Modified SEIR model for COVID-19 outbreak in Spain

Javier Montoliu and Paula Ruiz de la Cuesta
ETSETB, ***UPC***
Barcelona, Spain
(Dated: June 19, 2020)

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak has been ongoing for over 5 months now and has spread to nearly every country in the world from its initial outbreak in Wuhan, China. Europe, specially Italy and Spain, has been one of the most affected areas by the virus, although there are now many other epicentres surging in other areas, such as North and South America. In this project we will try to model the disease spread in our home country Spain and try to find some useful parameters to analyse and characterise the outbreak. We have thus tried to implement a modified SEIR model that includes the effects of the lockdown imposed on March 14th by the Spanish government.

Creating and estimating the model parameters has proven to be a rather difficult task for us, since the data involving this disease is often unreliable and under continuous modification of criteria. That's the reason why we have only achieved to model the outbreak up to a point and, given time and difficulty constraints, we have been forced to settle for the results obtained. The algorithm for the fitting of the model and estimation of the parameters was initially implemented in R language, but due to complications and the authors' familiarity with the MATLAB environment, the latter was chosen in the end.

I. DESCRIPTION OF THE MODEL

We have followed the same approach as [1] and have tried to implement a modified SEIR model that takes the latent period of this new virus into consideration. SARS-CoV-2 virus has shown to have significantly long latent period, the time between the infection with a virus and the onset of the first symptoms.

The model in question considers 7 compartments or populations:

- Susceptible (S): individuals liable to be infected with COVID-19.
- Exposed (E): infected indiviuals but not yet infectious themselves, in a latent period.
- Infective (I): infected individuals capable of infecting others.
- Quarantined (Q): infected people that have tested positive, have been quarantined and are thus not infectious anymore (corresponds to the active reported cases figures given by the government).
- Recovered (R): people recovered from the disease and removed from the epidemic dynamics.
- Dead (D): deaths caused by the virus.
- Confined (C): individuals that are removed from the dynamics of the epidemic through government lock-downs or social distancing measures.

When developing the model, several assumptions have been made. The population is assumed constant due to the fast evolution of the spread. The infection of susceptible individuals occurs upon contact between susceptible (S) and infective (I) at a rate β. We will also assume that infective, once tested positive, are all quarantined and can no longer infect other susceptible individuals. In turn, susceptible population can be protected by confinement through lock-down enforced by the government, in which case they are moved to the protected/confined (C) compartment. It is assumed, however, that this protection occurs at a characteristic rate, when in reality, it's closer to an instantaneous transfer of a portion of the S population to the C compartment at the day in which lock-down was enforced in Spain.

Hospitalizations and hospital infections will not be considered in our model, in other to simplify the parameter estimation process. Confined people are assumed to have no contact with the rest of the dynamics and cannot thus be infected, which, again, is not exactly true since there is still some contact between confined people and the rest of the populations.

The dynamics of the disease are modelled according to the following equation system:

$$\frac{dS}{dt} = -\alpha S(t) - \beta \frac{S(t)I(t)}{N} \tag{1a}$$

$$\frac{dE}{dt} = -\gamma E(t) + \beta \frac{S(t)I(t)}{N} \tag{1b}$$

$$\frac{dI}{dt} = \gamma E(t) - \delta I(t) \tag{1c}$$

$$\frac{dQ}{dt} = \delta I(t) - \lambda(t)Q(t) - k(t)Q(t) \tag{1d}$$

$$\frac{dR}{dt} = \lambda(t)Q(t) \tag{1e}$$

$$\frac{dD}{dt} = \kappa(t)Q(t) \tag{1f}$$

$$\frac{dC}{dt} = \alpha S(t) \tag{1g}$$

The main parameters of the model are the confinement/protection rate α, the infection rate β, the inverse of the latent period γ, the rate at which infectious people enter in quarantine δ, the recovery rate $\lambda(t)$ and the mortality rate $\kappa(t)$.

