

Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model

Elena Loli Piccolomini^b, Fabiana Zama^a

^a*Department of Mathematics, University of Bologna*

^b*Department of Computer Science and Engineering, University of Bologna*

Abstract

In this paper we propose a Susceptible-Infected-Exposed-Recovered-Dead (SEIRD) differential model for the analysis and forecast of the COVID-19 spread in some regions of Italy, using the data from the Italian Protezione Civile from February 24th 2020. In this study investigate an adaptation of the model. Since several restricting measures have been imposed by the Italian government at different times, starting from March 8th 2020, we propose a modification of SEIRD by introducing a time dependent transmitting rate. In the numerical results we report the maximum infection spread for the three Italian regions firstly affected by the COVID-19 outbreak (Lombardia, Veneto and Emilia Romagna). This approach will be successively extended to other Italian regions, as soon as more data will be available.

1. Introduction

The recent diffusion of the COVID-19 Corona virus has renewed the interest of the scientific and political community in the mathematical models for epidemic. Many researchers are making efforts for proposing new refined models to analyse the present situation and predict possible future scenarios.

With this paper we hope to contribute to the ongoing research on this topic and to give a practical instrument for a deeper comprehension of the virus spreading features and behaviour.

We consider here deterministic models based on a system of initial values problems of Ordinary Differential Equations (ODEs). This theory has been studied

since about one century by W.O. Kermack and A. G. MacKendrick [1] that proposed the basic Susceptible-Infected-Removed (SIR) model. The SIR model and its later modifications, such as Susceptible-Exposed-Infected-Removed (SEIR) [2] are commonly used by the epidemic medical community in the study of outbreaks diffusion. In these models, the population is divided into groups. For example, the SIR model groups are: Susceptible who can catch the disease, Infected who have the disease and can spread it, and Removed those who have either had the disease, or are recovered, immune or isolated until recovery. The SEIR model proposed by Chowell et al. [3] considers also the Exposed group: containing individuals who are in the incubation period.

The evolution of the Infected group depends on a key parameter, usually denoted as R_0 , representing the basic reproductive rate. The value of R_0 can be inferred, for example, by epidemic studies or by statistical data from literature or it can be calibrated from the available data. In this paper we use the available data for determining the value of R_0 best fitting the data.

Compared to previous outbreaks, such as SARS-CoV or MERS-CoV [4], when the disease had been stopped after a relatively small number of infected people, we are now experimenting a different situation. Indeed the number of infected people grows exponentially, and apparently, it can be stopped only by a complete lockdown of the affected areas, as evidenced by the COVID-19 outbreak in the Chinese city of Wuhan in December 2019.

Analogously, in the Italian case, in order to limit the virus diffusion all over the Italian area, the government has started to impose more and more severe restrictions since March 6th 2020. Hopefully, these measures will affect the spread of the COVID-19 virus reducing the number of infected people and the value of the parameter R_0 .

The introduction of different levels of lockdown require an adaptation of the standard epidemic models to this new situation. Some examples about the Chinese outbreak can be found in [5, 4, 6]. Concerning the Italian situation, which is currently evolving, it is possible to model the introduction of restricting measures by introducing a non constant infection rate [7]. In this paper we propose

to represent the infection rate as a piecewise function, which reflects the changes of external conditions. The parameter R_0 , which is proportional to the infection rate, becomes a time dependent parameter R_t which follows a different trend each time the external conditions change, depending on the particular situation occurring in that period. For example, if new restrictions are applied to the population movements at time t_1 , we can hopefully argue that R_t starts to decrease when $t > t_1$.

Finally, we believe that relevant information is not only Infected but also Recovered and Dead numbers we modified SEIR model by splitting the Removed population into Recovered and Dead.

In section 2 we describe the details of SEIRD model with constant and with time dependent infection rate SEIRD(rm). Finally in section 3 we test the model on some regional aggregated data published by the Protezione Civile Italiana [8].

2. Numerical model and methods

The SIR model proposed by Kermack and A. G. MacKendrick [1] divides the population in three groups: Susceptible (S), Infected (I) and Recovered(R). The equations relating the groups are the followings:

$$\begin{aligned}\frac{dS}{dt} &= \frac{\beta}{N}SI \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

where N is the total population, β is the infection rate, a coefficient accounting for the susceptible people get infected by infectious people and γ is the parameter of infectious people which become resistant per unit time. A more refined model is the SEIR model where a new compartment E representing the exposed individuals that are in the incubation period is added.

