Dynamical modeling, calibration and robustness analysis of COVID-19 using Italian data

Chiara Antonini ICT4life Srl Perugia, Italy chiara.antonini.21@gmail.com

Claudio Dario Regional Government of Umbria Perugia, Italy cdario@regione.umbria.it Sara Calandrini ICT4life Srl and University of Perugia Perugia, Italy sara.calandrini@gmail.com Fabrizio Stracci University of Perugia Perugia, Italy fabrizio.stracci@unipg.it

Fortunato Bianconi, IEEE Member COVID19 epidemiological unit, Regional Government of Umbria Perugia, Italy fortunato.bianconi@gmail.com

Abstract-Italy was the first European country severely hit by the COVID-19 pandemic. In late February and March 2020, the number of people requiring hospitalization and mechanical ventilation has soared, putting a strain on the Italian health system. In the absence of pharmaceuticals therapies, the government implemented a set of mobility restrictions for transmission containment. Starting from the need of predicting hospitalization and ICU rates for the Umbria region in Italy, we propose the application of a computational framework to model the epidemic and analyze the effects of the imposed lock-down. We calibrate a compartmental model of COVID-19 clinical progression using a Bayesian method called Conditional Robust Calibration (CRC) against the daily epidemiological data. Then, we perform a robustness analysis on the calibrated model, in order to quantify the influence of model parameters on the hospital capacity and to draw possible scenarios of different containment measures. CRC confirms the hypothesis of underestimation of new positive cases and highlights how identifying presymptomatic transmission is crucial for reducing the contagion. Moreover, our results show the central importance of the lock-down timeliness and intensity, in order to curb the contagion and avoid a relapse.

Index Terms—Epidemiological model, Bayesian calibration, Robustness analysis, COVID-19, Italy.

I. INTRODUCTION

T HE new severe acute respiratory syndrome-coronavirus (SARS-CoV-2), the causative agent of COVID-19, was identified in China in December 2019 and was characterized as a pandemic virus on March 2020 by the World Health Organization (WHO) [1]. In Italy, the European country most affected by COVID-19, a cluster of pneumonia cases was detected on February 2020 [2]. The levels of spread and severity caused the implementation of drastic restrictions by the Italian government, like social distancing and total lock-down, imposed on 9 March [2]. Given this scenario, there is the urgent need to represent and predict the evolution of the outbreak in order to face the emergency and to make informed decisions.

In this context, mathematical models help to understand the spread mechanism and the dynamic of the pandemic, providing estimates of effectiveness of different interventions. [3]–[5].

Since the diffusion of COVID-19 is complex depending on many parameters and factors, different models are designed and implemented starting from traditional models used in the context of epidemiology. Many models are based on the so called Susceptible-Exposed-Infectious-Removed (SEIR) model, which divides the population into these four classes. To describe the outbreak in China, an extended SEIR is used in [6], while in [7] an estimation of the basic reproduction number, of the infection mortality and recovery rates is provided with a Susceptible-Infectious-Recovered-Dead (SIDR) model.

In this work, our attention is focused on the case study of Italy. To represent the disease evolution in Italy, in [4], the SIDARTHE model is used to distinguish infected individuals based on diagnosis while in [3] a metacommunity (SEIR)like transmission model is employed to estimate the effects of restrictions on the human mobility and contacts. Here, to represent the COVID-19 dynamics in Italy, we use a compartmental epidemiological model taken from [8] and presented also in [9], based on the classic SEIR model with lock-down (L) measures (SEIRL), which takes into account both asymptomatic and presymptomatic transmission.

This study started from the request to provide estimations on hospitalization and Intensive Care Unit (ICU) rates for the Umbria region in Italy, in order to support the Umbria Health Government in making decisions regarding staff resources and hospital beds. In this work, we apply the modeling strategy developed for the Umbria region to estimate model parameters for Italy. We present the results of the Bayesian method called Conditional Robust Calibration (CRC) [10]-[12], also applied to the Umbria region. Then, for Italy, we perform a Conditional Robustness Analysis (CRA) through our CRA Toolbox, in order to understand the influence of epidemiological parameters and progressive restriction measures on the hospitalization capacity and logistics [13], [14]. Our analysis shows how the combination of model calibration and robustness analysis can be used as a support to make reliable predictions on the scale of the outbreak and on the evaluation of the timescale and impact of interventions.