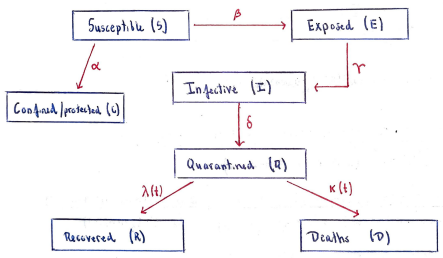

FIGURE 1. Diagram of the model.

The confinement rate α models the different actions taken by the government to control the spread and protect susceptible population. Exposed (E) individuals become infective (I) after an incubation time τ, the inverse of which we will call the incubation rate γ. It has been experimentally observed that the incubation period has a value of around ~ 4 days, giving us an initial expected value of $\gamma = 0.2$.

Mortality and recovery rate parameters have been assumed to be time-dependant and modelled by one of the following functions. For the mortality rate, the proposed modelling functions are:

$$\kappa(t) = \kappa_0 \exp\left(-\kappa_1 (t - \tau_\kappa)^2\right) \tag{2a}$$

$$\kappa(t) = \kappa_0 + \exp\left(-\kappa_1 (t - \tau_\kappa)\right) \tag{2b}$$

$$\kappa(t) = \frac{\kappa_0}{\exp\left(\kappa_1 (t - \tau_\kappa)\right) + \exp\left(-\kappa_1 (t - \tau_\kappa)\right)} \tag{2c}$$

where κ_0 and κ_1 are dimensionless constants and τ_κ has time dimensions. The recovery rate $\lambda(t)$ can either be modelled as:

$$\lambda(t) = \frac{\lambda_0}{1 + \exp\left(-\kappa_1 (t - \tau_\kappa)^2\right)} \tag{3a}$$

$$\lambda(t) = \lambda_0 + \exp\left(-\lambda_1 (t - \tau_\lambda)\right) \tag{3b}$$

where λ_0 and $_1$ are dimensionless constants and τ_λ has time dimensions. It is the MATLAB code implemented for the fitting which will be in charge of choosing the adequate models for these two parameters.

II. MODEL FITTING AND PARAMETER ESTIMATION

As previously mentioned, the chosen environment for the fitting finally ended up being MATLAB. At first, an attempt with R language was made, due to several recommendation of this language for data analysis fields. The only problem encountered with the R environment was the lack of knowledge of some of the functions required for the implementation of the model, since none of us had ever programmed with it before. It is true that we both had the impression that, in general, data analysis is much easier on R (it is similar to the *Pandas* environment in Python) and managed to get some decent fitting but it was very slow for us and due to the litte time available for developing this project we needed the speed that only our familiarity with MATLAB could provide.

The estimation of the parameters will be based on non linear curve fitting by minimization of the sum of square errors (least squares). The function used for the fitting of the data is *lsqcurvefit()*, which performs a trust-region-reflective algorithm and can also return the residual in order to chose the best model (for the recovery an mortality rate parameters). The model's ODE system of equations is solved using a 4th order Runge-Kutta method with a higher resolution time vector in order to improve the algorithm's convergence.

The data for the fitting is directly obtained from the website of Spain's Ministry of Health [2] which provides daily figures for the new reported cases, recoveries and deaths for each of the 17 autonomous communities (CCAA) of the country, in a .csv file that we can directly import into MATLAB. (as of 18 June 2020, the government has started releasing data at a provincial level, which was not available at the time of the development of this project).

The fitting of the parameters is performed on three stages:

1. **Preliminary fitting of the recovery rate λ.** As a first step, using the Recovered figures of the data available, we compute a first approximation of the daily recovery rate and preliminary fit the λ_0, λ_1 and τ_λ parameters. This is a necessary step in order to ensure a proper convergence of the subsequent general fitting of all the parameters. The fitting and the choice of the function is done using the lsqcurvefit() function from MATLAB.

2. **Preliminary fitting of the mortality rate κ.** An analogous procedure is done for the mortality rate, and the first estimations for κ_0, κ_1 and τ_κ are obtained.

3. **General fitting of the rest of the parameters.** Once the preliminary fitting is done and we have a first approximation for the λ(t) and κ(t) parameters, it is time to perform a general fitting for the rest of the parameters with a maximum number of iterations of 1200 and a tolerance of $1e-5$.

