

Epidemic analysis of COVID-19 Outbreak and Counter-Measures in France

Dr. Eren Unlu*
Dr. Hippolyte Léger*
Dr. Oleksandr Motornyj*
Dr. Alia Rukubayihunga*
Thibaud Ishacian*
Dr. Mehdi Chouiten*

*Datategy SAS
Paris, France
first.last@datategy.fr

Abstract—The COVID-19 pandemic has triggered world-wide attention among data scientists and epidemiologists to analyze and predict the outcomes, by using previous statistical epidemic models. We propose to use a variant of the well known SEIR model to analyze the spread of COVID-19 in France, by taking in to account the national lockdown declared in March 11, 2020. Particle Swarm Optimisation (PSO) is used to find optimal parameters for the model in the case of France. We propose to fit the model based only on the number of daily fatalities, where an R^2 score based error metric is used. As the official number of confirmed cases is not reliable due to the lack of widespread testing, especially in the first phases of the outbreak, we show that basing the model optimisation on the number of fatalities can provide legitimate results.

I. INTRODUCTION

In December 2019, novel coronavirus-sourced atypical pneumonia cases were reported in Wuhan, China. Rapidly, it evolved into an epidemic in its city of origin [1]. Despite taken counter-measures, the outbreak has gradually spread out globally. The World Health Organisation (WHO) declared it a pandemic in March 11, 2020 and called for augmented enforcing policies to all governments [2]. Countries have responded to outbreak with varying degrees of containment and other preventive actions; also with varying latency [3] [4] [5] [6]. The pandemic is ongoing by the date April 22, 2020 when this paper is written, with total confirmed cases more than 2.5 million and 180,000 deaths.

This work focuses on France, where the first confirmed case was reported on January 24, 2020. The situation then rapidly deteriorated, leading authorities to execute more draconian policies [7] [8]. A nation-wide, strict lockdown has been initiated on March 17 and was announced to continue till May 11. By April 22, there are around 159,000 confirmed cases and 20,800 deaths caused by COVID-19. French government has announced that [9] [10] lockdown will be lifted starting on May 11, with strict measures. The exact details of the lockdown lifting policy are not publicized yet, however it is known that the allowed proportion of population and the commercial

and public establishments will be gradually increased week by week. The reopening date of high risk establishments such as restaurants or schools is still unknown. It would not be speculative to state that the reproduction number R_0 will also increase gradually over this period, however we expect that it will be much lower compared to pre-lockdown values due to increased awareness and public health measures.

Considering this setting, we aim to analyze the current situation and its future evolution using a variation of a widely employed mathematical epidemiology model: the SEIR (Susceptible, Exposed, Infected, Recovered) model [11] [12]. We have made the assumption to use two different reproduction numbers: R_0^b and R_0^q . R_0^b is used to describe reproduction rate between the first confirmed death until the start of the lockdown and R_0^q from the start to the end of the lockdown. The parameters of the model are fitted using Particle Swarm Optimization (PSO) [13] [14]. However, unlike other approaches we have decided to base our optimization on the number of fatalities only. In addition, rather than Mean Squared Error (MSE) or Mean Squared Logarithmic Error (MSLE); we minimize an R^2 score based metric [15] for daily deaths. The rationale behind this approach is that due to largely unknown dynamics of the novel coronavirus such as degree of infectiousness, length of incubation period and limited testing capability, using confirmed cases with such limited information may highly disrupt the validity of the model. However, the number of deaths caused by COVID-19 is much more definitive, especially in the first days of the outbreak and thus expected to increase the accuracy. Note that, with this approach the initial state values are also defined in terms of proportion of fatalities, hence they also are parameters of the PSO optimization.

According to our initial results, we estimate the reproduction number at around 3.56 before quarantine and 0.74 after total lockdown, in agreement with various recent studies around the globe for different scenarii [16] [17]. Using the developed model we predict that if lockdown continues with strict

measures, the total number of COVID-19 fatalities should topple below 50,000 (which is currently around 20,000) by late August, 2020; where the effects of the epidemic start to significantly diminish. As it is not possible to predict the reproduction number for the forthcoming lockdown lift, we propose two scenarii with a reproduction number increase of respectively 5% and 10% per week during 3 consecutive weeks. For these two scenarii, it is estimated that total number of fatalities may reach up to 70,000-80,000, and that an epidemic situation could continue till November, 2020.

