Development and Testing of a New Epidemiological Model Applied to COVID-19 Pandemic

Project Team T88: Stefano Veroni and José Antonio Guzmàn Funck

# Supervisor: Kilian D Stenning

22 June 2020

**Abstract**

In 2020, COVID-19 pandemic has bent many countries all over the world. We developed an epidemiological model to study the spread and characteristics of COVID-19. First, information was gathered by consulting many specialised virological and epidemiological papers. Then, based on theoretical considerations, the mathematical model was developed. A code was written in order to fit our model to the current available data specifically for considered country, i.e. Italy, Spain, UK, Brazil, Sweden and India, and run simulations. Moreover, we have been able to make some short-term predictions about the number of infected and deceased people.

1. BACKGROUND

The main concept for the project was to create a model which could describe the evolution of COVID-19 pandemic. The idea was first inspired by the current situation we were and we are living in and the first project in the list we were originally given for our summer projects 2020, i.e. “Markovs for life, trading and diseases”. The work was conducted by first gathering information by consulting specialized papers, then by modelling the disease using computer simulations.

The initial goals were:

* to create a model of differential equations which, based on theoretical bases, could describe the evolution of COVID-19 pandemic (or ideally any pandemic);
* to gather information about the parameters related to the disease by consulting specialized virological and epidemiological papers and articles;
* to create a code which could be used to develop simulations based on our model;
* to compare simulations with the official data made available, in order to fit the parameters and give an indication of how well the model performs.

Extended goals were:

* to use the model to make short term, i.e. 1 or 2 months, predictions about the pandemic.

1. DESCRIPTION OF PROJECT WORK
2. *Project start-up*

Since the 8th of May we have kept an intense online WhatsApp communication reinforced by weekly meetings between us two on Microsoft Teams. The first meeting with the supervisor took place on the 18th of May. Until that day different options were considered for the project. In particular three should be mentioned: the physics of some sports (football, basketball and Formula 1 especially), the creation of an algorithm which could create a map of the visible sky and constellations based on the position, the period of the year and the time which was inputted, and the creation of an epidemiological model which could describe the past, present and possibly future of COVID-19 pandemic. After the meeting it was decided to proceed with the last option, i.e. the modelling of COVID-19.

After that day, the communication between the two team members proceeded constantly. Instead we only had one other meeting with the supervisor on the 16th of June. However, since the 27th of May he has always had the possibility of directly accessing our work, which was regularly uploaded onto Microsoft Teams files exchange.

1. *Initial information-taking and model development*

The two activities of data-taking and model development proceeded simultaneously, as, more than once, the new information found led to a modification of the model itself. Both the team members worked simultaneously on both the tasks. The information was taken from specialized papers in the fields of virology and epidemiology.

Through the creation of a shared word document, which served as a lab book, we could keep track of each other’s progress: it was fundamental to always read what the other team member had previously found out before beginning the research. Moreover, nothing of what was written down would be deleted: even if we both would had considered something to be mistaken, it was left in the document for future reference. We also kept track of all the sources which were consulted. Although the data-taking and model development activities never ended, as new useful information could have always been found and improvements to the model could have always been made, they have become much less intense after the 27th of May, when the work on the code began.

1. *Development of the code and data analysis*

The coding strategy we followed was simple: we both developed a code independently, although supporting each other, then we compared performances and chose Stefano’s code over José’s to perform the vast majority of the fits.

The development of the code was preceded by some online research, as the task we were facing was completely new to both of us. The language used for all the coding was python, the environment Spyder IDE. The fitting and simulation running tasks required the packages scipy, matplotlib, lmfit and emcee. The last two packages were used to fit the model and perform simulations implementing Markov Chain Monte Carlo methods.

Each analysed country demanded a slightly different code, because of the different ways the COVID-19 pandemic was faced in each of them, even if the bases and the fundamental working of all the codes was the same. On average, the code used to fit the parameters of the model with respect to the official data was 350 lines long for each country. The demanded simulations were run on our personal laptops. The fitting process was much less automated than anticipated: there was in fact a higher dependence than expected on the initial guesses for the parameters’ values, which resulted in the necessity for a frequent human intervention. Another crucial factor that affected the development of the project is the time taking by our computers to minimise the error function to optimise the parameters. A minimum of 10 minutes was taken by each run of the code. This was due to the complexity of the model and the number of parameters to fit.

Two issues arose when collecting the data for fitting. First, in some countries the intensity at which the identification processes were conducted was inconstant, which resulted in worse fitting, as the model assumes constancy in the procedures. Secondly, it was sometimes the case that local, national and international sources disagreed: among the countries we studied, it was Spain the one affected the most by this issue.

1. *Pursuing extended goal: short-term predictions*

After the fitting of the parameters was completed, it was easy to apply the model to make predictions for the future. However, due to the high level of unpredictability of reality, it was decided to only make predictions about the number of infected and deceased cases in the short-term.

