Modelling COVID-19 evolution in Italy with an augmented SIRD model using open data

Vincenzo Nardelli, Giuseppe Arbia, Andrea Palladino and Luigi Giuseppe Atzeni

Abstract We propose an augmented version of the traditional SIRD epidemic model and we estimate its parameters using the SaRs-Cov-2 Italian open-data. The model’s parameters are estimated partly using numerical optimization and partly with ABC. Our estimation procedure provides a good fit to real data.

Key words: epidemiological models, open data, Approximate Bayesian Computation, uncertainty evaluation

1 Introduction

The recent SARS-COV-2 epidemic is the first global pandemic in the big data era. Differently from other past epidemics, it developed even in technologically advanced countries and put the most innovative health systems in crisis. Moreover, this event brought to light different problems related to the quality of data and the related decision-making. Indeed, the public sector in most countries was not ready to collect, validate and distribute open data (Hua and Shaw, 2020) and the lack of statistical knowledge in the citizens and in most of the media led to the inability to clearly distinguish between “data” and “information” (Arbia and Nardelli, 2020) (Zarocostas, 2020). A large number of researchers during the Covid pandemic have...
unsuccessfully required the access to anonymous individual data. Many active research groups, (among which e. g. Covstat (2020)) developed models to predict the trend of the epidemic using all available open data, trying to mitigate the problems due to the poor data quality and to implement and estimate the model’s parameters together with its uncertainty. In the next section we will present our proposal.

2 Model definition

Historically, one of the first model used to predict the spread of the pandemic was the SIR model (Kermack and McKendrick, 1991) based on a system of ordinary differential equations that models 3 categories of population (Susceptible, Infected, Recovered). In any given moment of time $t$, $I(t)$ and $S(t)$ indicate respectively the number of infected people and the number of vulnerable people, while $R(t)$ (removed) represents the total of those who develop immunity (recovered) or died. Obviously in any moment of time we have: $t$, $S(t) + I(t) + R(t) = N$ with $N$ the total population. The SIR model describes the variation of $S(t)$, $I(t)$, and $R(t)$ and the transitions from one category to the other. The original model specification does not consider population mobility in response to possible lockdown measures nor the impact of the asymptomatic. In this paper we propose an adaptation of Khailaie et al. (2020) model which can be applied to model the spread of the epidemic in Italy using the available open-data diffused from Protezione Civile (Palladino et al., 2020).

Our model is based on 6 categories, namely: Susceptible (people that can still be affected by the virus); Infected (people that are currently infected); Hospitalized (people that need a medical treatment in hospital); ICU (people with severe symptoms that need to go to Intensive Care); Recovered (people that recovered from the illness) and Deaths. We will refer to this model with the acronym “SIHCRD”. whose schematic representation is reported in Fig.1.

The model is characterized by six non-linear ordinary differential equations:

$$\frac{dS}{dt} = -\beta(t) \frac{S}{N} I$$
$$\frac{dI}{dt} = \beta(t) \frac{S}{N} I - \gamma I - k_1 I$$
$$\frac{dH}{dt} = k_1 I - k_4 H - k_2 H$$
$$\frac{dC}{dt} = k_2 H - k_3 C - k_3 C$$
$$\frac{dD}{dt} = k_3 C$$
$$\frac{dR}{dt} = k_4 H + k_3 C + \gamma I$$


Modelling COVID-19 evolution in Italy with an augmented SIRD model

where $\beta(t) = \beta_0 \cdot e^{-t/\tau}$. The model is characterized by 8 free parameters. From the previous equations, we have $S + I + H + C + D + R = \text{constant}$. In what follows we describe the parameters in details.

- $\beta$ is related to the spread of the infection. Larger values of $\beta$ corresponds to a fast spread of the epidemic;
- $\gamma$ is related to the (inverse of) time necessary to move from the category “infected” to the category “recovered”, without passing through hospital;
- $k_1$ is the product between the fraction of infected people that need to go to hospital (roughly 5% in the Italian experience) times the inverse of the average time required to move from “infected” to “hospitalized”;
- $k_2$ denotes the product between the fraction of hospitalized people that need to go to intensive care units (roughly 10% in Italy) times the inverse of the average time required to move from “hospitalized” to “intensive care units”;
- $k_3$ denotes the product between the fraction of patients that die times the average time that they stay in ICU (Intensive Care Unit) before the death;
- $k_4$ denotes the product between the fraction of people that do not go to ICUs (roughly 90%) times the inverse of the average time required to recover;
- $k_5$ denotes the product between the fraction of people that do not die in ICUs (roughly 70% during the second pandemic wave in Italy) times the average time required to recover;
- $\tau$ denotes the timescale of the decreasing of the parameter $\beta$.