III. RESULTS AND DISCUSSION

As a start, we will use the general case of the whole Spanish region to estimate the model parameters and to illustrate the process of obtaining our results. We will subsequently attach table including the estimated parameters of each of the 17 autonomous communities and the graph of their populations' evolution.

The data for Spain has proven to be fairly unreliably and is under constant change of criteria and/or corrections and adjustments. In particular, following a change in the counting criteria, last few weeks' data has stopped showing recovered cases and the active cases do not still show a clear new tendency. We have decided to ignore, for the time being, the last 3 weeks to simplify the fitting, although we are aware that is not the adequate way to proceed. With our current resources, we feel unable to sort the data or try to adapt the new data the old criteria.

Applying our fitting code to the data of Spain, starting from Monday the 20th of March up until May 9th, yields the following results:

Parameter	Value
α	0.0428
β	0.7392
γ	0.1585
δ	0.4937
$[\lambda_0, \lambda_1, \tau_\lambda]$	$[0.0283, 0.0430, 93.2488]$
$[\kappa_0, \kappa_1, \tau_\kappa]$	$[[0.0004, 0.0497, 78.689]$

TABLE I. Estimated parameters for Spain.

In figure 2 we have plotted the reported official data against the solved model for the estimated parameters.

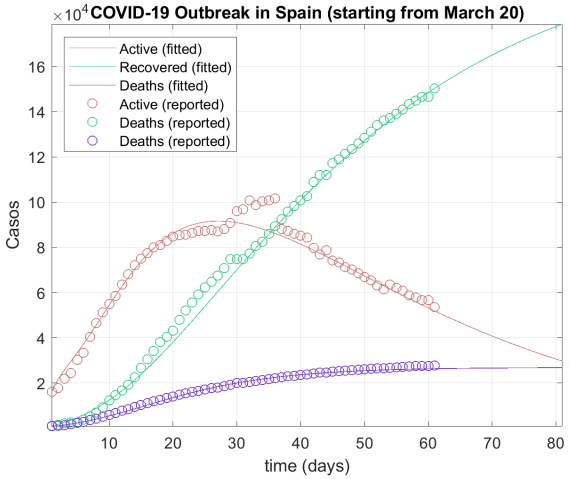

FIGURE 2. Fitted model vs reported data.

The results obtained for the data of Spain at a statal level provide a fairly plausible model. The trend shows a latent period of $1/\gamma = 6.309$ days for the exposed individuals to become infective, which is highly compatible with the current situation. The average infective individual is quarantined after $1/\delta = 2.03$ days of showing the first symptoms. The mortality shows a decreasing trend which is in accordance to the relief of pressure in hospitals and better treatment of the disease. The algorithm has been applied to each of the 17 CCAA of Spain and the estimated parameters are shown in TABLE II. Due to length limitations it is not possible to include the graphs for each of the communities, but we have uploaded them and linked them here for the interested reader.

IV. INTENDED IMPROVEMENTS FOR THE MODEL.L

After some online meetings with our tutor Blas Echebarría and Sergio Alonso from the Computational Biology and Complex Systems group at UPC, we decided upon a few correction to the model in order to better simulate the real spread in Spain. Unfortunately, we haven't had enough time to implement them. We have tried to adapt our code to this new guidelines but we were not successful in reaching a fully working program. We still want, however, to show the new model as we would have liked to implement. The main changes with respect to the previous one are detailed below.