II. RELATED WORK

The SEIR model has been one of the keystone components of statistical epidemiology for a long time, and has proven its validity also for relatively recent regional or global epidemics such as MERS, SARS and Ebola [18] [19] [20]. The SEIR model is classified as a *closed dynamical epidemiological model*, as it divides the total population into four distinct categories, where the proportion of individuals belonging to each category evolves over time, subjects passing from one state to other. The temporal transitions between states are defined by several differential equations [11]. Initially, the entire population is considered as susceptible except very few infected individuals. Gradually over time, according to base differential equations, susceptible individuals are exposed to disease, proportional to the reproduction number R_0 . Exposed subjects get infected, recover or die also based on the scalar parameters used in base differential equations. Generally, these parameters are calculated for the considered case using optimization algorithms [21]. In the past, researchers have also developed noteworthy number of extended SEIR/SIR models, where additional number of states are added, considering the special circumstances of the evaluated epidemic [18] [22]. We further explain the statistical formulation of an extended SEIR model in the next section.

COVID-19 pandemic has immediately gained attention among the research community and numerous different approaches using SEIR model have been proposed. First studies were published in January, 2020, for the initial epicenter Wuhan, less than a month after the novel coronavirus was identified [23] [24]. One particular work aims to extend the model according to characteristics of COVID-19, by a research group from the University of Basel [25] [26]. Authors identify 3 new states: H (Hospitalized), C (Critical) and D (Dead). We base our methodology by following this guideline, as it appears to fit the dynamics of the on going pandemic.

III. PROPOSED METHODOLOGY

As mentioned in the previous section, we developed a SEIR-HCD model as in [25]. The state transition diagram is shown in Fig. 1. Infected individuals may be hospitalized after a certain period. A proportion of the infected agents turn into critical cases, requiring intensive care; whilst the rest recovers. Among the critical cases, a certain proportion of individuals eventually dies. These proportion constants are among the parameters of the model to be optimized.

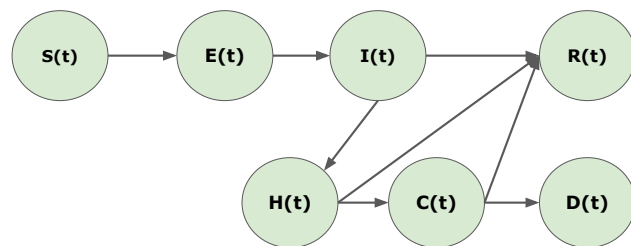


Fig. 1. Dynamical state transition diagram of the SEIR-HCD model. Infected individuals may be hospitalized after a certain period. A certain proportion of infected agents switches to the critical case state, requiring intensive care; whilst the rest recovers. Among the critical cases, a proportion of individuals eventually die.

Initially there are only a few infected agents, $I(t_0)$, while rest of the population is susceptible $S(t_0) = N - I(t_0)$; where N is the population of France (65 millions). Also note that, at any time the sum of all 7 states must be equal to N .

One of the innovative proposal we make is to base our model completely on the number of deaths. As mentioned previously, the most definitive data for COVID-19 case is the number of fatalities, due to current lack of pathological and epidemiological information about the disease and the low number of tests. Especially, in the first phase of dissemination of the virus in France, the number of tests was much lower, further decreasing the validity of using confirmed cases as a model initiator. Therefore, we propose to estimate the initial number of infected people, $I(t_0)$ from the initial number of fatalities $D(t_0)$, by simply reverse tracing the state transition diagram, using the proportion of hospitalized $H(t_0)$ and critical $C(t_0)$ cases. It is important to note that these proportion constants are parameters of the SEIR-HCD model we aim to optimize along a temporal axis. In other words, one novel outcome of our proposed algorithm is to be able to intrinsically calculate the initial number of infected citizens. It is highly reasonable to indicate that the number of infected people at the time where the first confirmed cases are announced shall be exponentially higher, due to very high proportion of asymptomatic cases [27].

The transitions between states are explained by this set of differential equations in terms of proportion of the total population [25] [26]:

$$\frac{\partial S(t)}{\partial t} = -\frac{R_0}{T_{inf}} I(t)S(t) \quad (1)$$

$$\frac{\partial E(t)}{\partial t} = \frac{R(t)}{T_{inf}} I(t)S(t) - T_{inc}^{-1} E(t) \quad (2)$$

$$\frac{\partial I(t)}{\partial t} = \frac{E(t)}{T_{inc}} - \frac{I(t)}{T_{inf}} \quad (3)$$

$$\frac{\partial H(t)}{\partial t} = (1 - p_a) \frac{I(t)}{T_{inf}} + (1 - p_f) \frac{C(t)}{T_{crt}} - \frac{H(t)}{T_{hsp}} \quad (4)$$

$$\frac{\partial C(t)}{\partial t} = \frac{p_c H(t)}{T_{hsp}} - \frac{C(t)}{T_{crt}} \quad (5)$$

$$\frac{\partial R(t)}{\partial t} = \frac{p_a I(t)}{T_{inf}} + (1 - p_c) \frac{H(t)}{T_{hsp}} \quad (6)$$

$$\frac{\partial D(t)}{\partial t} = \frac{p_f C(t)}{T_{crt}} \quad (7)$$

where T_{inc} is the incubation period of the coronavirus, T_{inf} is the infectiousness period of an infected agent, T_{hsp} is the duration it takes for an infected agent to check in to a health facility and finally T_{crt} is the duration it takes for an hospitalized person to turn into a critical case since the initial check-in. p_a, p_c, p_f respectively refer to the proportion of asymptomatic infected individuals, the hospitalized agents who switched to a critical case and the critical cases resulting in death. R_0 is the basic reproduction number for the coronavirus.