When $k_1 = 0$ we go back to the original SIR model. The model contains some working hypotheses. The first is that people die only in ICUs. The second is that once a patient is recovered is removed from the susceptible, it it cannot be infected again.
Vincenzo Nardelli, Giuseppe Arbia, Andrea Palladino and Luigi Giuseppe Atzeni (Wajnberg et al., 2020). As in the original SIR model the basic reproduction number, usually called $R_0$, can be obtained combining some of the previous parameters as follows:

$$R_0 = \frac{\beta_0}{\gamma + k_1}$$

(2)

The evolution of $R_0$ through time, usually indicated as $R_t$, is instead given by:

$$R_t(t) = \frac{\beta(t)}{\gamma + k_1} \frac{S(t)}{N}$$

(3)

The basic reproduction number gives important indications about the behavior of the epidemic. Indeed, when $R_t > 1$ the epidemic is expanding, when $R_t = 1$ the epidemic has reached the maximum, while for $R_t < 1$ the epidemic is in the decreasing phase.

3 Fitting procedure

In our study we fitted model (1) to the real Italian data during the second wave of the epidemic in the period October 1st and November 15th 2020. For the initial number of infected people we assumed the 6 million, estimated by Bassi et al. (2020). All the others variables in Equation (1) are initialized according to the data available for the previous day. In the loss function, we assign the same weights of the errors (absolute percentage error) to hospitalized (H), patients in ICU (C) and deaths (D), while we don’t use the number of infected people. This procedure allows us to make a more reliable estimate given the large uncertainties in evaluating the number of positive individuals and the irregularities in the transmission of data and in the testing procedures. A mixed approach was used to estimate the model parameters. For the parameters involving the transition between the categories of infected, the model was fitted through numerical optimization starting from the estimates published by the Istituto Superiore di Sanità (ISS) \(^1\) and other studies such as Richardson et al. (2020) and Grasselli et al. (2020). In particular, we used the optimizer algorithm Nelder-Mead (Gao and Han, 2012) implemented in SciPy (Virtanen et al., 2020) to tune 6 parameters of the model. In Table 1, we report the result of this optimization.

The remaining parameters ($\beta$ and $\tau$) were estimated through the Approximate Bayesian Computation - Sequential Monte Carlo (ABCSMC) (Toni and Stumpf, 2010) which allows to evaluate the uncertainty of the parameters considering the SIHCRD model as a black-box, starting from non-informative prior. In particular, the prior distribution was assumed to be Uniform between 0 and 1 for the $\beta$ parameter and uniform varies between 0 and 600 for the $\tau$ parameter. Fig. 2, shows the credible intervals of the estimations using PyABC (Klinger et al., 2018) with a population size of 400 and the stopping rule with minimum error set to 1.5%. In Table 2, we report the estimation of the parameters from the posterior distribution.

1 https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia
Table 1 Parameter fit after numerical optimization

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fitted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>0.200</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0.008</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.030</td>
</tr>
<tr>
<td>$k_3$</td>
<td>0.172</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.119</td>
</tr>
<tr>
<td>$k_5$</td>
<td>0.118</td>
</tr>
</tbody>
</table>

Fig. 2 Estimates and credible interval of the parameters $\beta$ and $\tau$

Table 2 Summary of the posterior distributions

<table>
<thead>
<tr>
<th>Quantile</th>
<th>0.025</th>
<th>0.5</th>
<th>0.975</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.33</td>
<td>0.34</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>$\tau$</td>
<td>123.73</td>
<td>228.66</td>
<td>326.67</td>
<td>226.20</td>
</tr>
</tbody>
</table>
4 Estimation results

Details about the proposed model are published online\(^2\) where the results are constantly updated. During the second wave, the average error of the fit never exceeded 10\% despite the great irregularity in the official data. The results are in agreement with other models published during the second wave.

The parameters’ estimation with ABCSMC allows to model the uncertainty. An example of the fit of the hospitalized curve (H) is reported in Figure 3.

![Fit H](image)

**Fig. 3** True and estimated of the posterior distribution of hospitalized patients (H)

5 Conclusions

In this paper we propose a mixed strategy to estimate the parameters of an augmented SIRD model combining numerical optimization and ABC procedures. In this way we can calculate credible intervals for the crucial epidemic parameters thus helping their interpretation and their use in the monitoring and surveillance of the pandemic diffusion.

\(^2\) https://dashboard.covstat.it/
Modelling COVID-19 evolution in Italy with an augmented SIRD model

References


Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jar-
