Forecasting COVID-19

Matjaž Perc1,2,3*, Nina Gorišek Miksić4,5, Mitja Slavinec1 and Andraž Stožer5

1 Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia, 2 China Medical University Hospital, China Medical University, Taichung, Taiwan, 3 Complexity Science Hub Vienna, Vienna, Austria, 4 Maribor University Medical Center, Maribor, Slovenia, 5 Faculty of Medicine, University of Maribor, Maribor, Slovenia

The World Health Organization declared the coronavirus disease 2019 a pandemic on March 11th, pointing to the over 118,000 cases in over 110 countries and territories around the world at that time. At the time of writing this manuscript, the number of confirmed cases has been surging rapidly past the half-million mark, emphasizing the sustained risk of further global spread. Governments around the world are imposing various containment measures while the healthcare system is bracing itself for tsunamis of infected individuals that will seek treatment. It is therefore important to know what to expect in terms of the growth of the number of cases, and to understand what is needed to arrest the very worrying trends. To that effect, we here show forecasts obtained with a simple iteration method that needs only the daily values of confirmed cases as input. The method takes into account expected recoveries and deaths, and it determines maximally allowed daily growth rates that lead away from exponential increase toward stable and declining numbers. Forecasts show that daily growth rates should be kept at least below 5% if we wish to see plateaus any time soon—unfortunately far from reality in most countries to date. We provide an executable as well as the source code for a straightforward application of the method on data from other countries.

Keywords: COVID-19, pandemic, disease dynamics, exponential growth, virality

1. INTRODUCTION

According to data in real time [1], confirmed coronavirus disease 2019 (COVID-19) cases are growing exponentially in most countries around the world. In Italy and Spain the pandemic is already overburdening the healthcare system [2], and shall the current trends persist, it will not take long before this becomes the grim reality also in many other European countries and the United States. Forecasting COVID-19 dissemination thus plays a key role [3–7]. In the first place, to inform governments and healthcare professional what to expect and which measures to impose, and secondly, to motivate the wider public to adhere to the measures that were imposed to decelerate the spreading lest a regrettable scenario will unfold [8, 9].

Research on epidemic processes has a long and fruitful history in statistical physics [10, 11]. Simple mathematical models that describe the essence of epidemic spreading can be used to fit the data with an overseeable number of parameters, and the obtained values can then be used to make informed predictions. In recent years, the research community has also accumulated overwhelming evidence in favor of complex and heterogeneous connectivity patterns in social networks [12–16]. These play a key role in determining the behavior of equilibrium and non-equilibrium systems in general, and the spreading of epidemics and finding optimal containment strategies in particular.

Interdisciplinary explorations at the interface of statistical physics, network science, and epidemiology, driven by massive amounts of data recording our health and way of life, have given
rise to digital epidemiology [17] and to the theory of epidemic
processes on complex networks [10]. From classical models that
assume well-mixed populations, to the more recent models that
account for behavioral feedback and the structure of our social
networks, we have come a long way in better understanding
disease transmission and disease dynamics. We are now able
to use this knowledge to develop effective prevention strategies
[11], and more broadly, we can use the synergies between these
different fields of research to improve our lives and societies
[18, 19].

Nonetheless, in times of urgency even the simplest model
can be too complicated, and the small gaps between different
fields of research can seem like gapping holes. In this paper,
we therefore present a simple iterative method to forecast
the number of COVID-19 cases, under the assumption that
governmental data is legitimate and truthful. The goal is not
to strive for meticulous accuracy nor to present our method
as the state of the art, but simply to provide first insights and
guidelines on elementary principles. We will be happy if our work
motivates further research to yield more elaborate and accurate
prediction methods.

2. METHOD

As input, our method requires only the readily available daily
values of confirmed cases. We denote these values as \(x_i\), where
\(i \in [0, n]\) is the index of days. Assuming we have \(n\) values available
in total, we take the last \(m\) values of the \(x_i\) series and determine
the average growth rate during this time according to

\[
G_\Delta = \frac{1}{m} \sum_{i=n-m}^{n-1} \left( \frac{x_i}{x_{i-1}} - 1 \right).
\]

(1)

We also record the minimal and the maximal growth rate during
the last \(m\) days as \(G_i\) and \(G_f\), respectively. The simple iteration

\[
x_{i+1} = x_i + G_\Delta \quad \text{(2)}
\]

already provides a decent forecast beyond \(i = n - 1\), assuming
the original \(m\) values are described well by exponential growth.

