbreak of the epidemic, as can be seen comparing figures 4, 5, and 6: Veneto region is more likely to reach the peak of active cases (Q) before the other regions.

Even though the PSO results may seem to provide a wide range for the SEIR parameters, we must stress two important aspects:

- As we already stated, the problem is underdetermined, so it is preferable to have an acceptable range of values than a unique point value, that could result to be uncertain, as could happen considering only a deterministic approach solution;
- The set of possible predicted scenarios, although related to different solutions with different sets of parameters, are quite similar, thus offering an acceptable level of variability of future predictions.

While it would be very useful to estimate a more narrow range of parameters like the infection ratio β or the latent time ($1/\gamma$), this is beyond the goals of our study, and the topic is being explored by researchers who focus also on the clinical aspects of the disease.

The model we studied has some limitations, some of which we discussed in the text. We summarize them to highlight possible needs in the further development of understanding and modeling.

- First, we have currently not sufficient information to say that, after recovery, an individual becomes totally immune to the disease, but we made this assumption: in the model we did not allow the passage from the Recovered category to the Susceptible category.
- Our model does not consider that the Exposed category may have a partial infection ability, as described in Shi (2020) [8], nor make a distinction from symptomatic and asymptomatic people, as studied in Shaikh (2020) [11].
- It does not consider the testing differences between different health system structures and country policies.
- While it fitted well Italian and Spanish data, it had more problems with South Korean ones: this evidences that different policies between countries can induce different trends in the spread of the epidemic, so the models should be adapted to different situations, with the introduction or removal of parameters. This will be especially true in analyzing the situation of least developed countries, that are not able to afford strict lockdown policies like developed countries.
- Except for the death rate parameter, the model does not have a strong link to the health resiliency of citizens: it could also be related to external factors like air pollution, which makes people more sensitive to respiratory diseases [18].
- The introduction of Google's COVID-19 Community Mobility Report represented a big data constrain that was implemented in the model in a quite simplistic manner. Further studies on the quality of those data and a rigorous implementation could be a novel interesting research topic.

5. Conclusions

We have applied two different approaches for solving the equations of the SEIR model to describe the evolution of the epidemic phenomenon in Italy and in most impacted regions of the North of Italy (Lombardia, Veneto, and Piemonte). We have considered all the possible available data on the 15th of April 2020. The main findings indicate that the deterministic approaches could be not appropriate to explore all the possible solutions of the space-domain of the solutions, because the mathematical problem is underdetermined. We recommend to fit the data of this epidemic using a

stochastic approach, such as the PSO method. According to the latter approach we have estimated different scenarios of the evolution in the next 30 days, each of which refers to a different set of parameters estimated by the algorithm. The different predicted scenarios are fairly similar and suggest that every region will soon reach the peak of the epidemic. Observing the influence that the time-varying infection rate β has on the equations may open interesting discussions about the effect of lockdown policies on the evolution of the epidemic in the near and far future.

Because the model was provided rapidly and the study was performed during the international emergency, we did not explore further the implications of different "reopening" scenarios. We can say that, if the β parameter remains at current values, e.g. if the lockdown policies are maintained or, better, the reopening of business is done with particular attention to health safety procedures, the prediction of the trend of the recovered and deaths could be considered reliable, with the approximations and the uncertainties that the PSO model have pointed out. At the Italian level, despite the great dispersion in the prediction of the quarantined and recovered cases, the number of deaths will reach a number of around 33000-35000 cases at the end of May and the number of active cases will gradually decrease. This prediction cannot take into account the impact of future decisions on social distancing.

The data and the model predictions confirm that some valuable lessons should be learned from the approaches of South Korea, which was able to contain the contagion very soon before a wide spreading of the infection. The Veneto region was one of the best examples in Italy about how integrated and synergic regional policies in social distancing and the health system can tackle the epidemic, and its epidemiological scenario is now more optimistic than those of Lombardia and Piemonte. We stress that tailored actions provide much better epidemiological outcomes than wide lockdowns, keeping in mind also the example of South Korea.