CCAA	α	β	γ	δ	λ_0	λ_1	τ_λ	κ_0	κ_1	τ_κ
Andalucía	0.2637	1.500	0.0955	0.3140	1.000	0.0612	84.2728	0.0162	0.0525	1.1526
Asturias	0.9993	1.2863	0.0990	0.1388	0.0287	0.0142	0.0039	0.0059	0.0009	18.6828
Illes Balears	0.0000	0.3538	0.3230	0.4884	0.0526	0.7313	4.1504	0.0282	0.1610	18.3138
Canarias	0.4164	1.5000	0.1055	0.2827	0.0311	0.7116	10.5861	0.0005	0.0706	68.0138
Cantabria	0.0006	0.2936	1.0000	0.3628	0.2837	0.1151	41.9589	0.0030	0.0477	116.2969
C-La Mancha	0.2710	1.5000	0.1613	0.2100	0.0179	0.6320	81.2933	0.0010	0.0608	61.1876
Castilla-León	0.0960	1.5000	0.0865	0.05138	0.0000	0.0379	74.2951	0.0300	0.1244	12.2097
Catalunya	0.0165	0.6850	0.1851	0.5924	0.0184	0.0405	99.9994	0.0204	0.0697	9.8432
Valencia	0.2652	1.5000	0.0993	0.3492	0.1010	00.1167	23.2489	0.0031	0.0590	76.6677
Extremadura	0.0920	1.5000	0.0812	0.7800	1.0000	0.1069	54.3770	0.3333	.3333	4102590
Galicia	0.0226	0.3546	0.5019	0.3048	0.4511	0.1541	47.3838	0.0045	0.0000	126.0275
La Rioja	0.3380	1.5000	0.0858	0.2340	0.0217	0.0494	62.0007	0.0081	0.0010	3.5594
Madrid	0.4417	0.5651	0.0891	0.1758	0.0033	0.0278	99.7998	0.0003	0.0440	90.8433
Navarra	0.1916	0.9220	0.1021	0.2384	0.0653	0.0682	33.2401	0.0007	0.0312	157.3723
País Vasco	0.0581	0.4073	0.1522	0.2995	1.0000	0.0368	90.3664	0.6194	1.9570	90.6745
Murcia	0.4499	1.5000	0.1435	0.3374	1.0000	0.0916	57.2761	0.0048	0.8189	185.3776
Aragón	0.0231	0.3863	0.5074	0.3361	0.2747	0.0295	99.9995	0.0228	0.0690	13.3596

TABLE II. Estimated parameters for each autonomous community.

a. ***Lockdown modelling:*** The process of confinement/protection of the susceptible population is modelled as an instantaneous transfer of a percentage of the total population from the S compartment to the C. The new parameters modeling this dynamic will thus be Dirac deltas $\delta(t - t_0)$ (Kronecker deltas in practice)

We will have one parameter governing the confinement through the lockdown:

$$\psi(t) = a\delta(t - t_{\text{lockdown}})$$

and another for the lifting of the lockdown:

$$\phi(t) = b\delta(t - t_{\text{lifiting}})$$

Where a and b represent the confined or deconfined population in each measure, respectively. The parameters t_{lockdown} and t_{lifting} represent (in days) the day in which the lockdown was enforced (March 14th in the case of Spain) and the day lockdown measures are lifted, respectively. The latter is a bit harder to determine, since the lifting of the confinement measures is being done progressively and is not uniform along the whole country. This progressive nature of the deescalation could be modelled by several partial lifting of the lockdown in each of which a portion of the confined population is released. We would therefore have a lifting parameter close to this:

$$\phi(t) = b_0\delta(t-t_{\text{phase},0})+b_1\delta(t-t_{\text{phase},1})+b_2\delta(t-t_{\text{phase},2})+...$$

The date of the transition to each of the phases varies not only within the CCAA student, but also within the provinces and an epidemiological study at the province level would be required.

b. ***Asymptomatic infective:*** It has been shown that there is a large group of non-tested asymptomatic infected population that are also infectious and greatly contribute to the spread of the virus. We can introduce a new compartment called asymptomatic (A). This new population will now contribute to the infection force within the term

$$\beta \frac{S(t)[I(t)+A(t)]}{N}$$

Under these conditions, a fraction p of the exposed (E) population will become infective symptomatic (I) and $(1-p)$ will turn into asymptomatic infective (A). The asymptomatic will have an independent recovery rate ν.

c. ***Age stratification:*** The key point of this new model is the splitting of the population in subcategories by age (for example, we propose 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80 and 80+ years old). This allows the study and possible implementation of the so called **vertical lockdown** of the population, that is, lockdown measures that depend on the age subcategory to which the individual belongs. The intention was to implement this model to the current Spanish data and from that, study the possible case scenarios comparing horizontal (equal confinement at any age group) to vertical lockdown measures, seeing up to which point the latter presented a better alternative.

