Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

Article

SEIR Modeling of the Italian Epidemic of SARS-CoV-2

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, 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 their results also with data and predictions of Spain and South Korea. 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.

---

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 or to try to predict possible contagion development scenarios.

It is based on a series of dynamic mathematical equations that consider the amount of the population potentially subject to contagion, considering that a portion of individuals is immune to contagion, the trend over time of individuals who recover after infection, and the individuals who unfortunately die [1].

In the most consolidated 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 [2]. Often these coefficients are considered constant [3] that is, they are unable to take due account of external influences, such as actions to contain 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 and pharmacological development capable of treating infected individuals. Like all models, the quality of the observed data and their validation is a crucial node in assessing the reliability of the model results.

This work is carried out during the crucial development phase of the epidemic in Italy (end of March 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 [4].

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

The generalized SEIR model we used is based on the system of differential equations found by Peng et al. (2020) [5] 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). Moreover, it introduces time-dependent model parameters and provides reliable results in the prediction of the peak of the epidemic in different China provinces.

Besides, we want to check the feasibility of using a stochastic approach, based on the particle swarm optimization approach (PSO) [6] over a deterministic method in the analysis of the evolution of an epidemic outbreak. In fact, citing the overview made by Parham (2012), 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.

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 collect the data from different official organizations such as OMS, USA Centers for Disease Control and Prevention, ECDC China, NHC, and other organizations. The Chinese data are collected through the DXY, a Chinese web-site able to aggregate the reports on the local situation furnished by local authorities, on the cases of confirmed deaths and infected. The Italian data are mainly collected through the bulletin of the *Protezione Civile Italiana*.

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 susceptible (S) is the part of the population that could be potentially subjected to the infection. 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 according to the kind of disease they are infectious, partially infectious or not infectious [7]. The infective (I) represents the infective population after the latent period. The recovered (R) are people after healing, and they are generally removed by the susceptible category, if it is considered that they became immune to the disease. The R category comprehends also the individuals died due to the disease.

The parameters β (infection ratio) controls how fast people move from being susceptible to exposed, and γ, the inverse of the latent time, controls how fast individuals pass from exposed to infective. Another parameter, δ, controls how fast people get recovered from being infected.

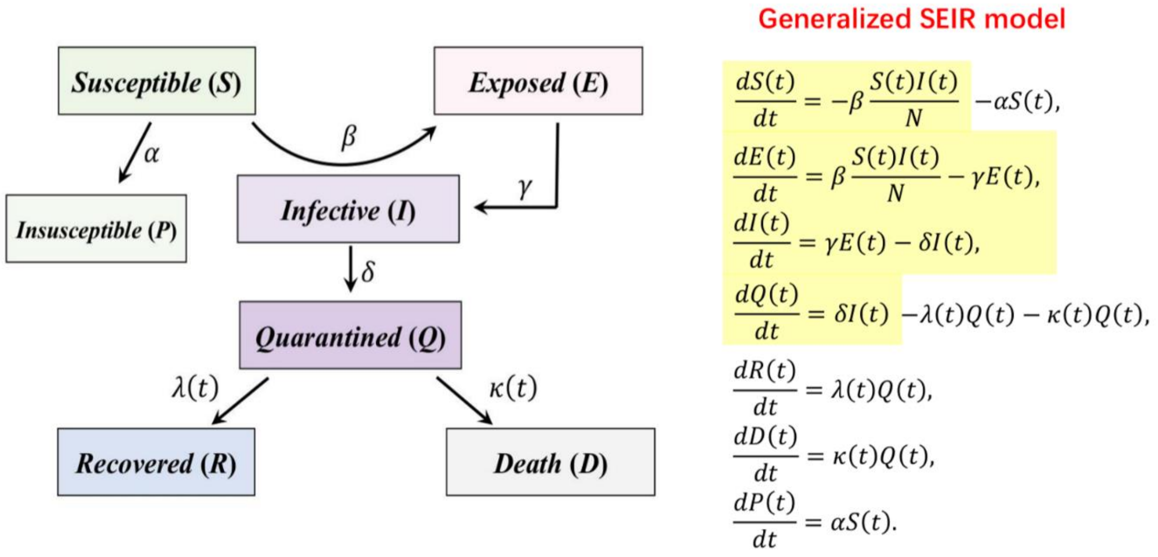
Given the complexity of the disease, we used a generalized SEIR model [5] which works with more parameters than the classical SEIR model:

- α is the protection rate,
- β is the infection rate,
- γ is the inverse of the average latent time,
- δ is the inverse of the average quarantine time,
- λ_0 and λ_1 are coefficients used in the time-dependent cure rate,
- κ_0 and κ_1 are coefficients used in the time-dependent mortality rate.