We have decided to use the reported data for France, starting from February 15, when the first fatality was confirmed. Without loss of generalization, we assume a binary reproduction number; R_0^b for the interval between February 15, 2020 and March 11, 2020 for the pre-quarantine period and R_0^q for lockdown period, which is still ongoing.

Finally, we need to optimize the model to find these 9 parameters : $R_0^q, R_0^b, p_a, p_c, p_f, T_{inc}, T_{inf}, T_{hsp}, T_{crt}$. We employ a Particle Swarm Optimization (PSO) for this task, which is a powerful evolutionary algorithm, well suited for this setting [13].

As we propose to set our model solely on the initially reported COVID-19 related fatalities, the the initial states of each dynamic component can be denoted as :

$$I(t_0) = D(t_0)(1 - p_a)(p_c)(p_f) \quad (8)$$

$$E(t_0) = R_0^b I(t_0) \quad (9)$$

$$H(t_0) = I(t_0)(1 - p_a) \quad (10)$$

$$C(t_0) = H(t_0)p_c \quad (11)$$

$$R(t_0) = H(t_0)p_c \quad (12)$$

$$E(t_0) = N - (H(t_0) + C(t_0) + R(t_0) + I(t_0) + D(t_0)) \quad (13)$$

Other than suggesting a model optimization based on fatalities, we have also observed that the most adapted error metric is the R^2 score of differentials of number of deaths (i. e. series of daily fatalities). This avoids overfitting, while preserving the parameter optimization in plausible ranges. The R^2 score

is a metric ranging from $[-\infty, 1]$, where 1 denotes the full accuracy, so we aimed to minimize the following value:

$$\epsilon = 1 - R^2\left(\frac{\partial D(t)}{\partial t}, \frac{\partial \hat{D}(t)}{\partial t}\right) \quad (14)$$

The gradual lockdown lift, starting on May 11, is modeled as a %5 to %10 increase of quarantine time reproduction number for each week for next 3 weeks (assuming reproduction number does not grow after 3 weeks), presenting two different scenarii for France.

IV. RESULTS

With PSO optimization based on the defined error metric, the optimal parameters for France are found in Table 1. We have estimated the reproduction number at 3.56 prior to lockdown and 0.74 after lockdown. In an independent research for France, these numbers are reported as 3.3 and 0.5, where their method is not exposed; confirming integrity of our approach from a parallel perspective [28]. We also report an approximate average reproduction number before lockdown that is consistent with the initial various studies on the issue [25] [26], where the average of the studies suggests a value around 3.6.

As shown in Table 1, all other parameters of the suggested model are within the range of other published research on COVID-19 [25]. For instance, asymptomatic ratio of 0.79, incubation period of 5.10 days, infectiousness period of 2.79 days, hospitalisation period of 5.14 days, transition to critical state period of 14.06, ratio of 0.12 for infected becoming critical and fatality ratio of 0.33 for critical patients are all close to the medians of the other reported work [25]. These results are particularly interesting, since our SEIR model optimization approach with suggested error metric for number of fatalities is able to converge to the optimal point, and directly coronavirus related metrics such as incubation period, hospitalisation period etc. (assuming demographics etc. of a country does not influence considerably) are close to the average of the previous works. This asserts the validity of binary reproduction number approach for lockdown and its calculated values. Note that, in complex dynamics of a SEIR model, slight variance of parameters may have a drastic impact on the outputs, such as the estimated number of deaths.

We have calculated the mortality ratio of infected people $(1 - p_a)p_c p_f$ as %0.83, where [28] proposes %0.5 for France. By the date April 22, our model estimates the proportion of already infected population as %6.6 (Fig. 2-3), while [28] also reports as %6. It is quite surprising to observe independent works show similar results, while our SEIR model optimization only takes fatalities into account for optimization.