The resulting equations in the SEIR model are the following:

$$\begin{aligned}\frac{dS}{dt} &= \frac{\beta}{N}SI \\ \frac{dE}{dt} &= \frac{\beta}{N}SI - \alpha E \\ \frac{dI}{dt} &= \alpha E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

where α represents the incubation rate. The difference between the exposed (E) and infected (I) is that the former have contracted the disease but are not infectious, and the latter can spread the disease. SEIR has been used to model breakouts, such as Ebola in Congo and Uganda [3]. In [2] the equations are modified by adding the quarantine and vaccination coefficients. In our case, unfortunately, vaccination is not available. A further model, the SIRD, considering the group of Dead (D) in place of the Exposed is analysed in [7] for the forecast of COVID 19 spreading.

In this paper we propose a SEIRD model accounting for five different groups, Susceptible, Exposed, Infected, Recovery and Dead. The system of equations is given by:

$$\begin{aligned}\frac{dS}{dt} &= \frac{\beta}{N}SI \\ \frac{dE}{dt} &= \frac{\beta}{N}SI - \alpha E \\ \frac{dI}{dt} &= \alpha E - (\gamma_R + \gamma_D)I \\ \frac{dR}{dt} &= \gamma_R I \\ \frac{dD}{dt} &= \gamma_D I\end{aligned}\tag{1}$$

In order to consider the restrictions imposed by the Italian government since March 8th 2020, we have partitioned the whole time interval $[0, T]$, considered for the integration of (1), into two sub-intervals: $[0, t_0]$ and $(t_0, T]$ where t_0 corresponds to time when the restrictions start to produce a valuable change

in the data trend. Moreover, since the applied restrictions should decrease the number of contacts between Infected and Susceptible, we model the coefficient β in (1) as a decreasing time dependent function $\beta(t)$. A similar model for the infection rate of SEIR equations can be found in [7], where the function is assumed to have a decreasing exponential form. However, observing the data trend, we believe that β_t has a smoother decreasing behavior and we choose to model it as a decreasing rational function:

$$\beta(t) = \begin{cases} \beta_0 & \text{if } t < t_0 \\ \beta_0 (1 - \rho(t - t_0)/t) & \text{otherwise} \end{cases}, \quad \rho \in (0, 1). \quad (2)$$

In the present work we use a constant value $\rho = 0.75$ but we might calibrate it in future. By substituting $\beta(t)$ (2) in the S and E equations in (1) we obtain SEIRD rational model SEIRD(rm).

We calibrate the parameters of SEIRD and SEIRD(rm) by solving non-linear least squares problems with positive constraints. For example, in the SEIRD model (1), we define the function $\mathbf{u}(t) = (S(t), E(t), I(t), R(t), D(t))$, depending on the vector of parameters $\mathbf{q} = (\beta, \alpha, \gamma_R, \gamma_D)$, and the vector \mathbf{y} of the acquired data at given times $t_i, i = 1, \dots, n$. Let $F(\mathbf{u}, \mathbf{q})$ be the function computing the numerical solution \mathbf{u} of the differential system (1), the estimation of the parameter \mathbf{q} is obtained solving the following non linear least squares problem:

$$\begin{aligned} \min_{\mathbf{q}} \frac{1}{2} \|F(\mathbf{u}, \mathbf{q}) - \mathbf{y}\|_2^2 \\ \mathbf{q} \geq 0 \end{aligned} \quad (3)$$

where we introduce positivity constraints on \mathbf{q} . The constrained optimization problem is solved with a trust-region based method implemented in the `lsqnonlin` Matlab function. For further details about the optimization problem for identification parameters in differential problems see for example [9].

3. Numerical Results

In this section we report the results obtained by using the SEIRD model to monitor the Covid-19 outbreak in Italy during the period 24/02/2020-20/03/2020.

The epidemic spread started on February 21st affecting the northern regions. Lombardia in particular registered the first epidemic outbreak followed by the Veneto region, Emilia Romagna the other Italian regions.

Since initially each region applied different containment measures to some restricted areas at different times, we chose to calibrate the SEIRD model on each region separately. This study considers Lombardia, Veneto and Emilia Romagna regions for which the largest amount of meaningful data have been collected in the GitHub repository [8].