This, however, does not take into account that after \(h \approx 14\)
days the majority of infected will recover, and that after \(d \approx 21\) days a fraction \(p \approx 0.04\) will die [1, 20–22] (see
also ourworldindata.org/coronavirus). By acknowledging these
case-recovery and fatality rates, we obtain a better forecast

\[
x_{i+1}^* = x_{i+1} - px_{i-d} - (1 - p)x_{i-h},
\]

(3)

where the asterisk emphasizes that \(x_{i+1}^*\) is not the value that enters
back into Equation (2) at the next iteration. If that was the case, the forecasted numbers of cases would drop fast. That might
be a reasonable assumption if the number of infected would
approach the population size, and if recovering from COVID-19
would mean becoming immune to the disease [23]. The former
is not yet the case, while the later is also questionable given that
there are reports of individuals being reinfected and the fact that
there are now more different strains of SARS-CoV-2 identified
and that the viral genome is evolving rapidly [24–26] (see also
nextstrain.org/#ncov). Also of note, the values \(h\), \(d\), and \(p\) for
COVID-19 vary significantly in the existing literature [1, 20–
22, 27–29], but it is not the scope of this paper to determine them
accurately. Rather, we use what seem to be reasonable estimates to
illustrate our point. Importantly, sensible variations in \(h\), \(d\), and \(p\)
do not affect the forecast that significantly. The key factor is the
average growth rate \(G_\Delta\), determined as per Equation (1).

We have found \(7 \leq m \leq 14\) to yield good results, whereby
the lower bound ensures a reasonable statistics on \(G_\Delta\) while the
upper bound should still satisfy \(n - 1 - m \geq d\) lest we run out
of data (\(i < 0\)) in \(x_{i-d}\) in Equation (3). We use \(m = 14\) for
the forecasts shown in Figure 1. Lastly, if we wish to rely on actual
data in Equation (3) beyond \(i = n - 1\), and taking into account
\(h < d\), we have to impose a forecasting horizon no longer than
\(n - 1 + h\).

We provide an executable as well as the source code in C for
a straightforward application of the above method on any
data. The executable searches for the file data.txt in the directory
and reads the daily values of confirmed cases, which should be
provided one number per line. The executable also asks for the
year, month, and day of the first entry in the data.txt file, and for
the value of \(m\). The first output file is actual.txt, which contains
three space separate columns, being the date, the number of cases
on said date (returns what is in data.txt minus those recovered and
dead up to then), and the growth rate during the previous
day. The second output file is forecast.txt, which also contains
three space separate columns, being the date, the forecasted
number of cases on said date, and the average daily growth rate
used for the prediction. The forecast is made for thirty different
average daily growth rates, starting from a 20% increased \(G_f\) (as
determined whilst calculating \(G_\Delta\) via Equation 1) and decreasing
in equal intervals toward growth rate zero. Forecasts obtained
with different growth rates are separated with an empty line.

3. FORECAST

Results of the method are shown in Figure 1 for the United States,
Slovenia, Iran, and Germany for 2 weeks onwards from March
29th. If the average growth rates during the past 14 days,
corresponding to \(\approx 30.6\%\) for the United States, \(\approx 9.0\%\) for
Slovenia, \(\approx 7.5\%\) for Iran, and \(\approx 18.7\%\) for Germany, persist,
we will be looking at \(\approx 3.9\) million cases in the United States,
\(\approx 1, 200\) cases in Slovenia, \(\approx 63, 000\) cases in Iran, and \(\approx 380, 000\)
cases in Germany by April 12th, as shown by the solid blue lines
in each graph. If the daily growth rates miraculously dropped
to zero overnight, we would see what is shown with the solid
green lines. That is of course completely unrealistic, but serves
to illustrate what would be the best-case scenario. Solid red lines
show the forecast obtained if the maximal daily growth rate
recorded during the past 14 days, corresponding to \(\approx 48.9\%\) for
the United States, \(\approx 15.5\%\) for Slovenia, \(\approx 9.9\%\) for Iran, and
\(\approx 34.2\%\) for Germany, would increase by 20%. This is not the
worst-case scenario, but it is arguably bad enough. According
to this, Slovenia would have \(\approx 7, 300\) cases by April 12th,
for example.
Given that the exponential growth still persists in all four examples considered in this work—note that the vertical scale in all graphs is logarithmic, and that straight lines thus correspond to exponential growth—the first goal is to arrest this very worrying trend. Between the green and the blue line we show forecasts obtained for daily growth rates between zero and the average of the past 14 days with dashed olive lines. By following the lines from bottom upwards, starting with the solid green line, we can identify the one that flattens out by April 12th. For the United States, for example, it is the 4th line, which corresponds to the $\approx 5.9\%$ daily growth rate from March 29th onwards. This would thus be the target if we wished to see a plateau in the next 2 weeks there. For Germany the same target is $\approx 5.5\%$ (5th line from the bottom), for Slovenia it is $\approx 3.7\%$ (7th line from the bottom), and for Iran it is $\approx 3.6\%$ (10th line from the bottom).