Based on our model, we estimate that if lockdown is maintained, the number of fatalities for France might never pass the limit of 50,000. If reproduction number grows by 5% each week during 3 weeks following the lockdown lift; this upper limit might reach 70,000 (Fig. 4). We observe that, if strict measures are respected till September, 2020 the epidemic almost terminates around this date. In case this growth rate

TABLE I

OPTIMAL VALUES OF 9 PARAMETERS OF THE SEIR MODEL FOUND BY A PSO ALGORITHM BASED ON THE R^2 SCORE OF DAILY FATALITIES IN FRANCE

Reproduction number before lockdown R_0^b	Reproduction number after lockdown R_0^a	Incubation Period T_{inc}	Infectiousness Period T_{inf}	Hospitalisation Period $T_{h.sp}$	Transition to Critical State T_{crt}	Ratio of infected without symptoms p_a	Ratio of hospitalised turning into critical state p_c	Ratio of critical patients resulting in death p_f
3.56	0.74	5.10 (days)	2.79 (days)	5.14 (days)	14.06 (days)	0.79	0.12	0.33

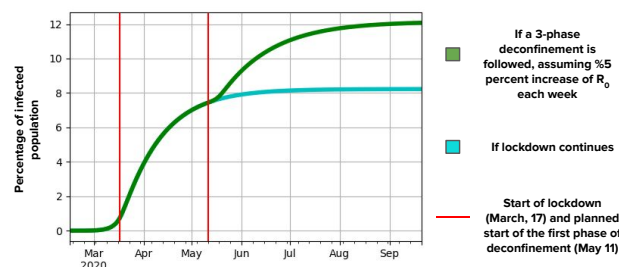


Fig. 2. Predicted number of infected proportion of France from February 15, 2020 to September 21, 2020; considering a scenario, where quarantine time reproduction number grows by 5% each week, during 3 weeks following lockdown lift.

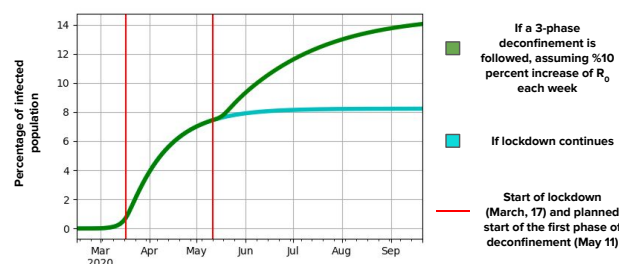


Fig. 3. Predicted number of infected proportion of France from February 15, 2020 to September 21, 2020; considering a scenario, where quarantine time reproduction number grows by 10% each week, during 3 weeks following lockdown lift.

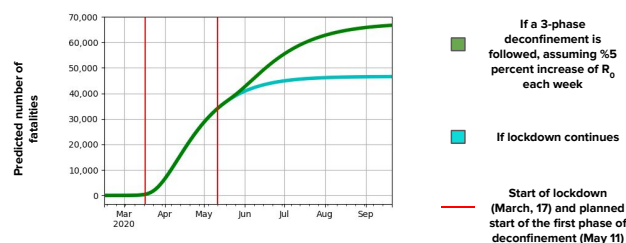


Fig. 4. Predicted number of fatalities for France from February 15, 2020 to September 21, 2020; considering a scenario, where quarantine time reproduction number grows by 5% each week, during 3 weeks following lockdown lift.

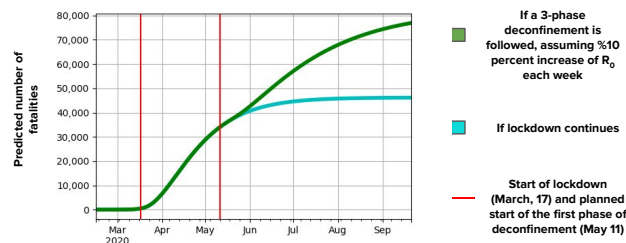


Fig. 5. Predicted number of fatalities for France from February 15, 2020 to September 21, 2020; considering a scenario, where quarantine time reproduction number grows by 10% each week, during 3 weeks following lockdown lift.