The purpose is to develop a model calibration and simulation method to be eventually extended to all the other Italian regions reached by the epidemic spread. All computations are performed using Matlab R2019b 2,9 GHz Intel Core i7 quad-core 16 GB ram. The method consist of two main steps:

- **Identification.** In this step the model parameters are estimated by means of the `trust-region-reflective` algorithm implemented in the `lsqnonlin` matlab function.
- **Simulation.** This step applies the SEIRD-SEIRD(rm) models with the identified parameters in order to monitor the process for a longer time (up to 240 days). The differential system is solved applying the `ode45` matlab function with the following initial condition: $S(0) = N, E(0) = I(0) = I_{init}, R(0) = D(0) = 0$ where the value I_{init} corresponds to the Infected individuals in the first measurement day.

3.1. SEIRD simulation

In this section we consider the SEIRD model (1) and use the following different measurement subsets from [8] relative to Lombardia region:

- **S1** 10 days measurements: 24/02/2020-04/03/2020
- **S2** 18 days measurements: 24/02/2020-12/03/2020
- **S3** 23 days measurements: 24/02/2020-17/03/2020

for identification of the parameters and then apply the identified parameters to model the Infected-Recovered-Dead populations up to June 22nd (120 days). Besides population plots in figures 1 ,2 we collect some meaningful quantitative information about the model parameters (table 1) and the peak values for Infected, Recovered and Dead populations (table 2).

The results can be summarized as follows:

Measured days	β	α	γ_R	γ_D	R_0
S1	0.74	0.28	0.0632	0.0205	8.8
S2	0.29	3.28	0.0473	0.0272	3.9
S3	0.29	3.22	0.0560	0.0354	3.1

Table 1: SEIRD Model parameters, Lombardia Region

Data data	Peak Infected		Peak Recovered		Peak Dead	
	day	Number (%)	day	Number (%)	day	Number (%)
S1	46	4674665 (46.4%)	120	7601989 (75.4%)	120	2466103 (24.5 %)
S2	60	3931157 (39.0%)	120	6190842 (61.4%)	120	3554881 (35.2%)
S3	65	3108534 (30.8%)	120	5792528 (57.4%)	120	3664115 (36.3%)

Table 2: SEIRD, Infected, Recovered Dead values. Lombardia region.

- The maximum Infected population is reached on April 9th 2020 for S1, April 23rd 2020 for S2 and April 28th 2020 for S3.
- Concerning the values of the model parameters we have that the transmitting rate β can be estimated as $\beta_e = 0.3$, the incubation rate α is approximately $\alpha_e = 3$ while the recovery rate is approximately $(\gamma_R)_e = 0.06$ and the dead rate can be approximated as is approximately $(\gamma_D)_e = 0.04$.

The following conclusions can be drawn:

- The SEIRD parameters, computed from the first 10 days measurements, do not model properly the measurements. Indeed the plots reported in figure 1 show that data of IRD populations have a slower increase rate, compared to the model predictions.
- Considering the parameters obtained from S3 we observe that even if the data fit is improved, the model does not seem to reproduce properly the increasing slope of the Infected data, hence the previsions from this model may not be entirely reliable.

Concerning Emilia Romagna region we have similar results. The plots in figures 3,4 show that the Infected population increase rate does not reproduce the data correctly hence the peak values reported in table 3 may not be accurate.

The SEIRD model seems to be more accurate for Veneto region where the plots

Data	Peak Infected		Peak Recovered		Peak Dead	
data	day	Number (%)	day	Number (%)	day	Number (%)
S1	48	3565588 (80.0)	240	968332 (21.7)	240	3345752 (75.0)
S2	57	3207633 (71.9)	240	1217341 (27.3)	240	3191080 (71.6)
S3	62	2729893 (61.2)	240	1272248 (28.5)	240	3179219 (71.3)

Table 3: SEIRD, Infected, Recovered Dead values, Emilia Romagna region.

in figure 5 reproduce the data quite accurately. The peak values reported in table 4 represent the largest percentage of Infected and the longest time required to reach the peak.

3.2. *SEIRD(rm) calibration and simulation*

In order to improve the data fit we split the parameter identification step into the following two phase process:

- **Phase 1** Identification of the parameters of the standard SEIRD model using a the data subset S1, (i.e. up to time $t_0 = 10$).

Data subset	Peak Infected		Peak Recovered		Peak Dead	
	day	Number (%)	day	Number (%)	day	Number (%)
S1	71	4024199 (82.0)	240	2107711 (43.0)	240	1966666 (40.1)
S2	77	3661977 (74.6)	240	2835238 (57.8)	240	1685415 (34.4)
S3	78	3547164 (72.3)	240	2913752 (59.4)	240	1698521 (34.6)

Table 4: SEIRD, Infected, Recovered Dead values, Veneto region.