Notice that λ and κ, the cure rate and mortality rate respectively, are considered time-dependent to enhance the capability of the model to follow the evolution of the outbreak. On the basis of the results achieved by Peng et al. (2020) [5], which suggested an exponential evolution of the two parameters, we constrained them to fit an exponential trend. The λ_0 represents the initial value of the cure rate, which is the velocity at which an average infective individual can recover. It is related to the initial health system ability to tackle the infection and depends also on other factors like the good health of the citizens. λ_1 measures how the rate has changed with time. Given the properties of the λ function, if λ_1 turns out to be positive, the cure rate will improve over time, and the higher it is, the better this improvement.

The parameter κ_0 represents the initial value of the mortality rate, which is the velocity at which an average infective individual could die. Like λ_0, κ_0 parameter 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. If κ_1 turns out to be positive, the mortality rate will decrease over time, and the higher it is, the better this decrease.

The generalized SEIR-model scheme is described graphically and mathematically in Figure 1.



$$\frac{dS(t)}{dt} = -\beta\frac{S(t)I(t)}{N} - \alpha S(t),$$
$$\frac{dE(t)}{dt} = \beta\frac{S(t)I(t)}{N} - \gamma E(t),$$
$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t),$$
$$\frac{dQ(t)}{dt} = \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t),$$
$$\frac{dR(t)}{dt} = \lambda(t)Q(t),$$
$$\frac{dD(t)}{dt} = \kappa(t)Q(t),$$
$$\frac{dP(t)}{dt} = \alpha S(t).$$

Figure 1. Generalized SEIR model. The highlighted equations are the classical SEIR model core equations (modified from [5]).

The main outputs of the model are the following data:

- S, target time-histories of the susceptible cases,
- E, the target time-histories of the exposed cases,
- I the target time-histories of the infectious 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 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 a temporary category: its individuals pass into the infective (I) category after a latent time (1/ γ), on average. The infective category generates newly infected people over time, taking them from the Susceptible category, and infected individuals are made quarantined (Q) to avoid spread. Then, they evolve into recovered or death cases, according to various causes, as health care system effectiveness, age, co-presence of other diseases.

The α, β, γ, δ, λ and κ parameters can be considered also an important output of the model, and the evolution over time of λ and κ could provide insightful information about the changes in the population and health system response to the contagion.

2.3. Implementation of the SEIR model and stochastic approach

The model equations were implemented using the MATLAB code provided by [8] 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 dead 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. At the beginning of modeling, their initial values are given as first esteem. Then, their values are calculated following a least square curve fitting approach that considers the observed data (Q, R, D) with time. The differential equations in Figure 1 are numerically solved by means of the Runge-Kutta method.

We modified the model by introducing a stochastic approach for solving the optimization problem. The six unknown parameters are calculated by minimizing an objective function that minimized the misfit between observed data (Q, R, D) and the response calculated by the SEIR model. This approach leads to the best optimized parameters, for several trials. The calculation of the parameters is based on the Particle Swarm Optimization (PSO) algorithm, which is a population-based nature-inspired metaheuristic method belonging to the family of Computational Intelligence [9, 10]. The main difference with respect to the standard linearized method (deterministic) is that the result is a set of most probable parameters, giving a set of most probable predicted curves.

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.

The analysis of the overall Italian situation 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 PSO approach gives a similar output in terms of data fitting and forecast of the development of the epidemic, even if the fitted parameters differ from those obtained by the deterministic approach, according to the global searching in the model domain parameter of the stochastic approach (Figure 3).