V. CONCLUSION

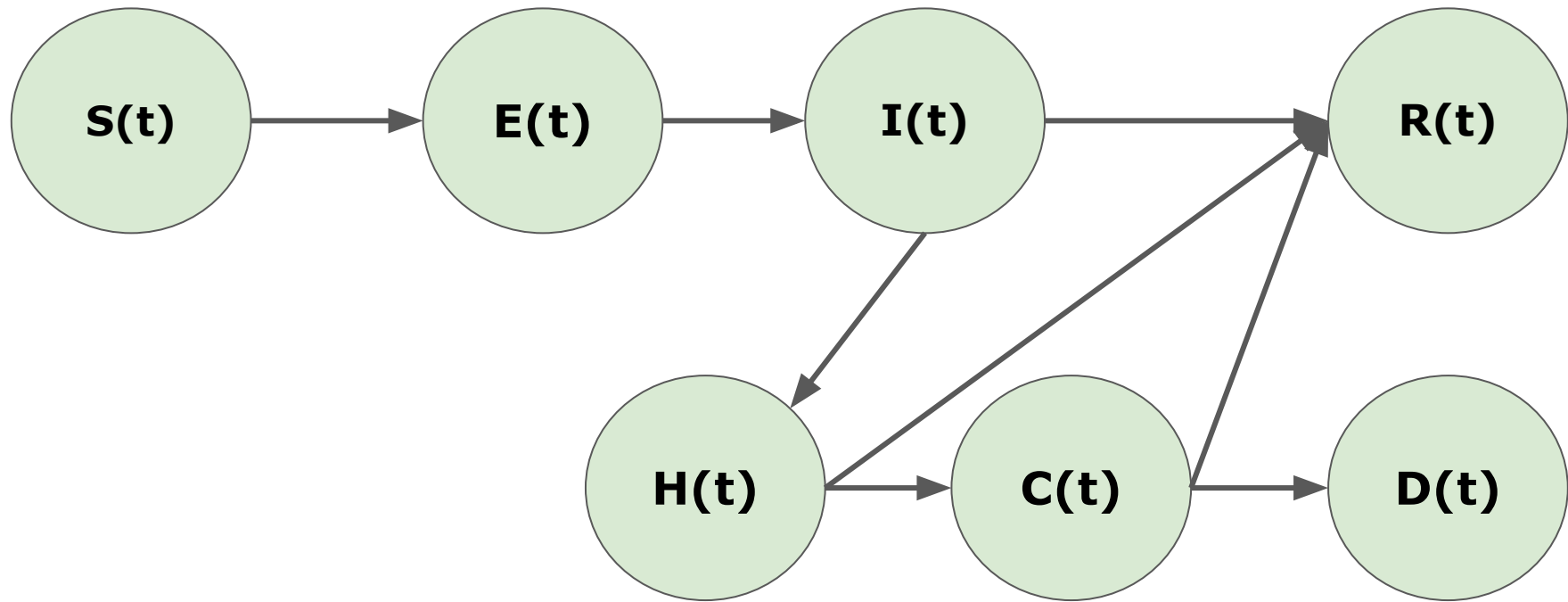
becomes 10% per week, the upper death toll limit might reach 80,000 (Fig. 5); where epidemic situation lasts till November, 2020. Considering the fact that, French government has already prepared a versatile and extensive plan for lifting, by limiting social gatherings, augmented surveillance and nationwide distribution face masks, these scenarii with %5-%10 per week increase of quarantine time reproduction number seem legitimate.

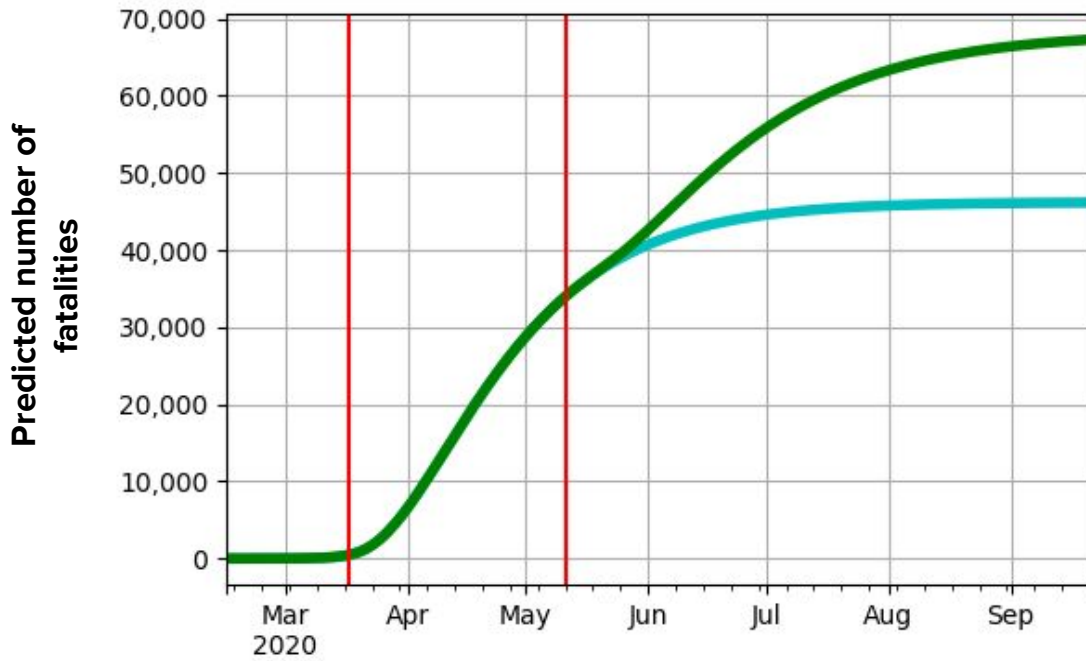
In this paper, we have proposed a SEIR model for COVID-19 epidemic in France, similar to [25]. Unlike other similar attempts, we have used only confirmed number of deaths as an optimization metric. Rationale behind this proposal is that deficiency in testing coverage, especially in the early phases of the outbreak, greatly underestimates the number of confirmed cases. However, the number of confirmed fatalities is a much more solid evidence in this setting. An error metric based on the daily death toll is presented and model parameters are


optimized using PSO algorithm. Also note that, the initial state of the model (initial values of each state) is also defined based on the proportion of fatalities, hence all initial state values are also parameters for the PSO optimization. We believe that, for epidemics at this scale starting a SEIR models with a hypothetical single infected individual may greatly lower the accuracy; especially considering their potential drastic impact.