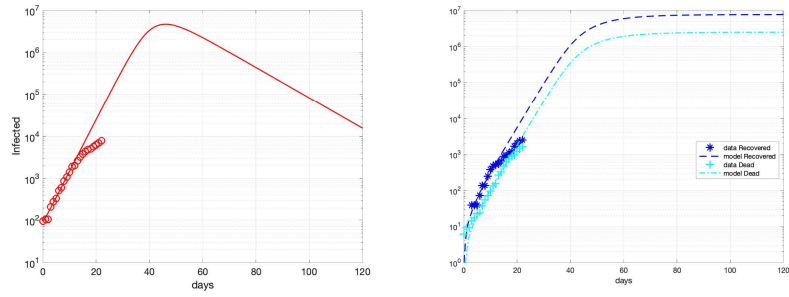


Figure 1: Infect, Recovered Dead model based on Lombardy data S1.

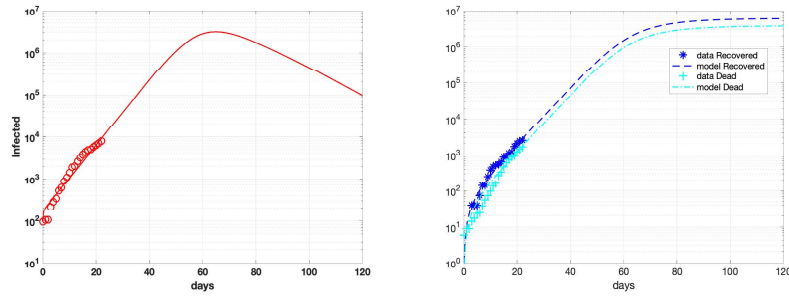


Figure 2: Infect, Recovered Dead model based on Lombardy data S3.

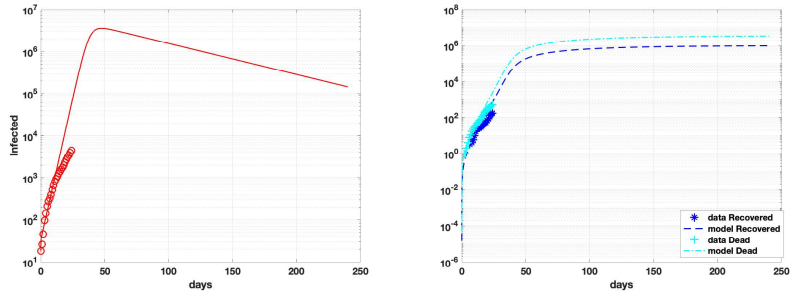


Figure 3: Infect, Recovered Dead model based on Emila Romagna data S1.

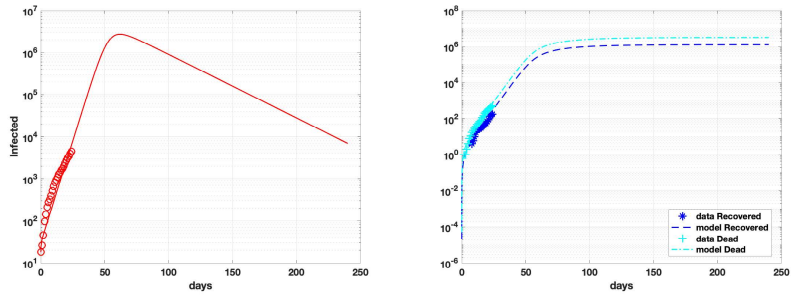


Figure 4: Infect, Recovered Dead model based on Emila Romagna data S3.

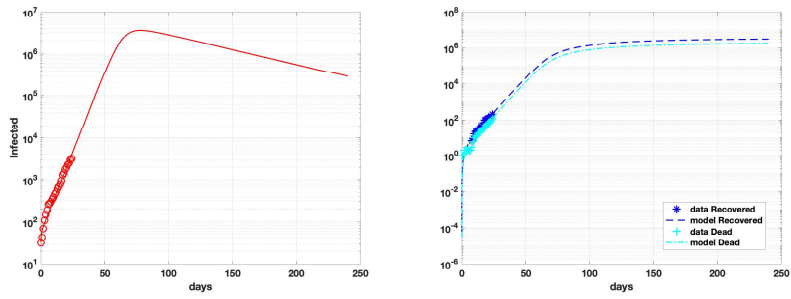


Figure 5: Infect, Recovered Dead model based on Veneto data S3.