1. *Equipment used and purchased*

All the activities were performed online using our laptops. The communication between team members required an intense use of WhatsApp on our smartphones. For file exchange and videoconference purposes we used Microsoft Teams.

1. SUMMARY OF RESULTS

We have been able to develop an epidemiological model, we called , to study the characteristics of the outbreak of COVID-19 in our societies. It can potentially be used to describe the evolution of any pandemic.

We have also been able to write a functioning code to solve and plot the system of differential equations that define our epidemiological model. Once the data was properly treated, we could fit the model’s parameters specifically for each considered country, i.e. Italy, Spain, UK, Brazil, Sweden and India, using the Markov Chain Monte Carlo method.

Because of the differences which exist between the imperfections and inconstancy of the real world and the substantial perfection the theoretical model assumes, standard goodness of fit tests, as the Chi Squared Test, were considered inappropriate to describe the accuracy and precision of our fits.

As a final achievement, the fitted model was used to make short term predictions.

1. CONCLUSION

The initial goals of the project were successfully met, as we were able to develop a model that, after some tuning, was successfully fitted specifically for each considered country, i.e. Italy, Spain, UK, Brazil, Sweden and India. Moreover, the model was then implemented to provide reasonable short-term predictions.

We believe in the importance of creating models and applying them to the real world, in order to better understand what has happened, what’s happening now and what will happen in the future. However, we also recognize that the study and analysis of a pandemic during the pandemic itself is necessarily limited by the relatively scarceness of certainties in the available information. Therefore, we believe that the model development and testing are tasks which should continuously carried out, in order to progressively improve their accuracy and precision.

We also reckon that improving the technical equipment, in particular the power of the computational devices, would lead to significant improvements regarding time-management and accuracy related to the fitting activity.

1. BIBLIOGRAPHY

The literature that we consulted is listed below, with a brief comment on each item’s role in the project.

REFERENCES

1. Uchida Y, Y1 Project Proposals (original). Available at <https://www.hep.ph.ic.ac.uk/~yoshiu/Y1Project2020/>, as of 20th June.

It represents the list of projects we were initially given.

1. Nabi KN, Forecasting Covid-19 pandemic: a data driven analysis. MedRxiv. 2020. Available from <https://doi.org/10.1101/2020.05.12.20099192>, as of 17th June 2020.

The model exposed in the paper represented the first inspiration for the project, although it wasn’t directly used in our work. It also provided likely ranges for the value of many parameters such as infectious contact rate, infectiousness factor for asymptomatic, recovery rates for different categories of infected individuals.

1. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. Dec. 2000; Vol. 42 (4): 599-653.

Contained precious information on the epidemiological MSEIR model, the basis for any more advanced attempt to model a pandemic.

1. Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Science Direct, Infectious Disease Modelling. 2020; Vol. 5: 248-255.

It represented further inspiration for our model. Moreover, it pvided a value for the probability of displaying symptoms.

1. Nishiura H, Kobayashi T, Suzuki A, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis. 2020; Vol. 94: 154-155.

It provided a new estimation for the probability of displaying symptoms.

1. Mizumoto K, Kagaya K, Zarebski A, Chowell G.Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 2020; Vol.25 (10).

It provided a new estimation for the probability of displaying symptoms.

1. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Int. Med. 2020; Vol. 172, pp. 577–582. doi: 10.7326/M20-0504.

It provided precious information about the average incubation period of the virus.

1. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet infectious diseases. 2020; 20 (6): 669-677. Available from: <http://dx.doi.org/10.1016/S1473-3099(20)30243-7>, as of 17th June 2020.

Provided estimates of the severity and fatality of the coronavirus

1. Basu A. Estimating the infection fatality rate among symptomatic COVID-19 cases in the United States. Health Aff (Millwood). published online ahead of print May 2020; doi:10.1377/hlthaff.2020.00455.
2. Worldmeters. Available at <https://www.worldometers.info>, as of 17th June 2020.
3. Paradisi M, Rinaldi G. An Empirical Estimate of the Infection Fatality Rate of COVID-19 from the First Italian Outbreak. SSRN Electronic Journal. 2020. Available from: doi: 10.2139/ssrn.3582811.
4. Villa M, Myers JF, Turkheimer F. COVID-19: Recovering estimates of the infected fatality rate during an ongoing pandemic through partial data. Medical Letter on the CDC & FDA. May 3, 2020: pp. 208.
5. Grewelle R, De Leo G. Estimating the Global Infection Fatality Rate of COVID-19. Medical Letter on the CDC & FDA. Jun 7, 2020: pp. 286
6. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates. MedRxiv. May 2020. Available from <https://doi.org/10.1101/2020.05.03.20089854>, as of 17th June 2020.
7. Oke J, Heneghan C. Global Covid-19 Case Fatality Rates. Mar 2020. CEBM.net. Available from <https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/>, as of 17th June 2020.