REFERENCES


- [1] W. H. Organization *et al.*, “Coronavirus disease 2019 (covid-19): situation report, 74,” 2020.
- [2] C. Sohrabi, Z. Alsafi, N. O’Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, “World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19),” *International Journal of Surgery*, 2020.
- [3] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, “How will country-based mitigation measures influence the course of the covid-19 epidemic?” *The Lancet*, vol. 395, no. 10228, pp. 931–934, 2020.
- [4] J. Jia, J. Ding, S. Liu, G. Liao, J. Li, B. Duan, G. Wang, and R. Zhang, “Modeling the control of covid-19: Impact of policy interventions and meteorological factors,” *arXiv preprint arXiv:2003.02985*, 2020.
- [5] T. Colbourn, “Covid-19: extending or relaxing distancing control measures,” *The Lancet Public Health*, 2020.
- [6] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun *et al.*, “The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak,” *Science*, 2020.
- [7] D. Fanelli and F. Piazza, “Analysis and forecast of covid-19 spreading in china, italy and france,” *Chaos, Solitons & Fractals*, vol. 134, p. 109761, 2020.
- [8] S. B. Stoecklin, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon, M. Mechain, L. Meurice, M. Nguyen, C. Bassi *et al.*, “First cases of coronavirus disease 2019 (covid-19) in france: surveillance, investigations and control measures, january 2020,” *Eurosurveillance*, vol. 25, no. 6, p. 2000094, 2020.
- [9] “Lifting lockdown: France looks ahead to options for easing coronavirus restrictions,” *France24*, Apr 2020. [Online]. Available: <https://www.france24.com/en/20200404-lifting-lockdown-france-looks-ahead-to-options-for-easing-coronavirus-restrictions>
- [10] R. R. Ania Nussbaum, “France to unveil end-of-lockdown plan within 2 weeks,” *Bloomberg*, Apr 2020. [Online]. Available: <https://www.bloomberg.com/news/articles/2020-04-19/france-to-unveil-end-of-lockdown-plan-within-2-weeks>
- [11] M. Y. Li, J. R. Graef, L. Wang, and J. Karsai, “Global dynamics of a seir model with varying total population size,” *Mathematical biosciences*, vol. 160, no. 2, pp. 191–213, 1999.
- [12] N. ÖZalp and E. Demirci, “A fractional order seir model with vertical transmission,” *Mathematical and Computer Modelling*, vol. 54, no. 1-2, pp. 1–6, 2011.
- [13] A. Lee, “pyswarm: Particle swarm optimization (psa) with constraint support,” 2014.
- [14] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95-International Conference on Neural Networks*, vol. 4. IEEE, 1995, pp. 1942–1948.
- [15] M. R. Veall and K. F. Zimmermann, “Pseudo-r2 measures for some common limited dependent variable models,” *Journal of Economic surveys*, vol. 10, no. 3, pp. 241–259, 1996.
- [16] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, “The reproductive number of covid-19 is higher compared to sars coronavirus,” *Journal of travel medicine*, 2020.
- [17] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell, “Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions,” *MedRxiv*, 2020.
- [18] P. Diaz, P. Constantine, K. Kalmbach, E. Jones, and S. Pankavich, “A modified seir model for the spread of ebola in western africa and metrics for resource allocation,” *Applied Mathematics and Computation*, vol. 324, pp. 141–155, 2018.
- [19] N. Jia and L. Tsui, “Epidemic modelling using sars as a case study,” *North American Actuarial Journal*, vol. 9, no. 4, pp. 28–42, 2005.
- [20] B. Yong and L. Owen, “Dynamical transmission model of mers-cov in two areas,” in *AIP Conference Proceedings*, vol. 1716, no. 1. AIP Publishing LLC, 2016, p. 020010.
- [21] A. Kaddar, A. Abta, and H. T. Alaoui, “A comparison of delayed sir and seir epidemic models,” *Nonlinear Analysis: Modelling and Control*, vol. 16, no. 2, pp. 181–190, 2011.
- [22] J. A. Weller and G. Dimitoglou, “Survey of models, methods and techniques for computational epidemiology,” in *MSV*, 2009, pp. 99–105.
- [23] J. T. Wu, K. Leung, and G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study,” *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.
- [24] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, “Epidemic analysis of covid-19 in china by dynamical modeling,” *arXiv preprint arXiv:2002.06563*, 2020.
- [25] R. Neher, I. Aksamentov, and N. Noll, “About covid-19 scenarios,” <https://covid19-scenarios.org/>, 2020.
- [26] A. Sayed, “Seir-hcd model,” 2020. [Online]. Available: <https://www.kaggle.com/anjum48/seir-hcd-model/>
- [27] M. G. Pedersen and M. Meneghini, “Quantifying undetected covid-19 cases and effects of containment measures in italy,” *ResearchGate Preprint (online 21 March 2020) DOI*, vol. 10, 2020.
- [28] S. et Avenir avec AFP, “Covid-19 : seuls 6% des français infectés par le coronavirus le 11 mai, un niveau très insuffisant,” 2020. [Online]. Available: <https://www.sciencesetavenir.fr/sante/covid-19-environ-6-des-francais-infectes-le-11-mai-143656>