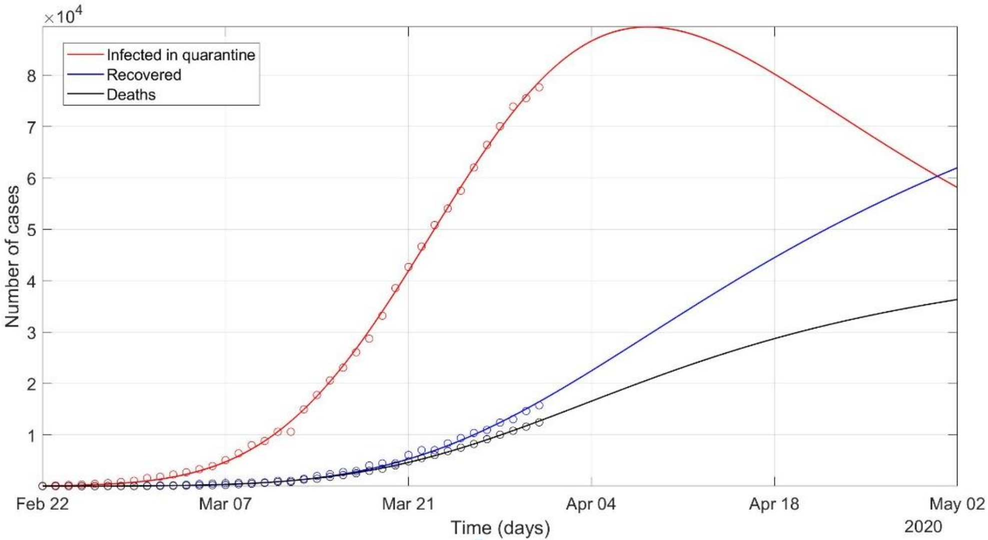


Figure 2. Observed data of infected in quarantine, recovered and deaths for Italy, updated on March 31st, 2020. Deterministic approach. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