452

II. METHODS

A. Mathematical model

The SEIRL model used in this work, shown in Fig.1, takes into account different clinical stages of the infection. Susceptible individuals (class S) start out in an exposed (class E) where they develop the infection but do not transmit it. Eindividuals enter the presymptomatic class (class P_S) since they may be able to transmit the virus before developing symptoms. The other classes are: asymptomatic infection (class A), mild infection (class M), severe infection (class H, hospital stay), critical infection (class ICU) and recover (R) or die (class D). The model assumes that only people in a critical stage die and that all individuals have equal susceptibility to infection. The total population size $N = S + E + P_S + A + M + M$ H + ICU + R + D is supposed to be constant. The dynamical system is represented by the Ordinary Differential Equations (ODEs) described in [8] (see Supplementary Materials at http://gitlab.ict4life.com/chiara.antonini/model_covid19).

Model parameters represent the transmission rates at the different stages of the disease $(b_i \ i = e, 0, 1, 2, 3)$, the different recovery rates $(q_i \quad i = 0, 1, 2, 3)$ and the death rate (u), the rate of exit from classes E and P_S , $(a_i \quad i = 0, 1, \text{ re-}$ spectively), the progression rate from mild to severe infection and from severe to critical infection $(p_1 \text{ and } p_2, \text{ respectively})$. These rates can be derived from the following input parameters according to the formulas in [8] (see Supplementary): Presym-Period is the length of the infectious phase of incubation period (IncubPeriod, days), DurMildInf and DurAsym are, respectively, the average duration of mild and asymptomatic infections (days). FracSevere, FracCritical and FracAsym are the average fraction of severe, critical and asymptomatic infections. DurHosp is the average duration of hospitalizations for people with severe infection while TimeICUDeath is the average duration of ICU admission (days) and CFR is the case fatality rate.

Moreover, the model takes into account the possibility of an intervention in order to reduce infection transmission. We introduce two more parameters, s_0 and s_1 , representing the reduction in transmission from presymptomatic/asymptomatic infection and mild infection, respectively. The employment of personal protective equipment (PPE) in hospitals may also reduce the probability of infection, represented with parameter s_2 . These parameters decrease transmission rate constants as follows: $b_e = b_e \cdot s_0$, $b_i = b_i \cdot s_i$, i = 0, 1, 2, $b_3 = b_3 \cdot s_2$.

B. Data

We perform parameter estimation against hospitalized, ICU and dead patients (H, ICU, D), which represent the most reliable measures. Indeed, we suppose that the number of mild infections is underestimated, due to under registration and testing only patients with symptoms or pneumonia [3], [16]. Data are available on the GitHub repository of the Italian Civil Protection Department [15].

C. Model calibration and conditional robustness analysis

Here, we summarize the main aspects of the methods used for model calibration and robustness analysis. CRC is an Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) approach for nonlinear model calibration that considers the parameter vector as a random variable [12]. It is an iterative algorithm based on the sampling of a proposal distribution and on the definition of multiple objective functions, one for each observable. CRC estimates the probability density function of parameters conditioned to the experimental measures and compared to other ABC-SMC approaches it has a reduced computational cost. After model calibration, we perform the Conditional Robustness Analysis (CRA) on model parameters, using the CRA Toolbox [13], [14]. The CRA measures the influence of parameters on a specific variable through computation of the Moment Independent Robustness Indicator (MIRI). Parameters with an high MIRI value have a significant variation on the selected output variable, represented through an evaluation function. Here, we use the CRA to evaluate the impact of model parameters on the hospitalization capacity, i.e. on H and ICU. More details of CRC and CRA can be found in [10], [12], [14].