Considering this new subdivision into age categories, and to simplify the notation, we will define a new magnitude, the infection force, as:

$$\mathcal{I}(t) = \sum_i \frac{S_i(t)+A_i(t)}{N_{total}}$$

where the subindex i represents one of the 9 age groups. This magnitude gives an idea of the total infectious population (including all age bins) per inhabitant of the region.

To sum up, taking into consideration the 8 compartments and the 9 age groups, this new model will be described by the following system of 72 uncoupled equations:

$$\frac{dS_i}{dt} = -\beta S_i(t)\mathcal{I}(t) - \psi_i(t)S_i(t) + \phi_i(t)C_i(t) \tag{4a}$$

$$\frac{dE_i}{dt} = \beta S(t)\mathcal{I}(t) - \gamma E(t) \tag{4b}$$

$$\frac{dA_i}{dt} = (1-p)\gamma - \nu_i(t)A(t) \tag{4c}$$

$$\frac{dI_i}{dt} = p\gamma E(t) - \delta I(t) \tag{4d}$$

$$\frac{dQ_i}{dt} = \delta I(t) - \lambda_i(t)Q(t) - k(t)Q(t) \tag{4e}$$

$$\frac{dR_i}{dt} = \nu_i A(t) + \lambda_i Q(t) \tag{4f}$$

$$\frac{dD_i}{dt} = k_i(t)Q(t) \tag{4g}$$

$$\frac{dC_i}{dt} = \psi(t)S(t) - \phi(t)C(t) \tag{4h}$$

Under the constraint $S_i+E_i+A_i+I_i+Q_i+R_i+D_i+C_i = N_i$, where N_i is the total population of that particular age group in the region studied. Notice how the lockdown measures ψ_i and ϕ_i, as well as the mortality and recovery rates (both for symptomatic and asymptomatic) depend on the age subcategory i.

We can benefit from this model and study how confinement of only older age groups can lead to a substantially lower peak of active cases and prevent the collapse of hospitalization system, with the total population requiring intensive care (ICU) falling within the region's capacity. In this regard, a further H_i group of hospitalized population could be added to the model in order to model the dynamics of the hospital and ICU beds availability.

V. CONCLUSIONS

In spite of investing many hours in this project trying to develop a robust algorithm for the parameter estimation, we haven't been able to achieve the desired results. The proposed improved model should have been implemented and studied, to compare possible lockdown scenarios and other control measures. As we have repeatedly stated during the project, we have faced a significant lack of available time, caused mainly by the inevitable readjustment and rearrangement of teaching or evaluating activities of the university following this exceptional situation.

We would have liked to propose a stratified vertical lockdown in which the peak hospitalizations never reached the maximum ICU beds available in Spain and simulate other scenarios with the model, thereby reaching some interesting conclusions. We were, however, unable to attain this target and consider this as a future improvement of the project. Reported data of COVID-19 cases has also proven to be very unreliable and we expect that during the following months (thanks, in part, to antibody testing), we all will be able to get clearer idea of what the situation really was at this time, understand how we should have acted and learn what measures should have been taken, in case a new wave of this or other virus was to come.

VI. REFERENCES

[1] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, Liu Hong. Epidemic analysis of COVID-19 in China by dynamical modeling. https://arxiv.org/abs/2002.06563 (2020)

[2] Spain's Ministry of Health official data https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos

[3] Yu Chen, Jin Cheng, Yu Jiang, Keji Liu. A Time Delay Dynamical Model for Outbreak of 2019-nCoV and the Parameter Identification. https://arxiv.org/abs/2002.00418 (2020)

[4] Begoña Cantó, Carmen Coll Elena Sánchez. Estimation of parameters in a structured SIR model. https://advancesindifferenceequations.springeropen.com/articles/10.1186/s13662-017-1078-5 (2020)