These are of course only approximate target values, but by and large, targeting daily growth rates below at least 5% seems reasonable and in line with what the countries that have thus far successfully responded to the COVID-19 pandemic have achieved.

4. OUTLOOK

As we hope the presented forecasts clearly show, epidemic growth is a highly non-linear process, where every day lost to inaction is a day too much. Even just a few days down the road not acting today can mean the difference between a manageable situation and a hopelessly overburdened healthcare system. The outlook very much depends on whether we take these facts to heart and act accordingly, or not. Governments can impose traveling bans, close down shops and restaurants, and encourage us to stay at home. Ultimately, however, it is on each one of us to respect these restrictions and to do all that we can to minimize the chances for further infections.

Keeping the daily growth rates at least below 5% is an important target for a promising outlook. Data from China,
where the COVID-19 pandemic seems to be coming to an end, confirm this prognosis. Around mid-February, the daily growth rates dropped to around 4% and then to 3% and lower. This marked the beginning of the plateau of confirmed cases, which together with recoveries and deaths led to declining numbers of infected individuals. Singapore, South Korea, and Hong Kong, have also successfully turned their epidemics around by employing strict tactics used in China. Unfortunately, this has not been the case in many other countries [30].

We have two options. The first is to show collective intelligence and restrict our behavior so that new COVID-19 cases will not grow as rapidly as they do now. The second is that we continue to let it slide, until the situation will become so dire that draconian governmental decrees will force us to restrict our behavior [30]. There is still time to act, but a rosy outlook is moving away from us exponentially fast.

**REFERENCES**

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. (2020). doi: 10.1016/S1473-3099(20)30120-1. [Epub ahead of print].

2. Remuzzi G, Remuzzi A. COVID-19 and Italy: what next? Lancet. (2020). doi: 10.1016/S0140-6736(20)30267-9. [Epub ahead of print].

3. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med. (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316

4. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J Evid Based Med. (2020) 13:3–7. doi: 10.1111/jebm.12376

5. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis. (2020) 92:214–7. doi: 10.1016/j.ijid.2020.01.050

6. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 compared to SARS and influenza. J Travel Med. (2020) 27:taaa021. doi: 10.1093/jtm/taaa021

7. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J Med Virol. (2020). doi: 10.1002/jmv.25723. [Epub ahead of print].

8. Ippolito G, Hui DS, Tsountou F, Mauerer M, Zumla A. Taming down the 2019-nCoV media hype – and restoring hope. Lancet Respir Med. (2020). doi: 10.1016/S2213-2600(20)30070-9

9. McCloskey B, Zumla A, Ippolito G, Blumberg L, Arbon P, Cicerò A, et al. Mass gathering events and reducing further global spread of COVID-19: a political and public health dilemma. Lancet. (2020) 395:1096–9. doi: 10.1016/S0140-6736(20)30881-4

10. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Rev Mod Phys. (2015) 87:925. doi: 10.1103/RevModPhys.87.925

11. Wang Z, Bauch CT, Bhattacharyya S, d’Onofrio A, Manfredi P, Perc M, et al. Statistical physics of vaccination. Phys Rep. (2016) 664:1–113. doi: 10.1016/j.physrep.2016.10.006

12. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D. Complex networks: structure and dynamics. Phys Rep. (2006) 424:175–308. doi: 10.1016/j.physrep.2005.10.009

13. Holme P, Saramäki J. Temporal networks. Phys Rep. (2012) 519:97–125. doi: 10.1016/j.physrep.2012.03.001

**DATA AVAILABILITY STATEMENT**

Publicly available datasets were analyzed in this study. The executable, source code and data are available at: http://www.matjazperc.com/COVID-19.

**AUTHOR CONTRIBUTIONS**

MP and AS designed and performed the research. MP, NG, MS, and AS wrote the manuscript.

**FUNDING**

This work was supported by the Slovenian Research Agency (Grant Nos. J4-9302, J1-9112, J2-1725, J3-9289, P1-0403, P3-0396, N3-0048, N3-0133, and I0-0029).

We have two options. The first is to show collective intelligence and restrict our behavior so that new COVID-19 cases will not grow as rapidly as they do now. The second is that we continue to let it slide, until the situation will become so dire that draconian governmental decrees will force us to restrict our behavior [30]. There is still time to act, but a rosy outlook is moving away from us exponentially fast.
28. Jung S, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, et al. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *J Clin Med.* (2020) 9:523. doi: 10.3390/jcm9020523

29. 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. *Lancet.* (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3

30. Cohen J, Kupferschmidt K. Countries test tactics in ‘war’ against COVID-19. *Science.* (2020) 367:1287–8. doi: 10.1126/science.367.6484.1287

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Perc, Gorišek Miksić, Slavinec and Stožer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.