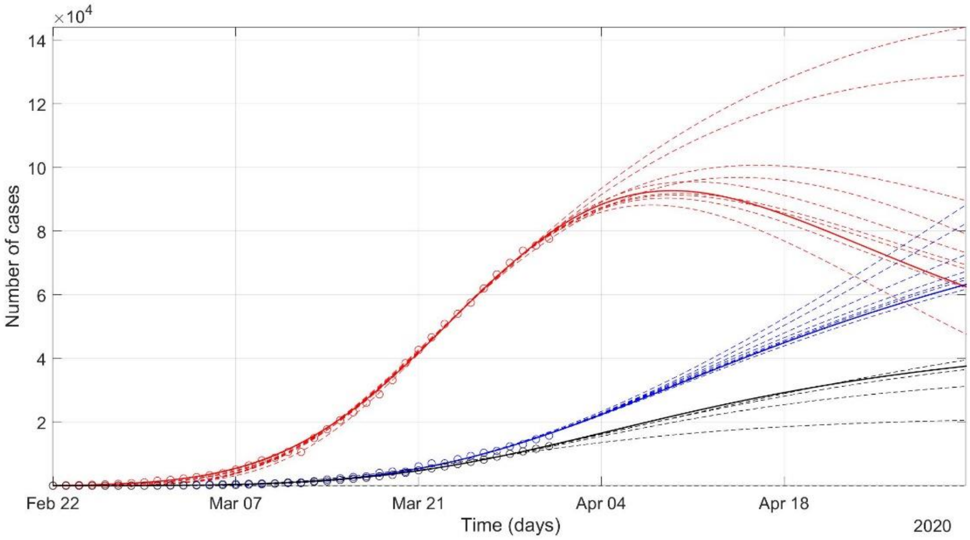


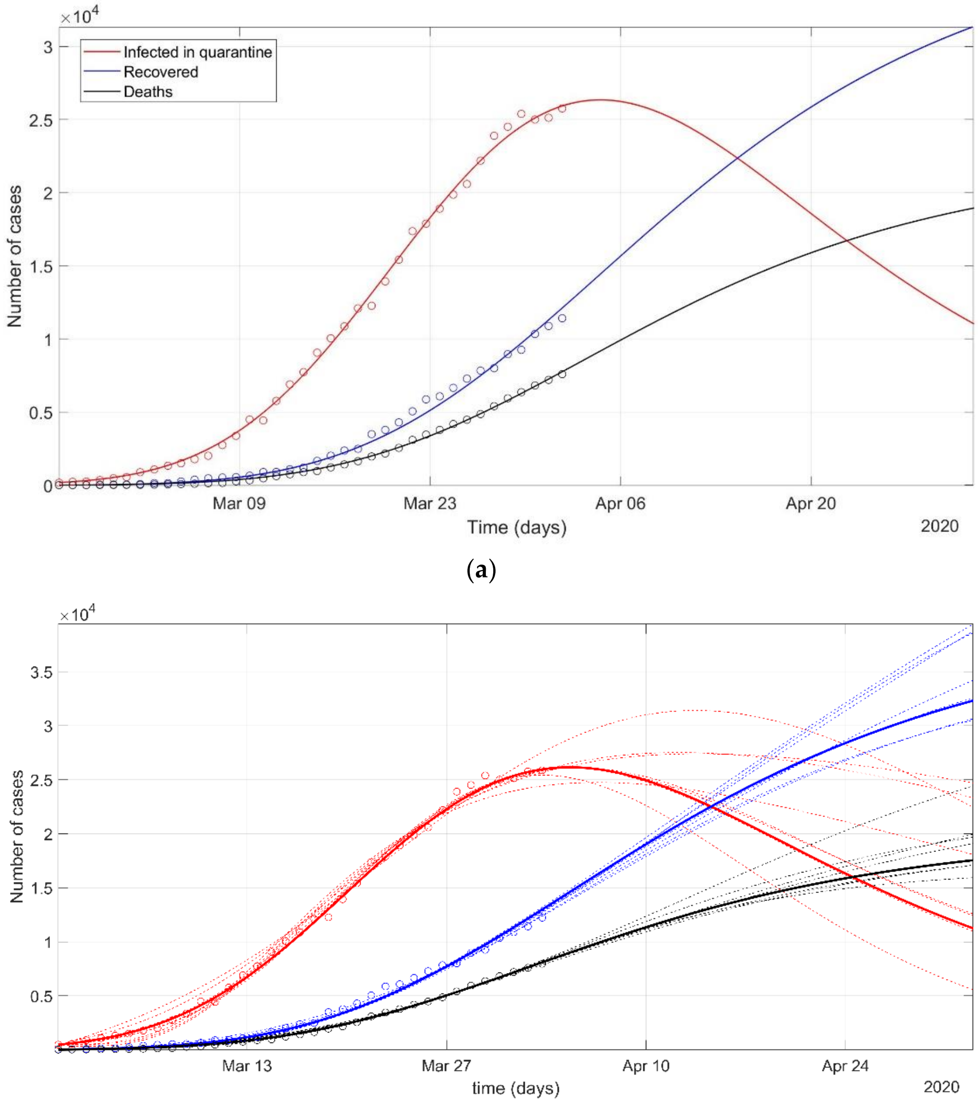
Figure 3. Observed data as in Figure 2. Observed data of infected in quarantine, recovered and deaths for Italy, updated on March 31st, 2020. Stochastic approach. The dashed lines are the predicted trends in 30 days according to the set of probable scenarios for different optimized parameters. The solid lines are the most probable scenarios for Q, R, D.

In these graphs, the points refer to the data provided by official institutions, and they are plotted together with the model-predicted trend lines. We showed in red the number of individuals tested positive and placed in quarantine (at home, or hospitalized, or in intensive care). 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 number. Blue and black colors represent the number of recovered and the number of deaths, respectively. 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 analysis of the situation of Lombardia, Veneto and Piemonte regions allows us to detect the following elements. 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. The predicted evolution of the phenomenon is depicted in Figures 4, 5 and 6 for Lombardia, Veneto and Piemonte, respectively.