- **Phase 2** Identification of the parameters of SEIRD(rm) model using all the measures available and modelling the infection rate as in (2). The value of the infection rate is constant when $t \leq t_0$ otherwise it has decreasing values.

The parameters obtained from the of the Phase 2 of each region are reported in table 5. Representing the infection rate as a time dependent function causes the

Region	α	γ_R	γ_D
Lombardia	0.41	0.040	0.025
Veneto	2.5	0.0069	0.0030
Emilia Romagna	1.35	0.011	0.025

Table 5: SEIRD(rm) parameters.

change of the Reproduction parameter (R_0) into a time dependent Reproduction function defined as follows:

$$R_t = \frac{\beta(t)}{\gamma_D + \gamma_R} \quad (4)$$

The plots of R_t for the three regions are reported in figure 6. In table 6 we

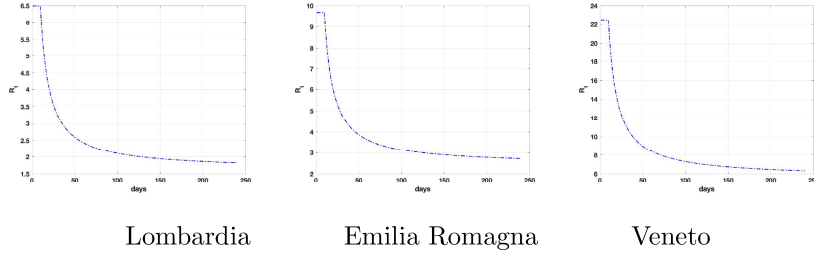


Figure 6: Reproduction Rate R_t of SEIRD(rm) model.

report the peak values of the IRD populations for the proposed SEIRD(rm) model. We observe that the Infected people should reach its peak on June 21-22 in Lombardia and Emilia Romagna while Veneto requires the longest time (August 1st 2020) and it seems to have the largest number of infected people. However, we highlight that Veneto applies a different testing policy with respect

Data	Peak Infected		Peak Recovered		Peak Dead	
Region	day	Number (%)	day	Number (%)	day	Number (%)
Lombardia	107	1615368(16.0%)	240	5047743(50.1%)	240	3116082(30.9%)
Emilia Romagna	109	1465451(32.9%)	240	1260782(28.3%)	240	2890963(64.8%)
Veneto	160	2833884(57.8 %)	240	2208955(45.0%)	240	979229(20.0%)

Table 6: Infected, Recovered Dead values SEIRD(rm) model

to the other regions which tests a larger part of population. Therefore the Infected people measured in Veneto contain a greater number of individuals who do not have symptoms but can spread the infection. Following a recent study on Wuhan Covid outbreak [10] the undocumented infectious can have a great influence in the virus spread. These undocumented infectious often experience mild, limited or no symptoms and hence go unrecognized, and, depending on their contagiousness and numbers, can expose a far greater portion of the population to virus than would otherwise occur.

The improved modeling properties can be appreciated in population plots reported in figures 7 , 8 and 9. We observe that SEIRD(rm) reproduces the data trends more precisely compared to the SEIRD model.

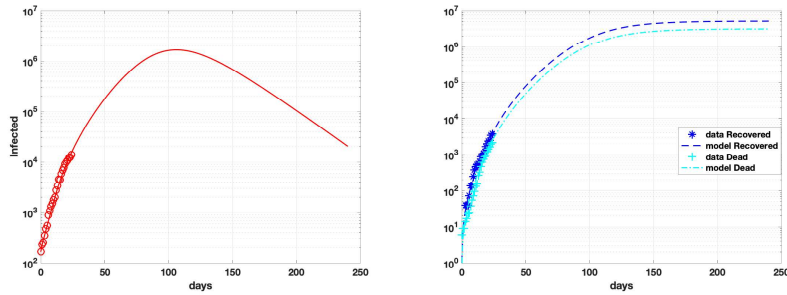


Figure 7: Infect, Recovered Dead model based on Lombardia region data (SEIRD(rm)).