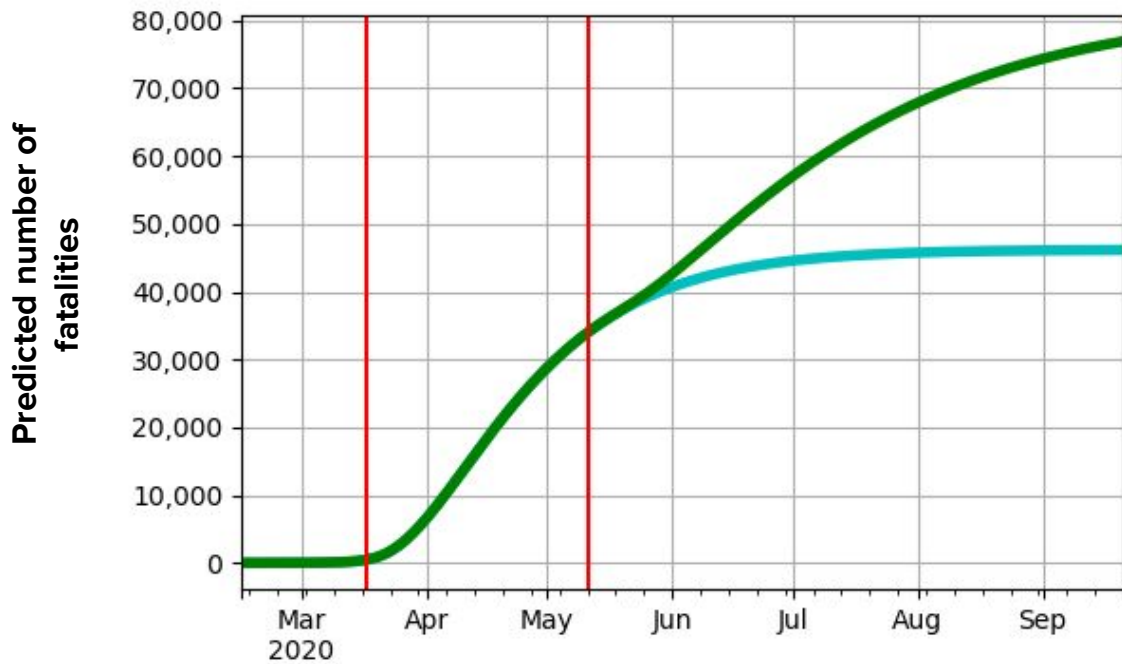





 If a 3-phase deconfinement is followed, assuming %10 percent increase of R_0 each week