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; 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 scenario, according to the PSO analysis (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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1



(**a**)

(**b**)

Figure 4. Observed data of infected in quarantine, recovered and deaths for Lombardia region, updated on March 31st, 2020. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model: (**a**) Deterministic approach; (**b**) Stochastic approach.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

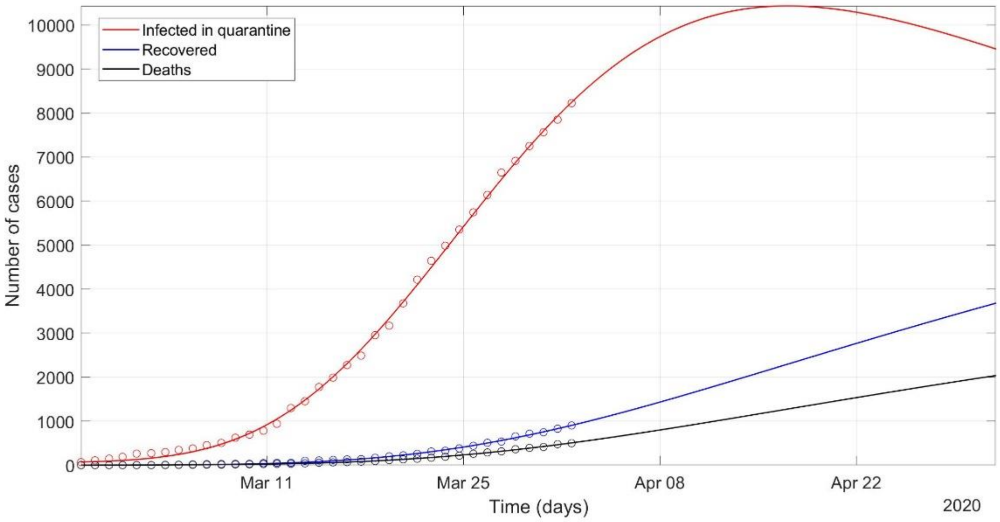


Figure 5. Observed data of infected in quarantine, recovered and deaths for Veneto, updated on March 31st, 2020. Deterministic approach. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model.

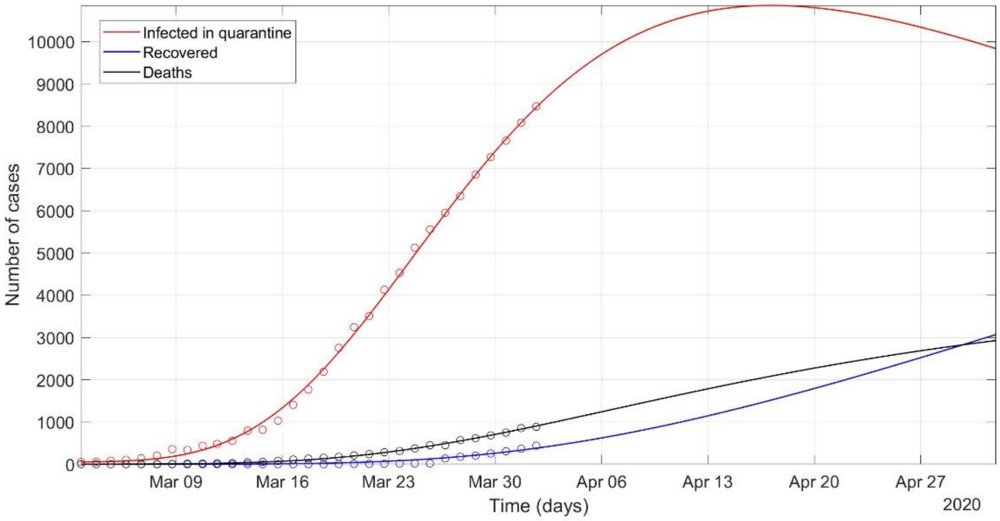


Figure 6. Observed data of infected in quarantine, recovered and deaths for Piemonte region, updated on March 31st, 2020. Deterministic approach. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model.

The picture of the situation in Spain is shown in Figure 7. The crisis exploded in few days, just 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; anyway, in order to give a wider overview of the possible scenarios a stochastic analysis could be more appropriate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