III. RESULTS

First of all, we calibrate the ODE model using the data from 24 February to 3 May 2020, assuming that the initial day of virus introduction is 30 days prior to the first ten registered deaths [16]. We also introduce the following interventions: 21 February 2020, creation of two quarantined red areas under strict lockdown in Lombardy and Veneto; 24 February 2020, school closure in most regions in the Northern of Italy (Lombardy, Veneto, Emilia-Romagna, Friuli Venezia Giulia, Liguria, Piedmont and part of Marche); 5 March 2020, school closure in the entire country; 8 March 2020, total lock-down area in the Northern of Italy; 10 March 2020, total lock-down area extended to all Italian regions. All these containment measures are implemented in the model through parameter vector $s_0 = [s_{01}, s_{02}, s_{03}, s_{04}]$. Since the lock-down was imposed at a distance of two days in the Northern and remaining part of Italy, we include both of them in a unique parameter (s_{04}) , considering also the delay in social acceptance of the restrictive intervention. Parameter s_1 represents the reduction of transmission of mild infected people due to a total ban on leaving their houses while parameter s_2 implements the employment of PPE in hospitals.

A. CRC results

The parameter vector to estimate for Italy consists of fifteen model parameters and six intervention parameters, i.e. $\mathbf{p} \in \mathbb{R}^{21}$, while for Umbria, there are only four intervention parameters to estimate (school closure and total lock-down), i.e. $\mathbf{p} \in \mathbb{R}^{19}$. The tuning parameters of CRC are set as follows: the number of samples in the parameter space is set to $N_S = 10^5$ for each iteration; the number of iterations is



Fig. 1: Graphic representation of SEIRL model. Clinical stages of infection are: Susceptible (S), Exposed (E), Presymptomatic (P_S) , Asymptomatic (A), Recovered (R), Mild infection (M), Severe infection (H), Critical infection (ICU) and Dead (D). Control measures: Lock-down measures (L) and personal protective equipment (PPE).

equal to 8; the number of realizations performed is set to 10, to ensure reliability of results.

Fig.2 depicts the behavior of output variables for Italy, when parameters are equal to the mode of the approximate posterior distribution returned by CRC in one of the final realizations (see Supplementary). Model simulations are shown together with the data, proving that CRC achieves consistent results for all variables. Moreover, we can see that there is a significant shift between mild infections predicted by the model and the observed data, because people were tested many days after the onset of the first symptoms, especially during the most critical phase of the epidemic.

Using the mode vector, the estimate of the initial reproduction number R_0 is equal to 4.65. We also estimate that containment measures have progressively reduced the presymptomatic and asymptomatic transmission by 40% in the first phase up to 90% during the total lockdown. On the other hand, quarantine and isolation of mild infected and wearing PPE have taken to a reduction in transmission of about 85% and 80%, respectively. As shown in Supplementary, the presymptomatic transmission rate is estimated much higher with respect to all the other rates, meaning that people in the presymptomatic phase play a central role in the COVID-19 spread.

Fig. 3 shows the result of the application of CRC to the Umbria region. Also in this example, the method is successful in replying the behavior of H, ICU and D while we notice the same shift as before for mild infected people. This mismatch finds a complete confirmation in Fig. 4 where the model prediction of new positive cases is in agreement with the data of the Istituto Superiore di Sanità (ISS). Indeed, the peak of new cases predicted by the model and the peak of symptom onset registered by the ISS are both around mid-March. The estimate of R_0 for Umbria region is equal to 5. CRC estimates similar values for process parameters for both Umbria and Italy (see Supplementary).

As regards the computational cost, CRC takes around 1 hour to complete ten realizations of one iteration. All the simulations are performed using Matlab (R2019a) on a Intel Core i7-4700HQ CPU, 2.40GHz 8, 16-GB memory, Ubuntu 18.04 LTS (64 bit).

B. CRA results

After model calibration, we run the CRA algorithm on the Italy case, through the CRA Toolbox software [14], in order to identify those model parameters that influence most the healthcare capacity. For this purpose, we choose as evaluation function the area under the curve of H. The lower and upper boundaries of the sampling intervals for the parameter space are fixed equal to the lower and upper 90 percentile of the final pdf estimated by CRC (see Supplementary). We perturb the parameter space with Linear LHS generating 10^4 samples and we set equal to 1000 the dimension of the upper and lower tail of the evaluation function pdf, in order to guarantee a stable estimation of the conditional parameter pdfs [13]. We perform 10 realizations of the entire procedure to ensure invariance and stability of results.