The main purpose of our work is to provide a fresh discussion and new tools able to support the policymakers in their decision about the action to minimize the impact of the disease. Our analysis demonstrated that, because the Italian health care system is highly decentralized, different regions managed different policies, which highly influenced the evolution of the epidemic in its first months: the data and the model prediction well reflected the different approaches taken by Lombardia and Veneto, two regions with similar socio-economic tissue. The overall lesson that could be learned from this analysis is out the scope of the mathematical modeling itself, and will require a wider analysis on all the possible socio-economic and political factors, even if the data analysis of the Veneto situation could be used to revisit regional and central policies early on; if so, the regions are going to emulate the main virtuous approaches of Veneto, including more demanding requests to support them in their diagnostic capacity that will weight on the central government.

Author Contributions: All authors provided a contribution to the development of the methodology and conceptualization. Code implementation and modeling: F.P. and A.G; validation and interpretation, A.G. and A.V.; writing—original draft preparation, A.G.; writing—review and editing, F:P. and A.V.; supervision, A.G. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: Computational resources provided by hpc@polito (http://hpc.polito.it)

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

References

1. Parham, P.E.; Michael, E. Outbreak properties of epidemic models: The roles of temporal forcing and stochasticity on pathogen invasion dynamics. *Journal of Theoretical Biology* **2011**, *271*, 1–9, doi:10.1016/j.jtbi.2010.11.015.

2. Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K.S.M.; Lau, E.H.Y.; Wong, J.Y.; et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* **2020**, *382*, 1199–1207, doi:10.1056/NEJMoa2001316.
3. Peng, L.; Yang, W.; Zhang, D.; Zhuge, C.; Hong, L. *Epidemic analysis of COVID-19 in China by dynamical modeling*; Epidemiology, 2020;
4. Bacaër, N. *A Short History of Mathematical Population Dynamics*; Springer London: London, 2011; ISBN 978-0-85729-114-1.
5. Calafiore, G.C.; Novara, C.; Possieri, C. A Modified SIR Model for the COVID-19 Contagion in Italy. *arXiv:2003.14391 [physics]* **2020**.
6. COVID-19 Community Mobility Report Available online: https://www.google.com/covid19/mobility (accessed on Apr 29, 2020).
7. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the Proceedings of ICNN'95 - International Conference on Neural Networks; IEEE: Perth, WA, Australia, 1995; Vol. 4, pp. 1942–1948.
8. Shi, P.; Cao, S.; Feng, P. *SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration*; Infectious Diseases (except HIV/AIDS), 2020;
9. Cheynet, E. Generalized SEIR Epidemic Model (fitting and computation) Available online: https://it.mathworks.com/matlabcentral/fileexchange/74545-generalized-seir-epidemic-model-fitting-and-computation (accessed on Apr 29, 2020).
10. Dandekar, R.; Barbastathis, G. Neural Network aided quarantine control model estimation of global Covid-19 spread. *arXiv:2004.02752 [physics, q-bio]* **2020**.
11. Shaikh, A.S.; Shaikh, I.N.; Nisar, K.S. *A Mathematical Model of COVID-19 Using Fractional Derivative: Outbreak in India with Dynamics of Transmission and Control*; MATHEMATICS & COMPUTER SCIENCE, 2020;
12. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases* **2020**, *93*, 211–216, doi:10.1016/j.ijid.2020.02.058.
13. Iwata, K.; Miyakoshi, C. A Simulation on Potential Secondary Spread of Novel Coronavirus in an Exported Country Using a Stochastic Epidemic SEIR Model. *JCM* **2020**, *9*, 944, doi:10.3390/jcm9040944.
14. Engelbrecht, A.P. *Computational Intelligence*; John Wiley & Sons, Ltd: Chichester, UK, 2007; ISBN 978-0-470-51251-7.
15. Ratnaweera, A.; Halgamuge, S.K.; Watson, H.C. Self-Organizing Hierarchical Particle Swarm Optimizer With Time-Varying Acceleration Coefficients. *IEEE Trans. Evol. Computat.* **2004**, *8*, 240–255, doi:10.1109/TEVC.2004.826071.
16. Pace, F.; Santilano, A.; Godio, A. Particle swarm optimization of 2D magnetotelluric data. *GEOPHYSICS* **2019**, *84*, E125–E141, doi:10.1190/geo2018-0166.1.
17. Pisano, G.P.; Sadun, R.; Zanini, M. *Harvard Business Review*. March 27, 2020.
18. Wu, X.; Nethery, R.C.; Sabath, B.M.; Braun, D.; Dominici, F. *Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study*; Epidemiology, 2020;