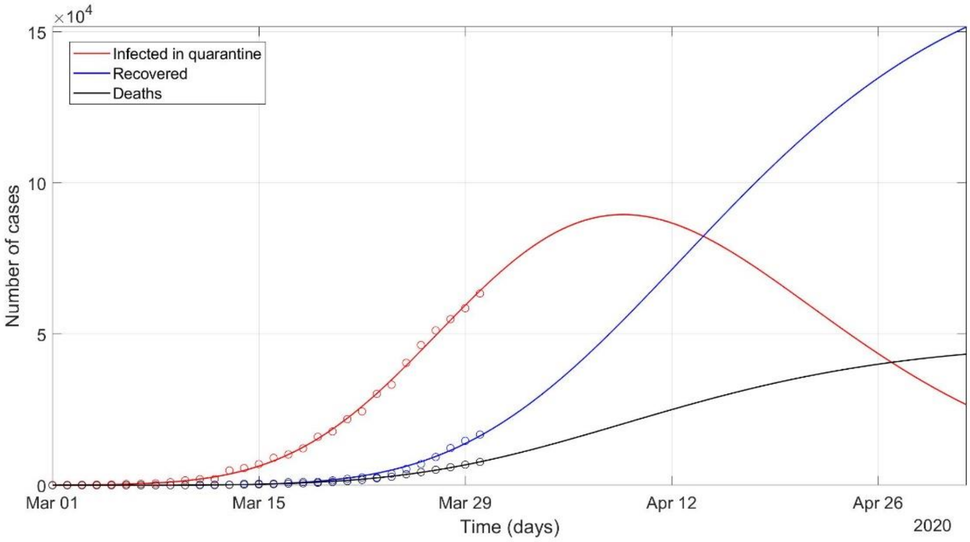


Figure 7. Observed data of infected in quarantine, recovered and deaths for Spain, updated on March 31st, 2020. Deterministic approach. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model.

4. Discussion

We adopted a generalized SEIR method 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 with a deterministic approach, and then with a stochastic approach, using a Particle Swarm Optimization algorithm, a novelty in the field of epidemiological studies. The main aim is to offer an opportunity to the policymaker in Italy or elsewhere to think about the importance that any decision in an emergency status must avoid any wasted time.

The analysis of the result given by the simulation using the stochastic approach (PSO) gives an overview of the most probable scenario in the days after April, 1st. It can be well noticed that the different predicted model responses are obtained with approximately an equivalent goodness of the data fitting (the normalized, with respect to the mean value of the quarantined, L_2-norm < 0.01), even if they are forecasting very different scenarios for the incoming days. 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 selection of the rate of infected, recovered, and deaths.

The data seem to confirm that while Lombardy and Piemonte applied similar approaches to social distancing and retail closures, Veneto's strategy considers a much more proactive effort in order to limit the contagious, by 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.

The behavior of Piemonte deals with a peculiar trend, introducing a time-delay of the recovered with respect to the death cases, and the rates also show a weird behavior, where the number of the recovered in the month of March is always lower than the deaths. This involves that the intersection of the trends of recovered and deaths will be reached at a later time with respect to other regions herein analyzed. We cannot say if all these peculiarities are caused by different approaches in computing the cases, with an underestimate of the recovered versus the infected or they are affected by some inertia in the health system.

A recent analysis [11] 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.

The expected trend of these two regions is controlled by many factors outside the control of policymakers, including Lombardy'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.

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. We could not fit the observed data, because of the marked oscillations both in the time-series of the quarantined and in the recovered one. The abrupt changes of the trends will require more careful analysis, taking into account the time and the effects of the action of the containment of the infection and also the effects of any new pharmacological treatment methods introduced by the health system.