Through the CRA, we analyze two different temporal scenarios: a short one from the end of January until mid-May and a longer one until the end of October. In the second case, we add two more intervention parameters for asymptomatic and presymptomatic transmission, i.e. s_{05} and s_{06} , in order to take into account the progressive lock-down release implemented by the Italian government in May. In more detail, from May 4 some businesses were allowed to reopen while from May 18 lock-down measures were almost completely lifted, maintaining only social distancing, mandatory PPE in closed places and ban of most public events. Because these two parameters are not estimated, their range of variation is set to [0.1-0.9]. Since there is a pretty long delay between the implementation of social distancing measures and the decline or rise in cases and deaths, depending on whether these



Fig. 2: Italy case. Time behavior of state variables when the parameter vector is equal to the final mode vector computed by CRC (black line); dots represent the data [15]. Data and simulations are in log-scale, normalized over the whole Italian population (~ 60 million) and multiplied by 10^5 . The colored area reproduces the variation of the temporal behavior when parameters vary between the 60, 70 and 90 confidence intervals of their corresponding conditional pdfs (see Supplementary). Time starts from 27 January 2020.



Fig. 3: Umbria case. Time behavior of state variables when the parameter vector is equal to the final mode vector computed by CRC (black line); dots represent the data [15]. Data and simulations are in log-scale, normalized over the whole population of the region (~ 882000) and multiplied by 10^5 . The colored area reproduces the variation of the temporal behavior when parameters vary between the 60, 70 and 90 confidence intervals of their corresponding conditional pdfs (see Supplementary). Time starts from 21 February 2020.



Fig. 4: Comparison of new positives cases for Umbria. (a) Epidemic curve estimated by the model in comparison with the observed data of the Italian Civil Protection Department [15]. (b) Epidemic curve based on date of sample and symptom onset estimated by the ISS.

measures are enforced or lifted, we are just starting to see the outcome of these government actions [9]. Through a wide perturbation of parameters s_{05} and s_{06} , we seek to understand the impact of these interventions in the present and future months, when a possible second wave of COVID-19 cases may occur.

Fig. 5 shows MIRI values returned by the CRA in the two different scenarios while in Supplementary the corresponding parameter pdfs are reported. In the first example, MIRIs are strongly influenced by presymptomatic transmission through parameters b_e and PresymPeriod. MIRIs highlight also the importance of parameter s_{02} , which represents the school

closure in the Northern Italian regions about one month after the starting of the epidemic. As regards the second example, the most relevant parameter is s_{06} which can determine or not a relapse of the disease. Indeed, by looking at the pdf in Supplementary, we can see that a value above ~ 0.25 may cause the epidemic to start again with full intensity. In Fig. 6, three different scenarios are presented, varying the strength of the restrictive measure release.

The computational cost of the CRA Toolbox is lower than that of CRC, since the CRA runs ten realizations in around 10 minutes with the same computational power used for CRC.



Fig. 5: Boxplot of the 10 realizations of the MIRIs in output from the CRA for all model parameters.



Fig. 6: Predictions of H and ICU time behavior by varying intervention parameters s_{05} and s_{06} .

IV. DISCUSSION

In this study, we successfully calibrate a SEIRL model of the COVID-19 dynamic against data for Italy and Umbria region, using the Bayesian method CRC.