 If lockdown continues

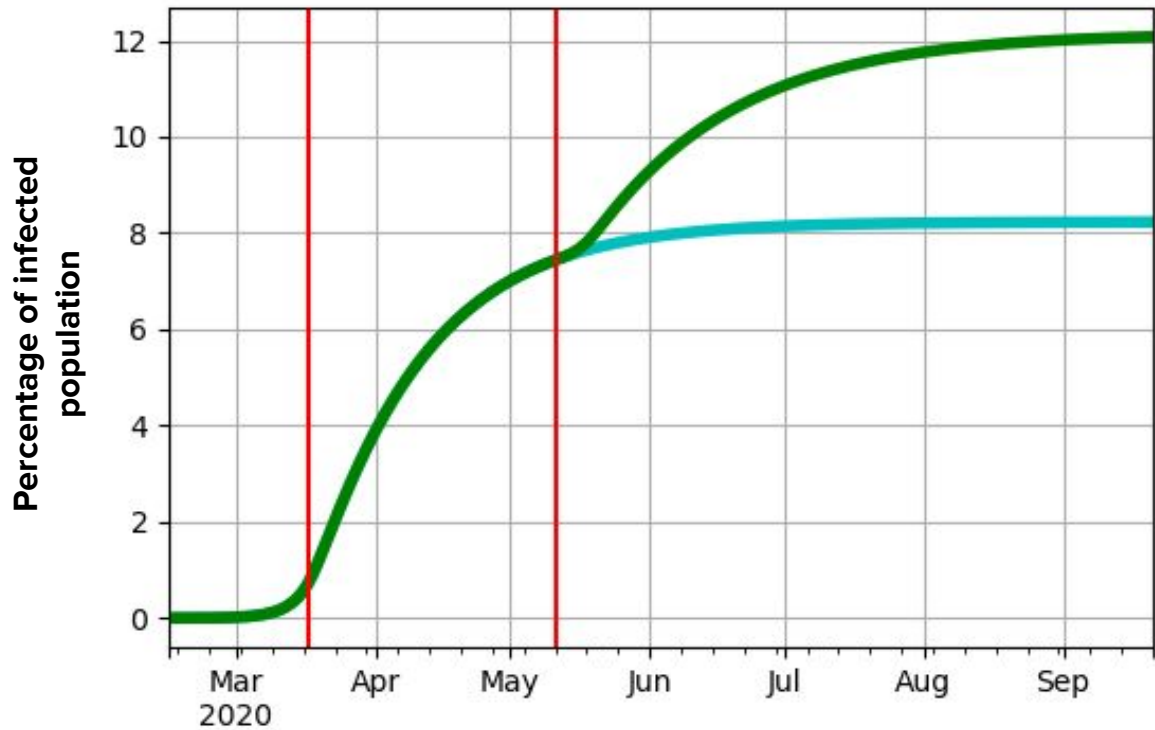
 Start of lockdown (March, 17) and planned start of the first phase of deconfinement (May 11)





 If a 3-phase deconfinement is followed, assuming %10 percent increase of R_0 each week


 If lockdown continues

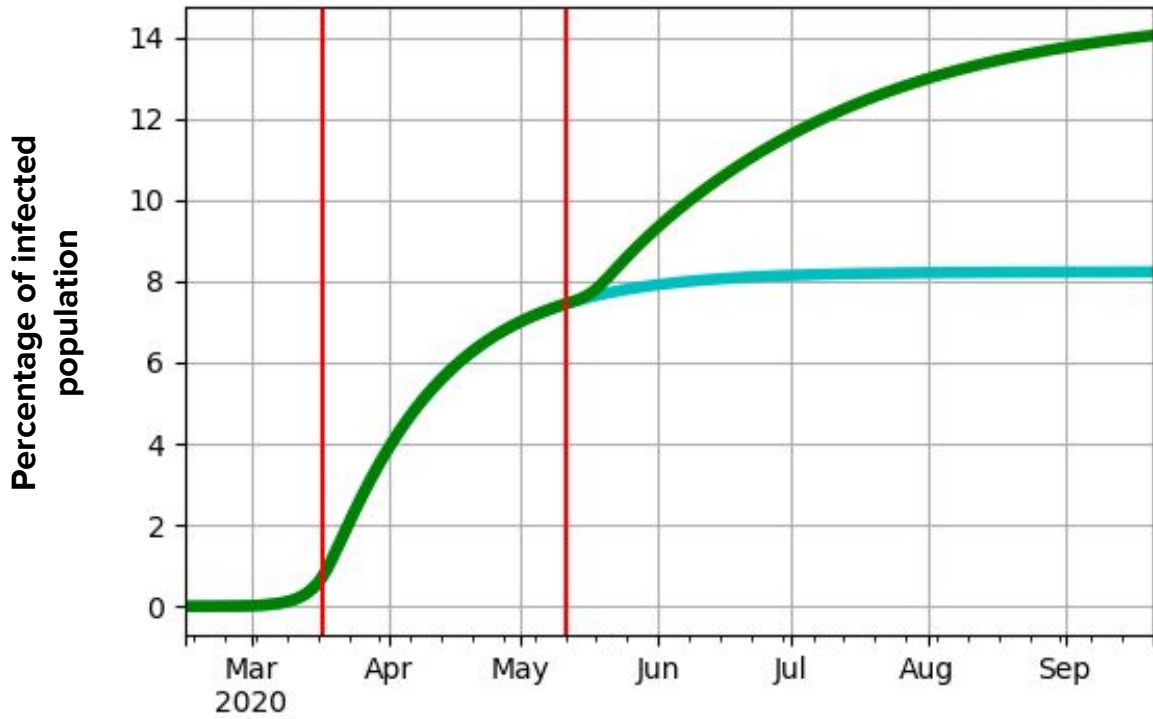
 Start of lockdown (March, 17) and planned start of the first phase of deconfinement (May 11)





 If a 3-phase deconfinement is followed, assuming %5 percent increase of R_0 each week


 If lockdown continues

 Start of lockdown (March, 17) and planned start of the first phase of deconfinement (May 11)



 If a 3-phase deconfinement is followed, assuming %10 percent increase of R_0 each week

 If lockdown continues

 Start of lockdown (March, 17) and planned start of the first phase of deconfinement (May 11)