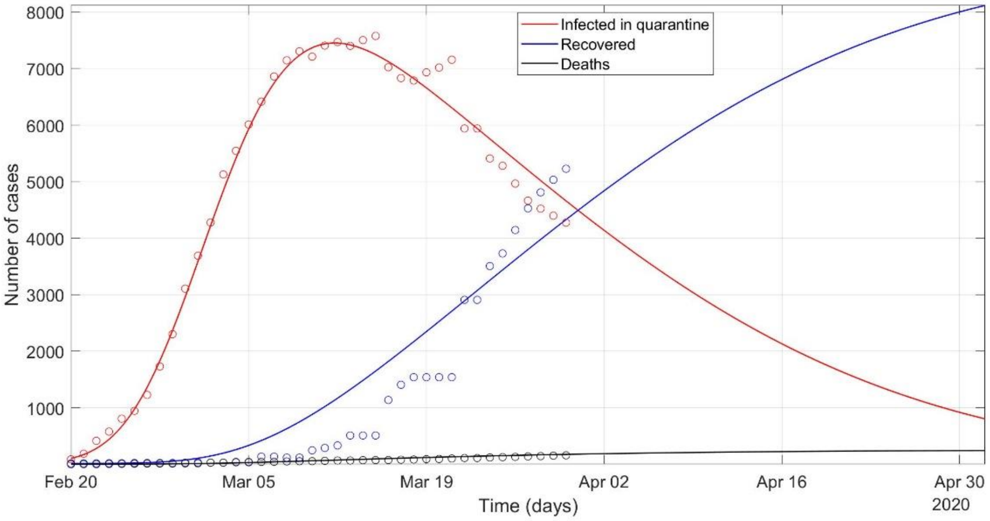


Figure 8. Observed data of infected in quarantine, recovered and deaths for South Korea, updated on March 31st, 2020. Deterministic approach. The continuous lines refer to the trend of the predicted evolution in 30 days according to the SEIR model.

To sum up our results, the estimated parameters for Italy, Spain and South Korea are shown in Table 1. They all refer to the deterministic approach. The β parameter, that governs the infection rate of the disease, appears to be quite similar in the various countries. It is interesting to notice the λ parameters: λ_0 is much smaller for Italy than for the other two countries, and this could represent an initial struggle to treat the patients, or also simply longer average quarantine time at the beginning of the outbreak. Italy has the highest value of the λ_1 parameter, thus meaning that the ability of the health system to cure patients has improved over time. This is consistent with the latest news from *Protezione Civile Italiana*.

Table 1. The final calculated values of α, β, γ, δ, λ and κ for Italy, Spain and South Korea.

Country	Prediction	α	β	γ	δ	λ_0	λ_1	k_0	k_1
Italy	30 days	0.0292	1.2687	0.7631	0.5086	0.0181	2.0000	0.0234	0.0174
Spain	30days	0.0636	1.3388	0.5329	0.2110	0.4814	0.0031	0.0170	0.0032
S. Korea	30 days	0.1044	1.3511	0.8486	0.4093	0.2068	0.0051	0.0010	0.0000

5. Conclusions

We have applied two different approaches for solving the equations of the SEIR model to focus on 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 at the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2020 doi:10.20944/preprints202004.0073.v1

end of March 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 solution; we suggest providing the fitting of the data using a stochastic approach, based on the PSO method. According to the latter approach we have estimated different scenarios of the evolution in the next 30 days.

The data and the model predictions confirm that some valuable lessons should be learned from the approaches of South Korea, and in Taiwan, and Singapore, which were able to contain the contagion very soon before a wide spreading of the infection.

The main purpose of our work is to provide 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. We pointed out how the data and the model prediction well reflected the different approaches taken by Lombardy and Veneto, two regions with similar socio-economic tissue.

The overall lesson that could be learned from this analysis is out the scope of our modeling and will require a wider analysis on all the possible socio-economic and political factors.

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.

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.
2. Bacaër, N. *A Short History of Mathematical Population Dynamics*; Springer London: London, 2011; ISBN 978-0-85729-114-1.
3. Calafiore, G.C.; Novara, C.; Possieri, C. A Modified SIR Model for the COVID-19 Contagion in Italy. arXiv:2003.14391 [physics] **2020**.
4. 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.
5. Peng, L.; Yang, W.; Zhang, D.; Zhuge, C.; Hong, L. *Epidemic analysis of COVID-19 in China by dynamical modeling*; Epidemiology, **2020**.
6. Kennedy, J. and Eberhart, R. Particle swarm optimization. International Conference on Neural Networks IV (IEEE) proceedings, extended abstracts, **1995**, 1942–1948. doi: 10.1109/ICNN.1995.488968.
7. 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**.
8. Cheynet E. (2020). *Generalized SEIR Epidemic Model (fitting and computation)* (https://www.github.com/ECheynet/SEIR), GitHub. Retrieved 2 April 2020.
9. Engelbrecht, A. P. *Computational intelligence: An introduction*. John Wiley and Sons Ltd, **2007.**
10. Ratnaweera, A., Halgamuge, S. K. and Watson, H. C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Transactions on Evolutionary, **2004.**
11. Pisano, G.P., Sadun R., Zanini M., 2020. Lessons from Italy's response to Coronavirus, Harvard Business Review, 27 March 2020