First of all, our analysis shows the crucial role played by presymptomatic individuals in transmitting the infection. Indeed, the presymptomatic transmission rate b_e is estimated much higher with respect to all the other transmission parameters (see Supplementary), meaning that infected people at this stage of disease notably speeds up the contagion. The other parameter related to presymptomatic transmission is the duration of the presymptomatic period. Given an incubation period of about 5 days estimated by CRC, in accordance with [17], the presymptomatic phase lasts about 3 days, during which infected people may transmit the virus without having symptoms. Moreover, the CRA reveals that these two parameters must be put under control in order to contain the number of hospitalizations. While PresymPeriod cannot be reduced, parameter b_e can be mitigated through face masks, higher hygiene standards and social distancing. According to the conditional pdf of b_e (see Supplementary), the number of hospitalized and ICU patients can be limited trying to keep b_e under 1.05. This result also highlights the urgent need of a policy of population-wide testing and contact tracing in order to detect and isolate presymptomatic cases as soon as possible. Containment measures such as swab testing and centralized quarantines are essential also for confining asymptomatic people, estimated at about 31% of the total infections in Italy (29% in Umbria) by our calibration procedure, consistent with [17], [18]. However, estimating asymptomatic infections is a very challenging task, since it may include also presymptomatic cases who will develop symptoms later [19].

The time behavior of mild infections predicted by our model is in agreement with [17], where it is stated that the median time between onset of symptoms and positive diagnosis ranges between 2 and 6 days. Moreover, CRC estimates a slightly higher number of people with mild symptoms, confirming the fact that released data are strongly dependent on the testing capacity of each Italian region, which was not always able to increase at a similar rate as the epidemic spread and was sometimes impaired by the lack of kits and reagents [3]. In addition, the model correctly predicts the peak of the curve of new positive cases in accordance with ISS data for Umbria region. If the model is correctly specified, then the difference between observed and model based incidence of SARS-CoV-2 infections represents the estimate of undiagnosed and unreported cases.

Then, the CRA applied to Italy reveals the extreme importance of the lock-down timeliness, through the higher value of parameter s_{02} compared to the other intervention parameters. Acting about one month after the start of the epidemic, trying to reduce transmission of at least 50%, is determinant for avoiding the overload of the health system's capacity while a too early lock-down would only postpone the outbreak but not its strength [20]. While these restrictions were adopted at the end of February, we started seeing their effectiveness around April due to clinical and epidemiological characteristics of COVID-19 [9]. Indeed, both the natural progression of the disease and the residual transmission during the lock-down phase, such as household transmission, may generate long delays between the beginning of restriction measures and the observation of their efficacy. This tardiness may impair the analysis of the influence of each containment measure, when they are implemented within a few days between each other. Thus, also the institution of two red areas in Lombardy and Veneto, corresponding to parameter s_{01} , is incisive for reducing the number of hospitalized people. However, its effect is hidden by the successive restrictive measure (parameter s_{02}), since they occur at a distance of three days one from the other. Through the CRA, we are also able to make some predictions about the future evolution of the epidemic, depending on the adoption of different distancing measures. We predict that loosing the lock-down significantly may cause a second wave of COVID-19 spread in the next autumn. However, these longterm predictions are strongly correlated to people's behavior in private and work life and to their contact rates.

Finally, this study has some limitations that, however, do not compromise our main conclusions. The model does not take into account the seasonality in transmission rates, as occurs with other respiratory diseases such as influenza. Furthermore, it must be noted that the Italian health system is highly decentralized, since most administrative and organizational decisions are taken at a regional level [21]. During the pandemic, the cross-regional movement was almost completely prohibited and, thus, a natural future development of this work would be to study each Italian region separately using the same framework, following the example of the Umbria region.

V. CONCLUSION

We presented the application of a new framework, based on the calibration method CRC and on the CRA algorithm, to a SEIRL model using COVID-19 contagion data for Umbria and Italy. CRC allow us to estimate both epidemiological parameters and the effect of progressive restriction measures imposed to the Italian population. Then, robustness analysis evaluates how much each of the model parameters influences the hospital demand.

ACKNOWLEDGMENT

We thank all the components of the COVID-19 epidemiological unit, Department of Health, Regional Government of Umbria (D.G.R. 206/2020, 25/06/2020). We also thank ICT4Life Srl for free support on mathematical modeling and for providing computational resources.