Wilson L. SARS-CoV-2, COVID-19, Infection Fatality Rate (IFR) Implied by the Serology, Antibody, Testing in New York City. SSRN Electronic Journal. May 2020. Available from: doi: 10.2139/ssrn.3590771, as of 17th June 2020.

The references [7-15] were useful in obtaining a reasonable range of value for COVID-19 infection fatality rate, a controversial parameter whose estimate highly varies between different papers.

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun. May 2020; Vol.109: 102433

Provided a good estimate for the average time in between the onset of symptoms and death.

1. He D, Zhao S, Lin Q, et al. The relative transmissibility of asymptomatic COVID-19 infections among close contacts. Int J Infect Dis. 2020; Vol. 94:145-147. doi:10.1016/j.ijid.2020.04.034

Provided a new estimate for the infectiousness factor for asymptomatic individuals.

1. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet. 2020; Vol. 395 (10229): 1054-1062. Available from: <http://dx.doi.org/10.1016/S0140-6736(20)30566-3>, as of 17th June 2020.
2. He X, Eric L, Peng W, Xilong D, Jian W, Xinxin H, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Medicine. 2020; Vol. 26. 10.1038/s41591-020-0869-5.
3. McIntosh K. Coronavirus disease 2019 (COVID-19): Epidemiology, virology, clinical features, diagnosis, and prevention. 2020. Available at <https://www.uptodate.com/contents/coronavirus-disease-2019-covid-19-epidemiology-virology-and-prevention>, as of 17th June 2020.

Articles [17-19] provided useful information about the evolution in time of viral shedding in an infected patient.

1. Pedersen M, Meneghini, M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy: Predicting phase 2 dynamics. 2020. 10.13140/RG.2.2.11753.85600.

Provided a useful quantitative insight regarding the relationship between the numbers of the total and the detected infected individuals.

1. ISCIII (*Carlos III’s Health Institute from Spain*). Data for new cases of COVID –19 in Spain. The data is separated by regions. Data available at <https://cnecovid.isciii.es/covid19/resources/datos_ccaas.csv> as of 18th June 2020.

It provided information about the distribution of new daily cases. This distribution was then scaled up to match the total number of cases provided by reference [9].

1. EpData. Information gathered from the Health Ministry of Spain about recovered cases of COVID-19 until the 17th of May. After that, no more data on this was published. Data available at <https://www.epdata.es/casos-curados-coronavirus-espana/58d0919c-8ad1-4a3f-9255-55f5b116da23> as of 19th June 2020.

This source provided data for the recovered cases in Spain.

1. ISCIII’s MoMo (*System for the Daily Surveillance of Mortality for Spain of the ISCII*). Information about the excess of mortality in Spain during the pandemic. Data available at <https://momo.isciii.es/public/momo/data> as of 18th of June 2020

This source provided data of the deceased cases in Spain until the 5th of May, when the death excess is not significant anymore.

1. Matthew Bennett. Independent journalist at <https://twitter.com/matthewbennett> . Gathering of deceased reported by regions in Spain from 5th of May to 14th of June. Data available at <https://bit.ly/2UTDlJ4> as of 18th June 2020.

This source was used to gather information about the deaths in Spain in between the 5th of May and 14th of June.

1. Rapporto ISS-ISTAT. Caratteristiche dei pazienti deceduti positivi all'infezione da SARS-CoV-2 in Italia. Dati al 11 giugno 2020. 2020. Available at <https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia>, as of 17th June 2020.

Such report was useful to gather important information about the average amount of time which characterized the detection or recovery of death of infected individuals in Italy.

1. Rigby S. Coronavirus immunity appears to last for ‘at least two months’ after diagnosis. Jun 2020. Available at <https://www.sciencefocus.com/news/coronavirus-immunity-appears-to-last-for-at-least-two-months-after-diagnosis>, as of 17th June 2020.
2. Simon M. Here’s How Covid-19 Immunity Compares to Other Diseases. Apr. 2020. Available at <https://www.wired.com/story/covid-19-immunity/>, as of 17th June 2020.

The articles [23-24] were consulted to obtain information about the possibility of obtaining or not immunity after recovery from the disease.

1. Cleb, <https://stackoverflow.com/questions/11278836/fitting-data-to-system-of-odes-using-python-via-scipy-numpy>, last visited on the 17th of June.

The answer of the user Cleb represented the starting point for the creation of the code which has been used to fit the model and run simulations.

1. WHO Coronavirus Disease (COVID-19) Dashboard. Available at <https://covid19.who.int/?gclid=Cj0KCQjwirz3BRD_ARIsAImf7LOw4SLL2aRhRERNNHOrDFVyQ6NSDU4JLQI56Hi0yPgAQvVc_YuyWUgaArtcEALw_wcB>, as of 20th June 2020.

Consulted in the dates between 27th May and 20th June, together with reference [9], provided official number about the pandemic in the world.