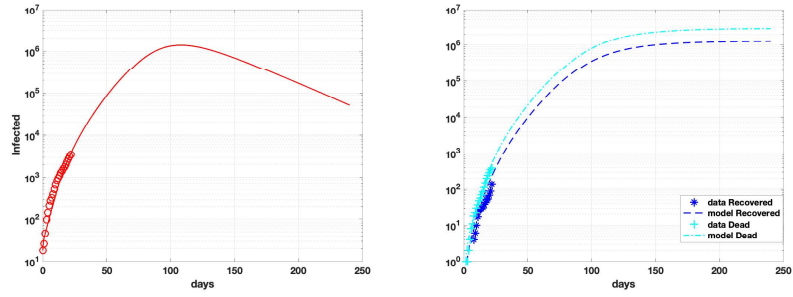


Figure 8: Infect, Recovered Dead model based on Emilia Romagna region data (SEIRD(rm)).

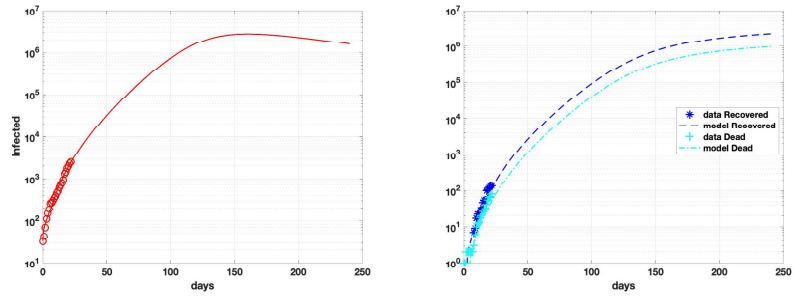


Figure 9: Infected, Recovered Dead model based on Veneto region data (SEIRD(rm)).

4. Conclusion

In this paper we proposed a SEIRD model for the analysis of the COVID-19 outbreak diffusion in Italy. In our new formulation the infection rate coefficient has been adaptively modeled as an inverse function of the time, to take account the restrictions imposed by the Italian Government in the social life on March 8th, 2020. The results obtained by fitting the data of three Italian regions, Lombardia, Emilia-Romagna and Veneto, available since February 24th 2020, show a very good fit to the data and give a prediction on the behaviour of Infected-Re individuals. The forecasts about the maximum infection spread report quite homogeneous results for both Lombardia and Emilia Romagna (about June 20th), while Veneto has its infection peak around the end of July 2020, probably due to different testing modalities.

We highlight that it is only 12 days since restrictions started in Italy and, maybe in the next few days, the effects of such measures will become more evident, hopefully causing a further decrease in the infection trend. In this case, the previsions shown in this paper should be updated by introducing a new time t_1 at which the decreasing slope of β_t should change, for example by estimating the parameter ρ in (2) with new data.

The proposed model is flexible and we believe it could be easily adapted to monitor various infected areas with different restriction policies.

References

- [1] K. W.O., M. A.G., A contribution to the mathematical theory of epidemics, in: Proceedings of the Royal Society of London, Vol. A, 1927, pp. 700–721.
- [2] H.Boujakjian, Modeling the spread of ebola with seir and optimal control, in: SIAM Under- graduate Research Online, no. 9, SIAM, 2016, pp. 299–310.
- [3] G. Chowell, N. Hengartner, C. Castillo-Chavez, P. Fenimore, J. Hyman, The basic reproductive number of ebola and the effects of public health

- measures: the cases of congo and uganda, *Journal of Theoretical Biology* (229) (2004) 119–126.
- [4] A. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. Eggo, Early dynamics of transmission and control of covid-19: a mathematical modelling study, *Lancet* (2020). doi:[https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).
 - [5] J. Wu, K. Leung, G. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, *Lancet* (395) (2020) 689–697. doi:[https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
 - [6] Z. Tanga, X. Lib, H. Lic, Prediction of new coronavirus infection based on a modified seir model. doi:medRxiv preprint doi: <https://doi.org/10.1101/2020.03.03.20030858>.
 - [7] D. Fanelli, F. Piazza, Analysis and forecast of covid-19 spreading, https://www.researchgate.net/publication/339943580_Analysis_and_forecast_of_COVID-19_spreading_UPDATE.
 - [8] D. Repository, <https://github.com/pcm-dpc/COVID-19>.
 - [9] F. Zama, D. Frascari, D. Pinelli, A. E. M. Bacca, Parameter estimation algorithms for kinetic modeling from noisy data, in: *System Modeling and Optimization*, Springer International Publishing, 2016, pp. 517–527.
 - [10] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov2), *Science* (2020). doi:10.1126/science.abb3221.