REFERENCES

[1] (2020) WHO timeline of whos response to covid-19. [Online]. Available: https://www.who.int/news-room/detail/29-06-2020-covidtimeline

- [2] (2020) Ministero della Salute website. [Online]. Available: http: //www.salute.gov.it/nuovocoronavirus
- [3] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, "Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures," *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10484–10491, 2020.
- [4] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, "Modelling the covid-19 epidemic and implementation of population-wide interventions in italy," *Nature Medicine*, pp. 1–6, 2020.
- [5] J.-D. Van Wees, S. Osinga, M. Van Der Kuip, M. W. Tanck, and A. Tutu-van Furth, "Forecasting hospitalization and icu rates of the covid-19 outbreak: An efficient seir model," *Bulletin of the World Health Organization*, 2020.
- [6] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S. S. Musa, M. H. Wang, Y. Cai, W. Wang, L. Yang *et al.*, "A conceptual model for the outbreak of coronavirus disease 2019 (covid-19) in wuhan, china with individual reaction and governmental action," *International journal of infectious diseases*, 2020.
- [7] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the covid-19 outbreak," *PloS one*, vol. 15, no. 3, p. e0230405, 2020.
- [8] (2020) modeling covid-19 spread vs healthcare capacity. [Online]. Available: https://alhill.shinyapps.io/COVID19seir/
- [9] A. Nande, B. Adlam, J. Sheen, M. Z. Levy, and A. L. Hill, "Dynamics of covid-19 under social distancing measures are driven by transmission network structure," *medRxiv*, 2020.
- [10] F. Bianconi, C. Antonini, L. Tomassoni, and P. Valigi, "Robust calibration of high dimension nonlinear dynamical models for omics data: An application in cancer systems biology," *IEEE Transactions on Control Systems Technology*, 2018.
- [11] F. Bianconi, C. Antonini, L. Tomassoni, and P. Valigi, "Application of conditional robust calibration to ordinary differential equations models in computational systems biology: a comparison of two sampling strategies," *IET Systems Biology*, vol. 14, no. 3, pp. 107–119, 2020.
- [12] F. Bianconi, L. Tomassoni, C. Antonini, and P. Valigi, "A new bayesian methodology for nonlinear model calibration in computational systems biology," *Frontiers in Systems Biology*, to be published.
- [13] F. Bianconi, E. Baldelli, V. Luovini, E. F. Petricoin, L. Crino, and P. Valigi, "Conditional robustness analysis for fragility discovery and target identification in biochemical networks and in cancer systems biology," *BMC systems biology*, vol. 9, no. 1, p. 70, 2015.
- [14] F. Bianconi, C. Antonini, L. Tomassoni, and P. Valigi, "Cra toolbox: software package for conditional robustness analysis of cancer systems biology models in matlab," *BMC bioinformatics*, vol. 20, no. 1, p. 385, 2019.
- [15] (2020) Dipartimento Protezione Civile, GitHub repository. [Online]. Available: https://github.com/pcm-dpc/COVID-19
- [16] M. Vollmer, S. Mishra, H. Unwin, A. Gandy, T. Melan, V. Bradley, H. Zhu, H. Coupland, I. Hawryluk, M. Hutchinson *et al.*, "Report 20: A sub-national analysis of the rate of transmission of covid-19 in italy," 2020.
- [17] (2020) Istituto superiore di sanità, sistema di sorveglianza integrata covid-19. [Online]. Available: https://www.epicentro.iss.it/coronavirus/ sars-cov-2-sorveglianza
- [18] H. Nishiura, T. Kobayashi, T. Miyama, A. Suzuki, S.-m. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A. R. Akhmetzhanov *et al.*, "Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19)," *International journal of infectious diseases*, vol. 94, p. 154, 2020.
- [19] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020," *Eurosurveillance*, vol. 25, no. 10, p. 2000180, 2020.
- [20] A. Scala, A. Flori, A. Spelta, E. Brugnoli, M. Cinelli, W. Quattrociocchi, and F. Pammolli, "Time, space and social interactions: exit mechanisms for the covid-19 epidemics," *arXiv preprint arXiv:2004.04608*, 2020.
- [21] B. Armocida, B. Formenti, S. Ussai, F. Palestra, and E. Missoni, "The italian health system and the covid-19 challenge," *The Lancet Public Health*, vol. 5, no. 5, p. e253, 2020.