International Journal of Statistical Sciences Vol. 22(2), November, 2022, pp 39-45 © 2022 Dept. of Statistics, Univ. of Rajshahi, Bangladesh

Forecasting SARS-CoV-2 Infections for SAARC Countries

Md. Kaderi Kibria¹⁺, Mohammad Rashedul Islam²⁺, Md. Giash Uddin², Md. Mostafizur Rahman¹, Mahbub Elahi¹ and Md. Nurul Haque Mollah^{1*}

¹Bioinformatics Lab, Dept. of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh ²Department of Pharmacy, University of Chittagong, Chittagong-4331, Bangladesh +Equal contribution

*Correspondence should be addressed to Md. Nurul Haque Mollah (Email: mollah.stat.bio@ru.ac.bd)

[Received March 05, 2022; Accepted August 28, 2022]

Abstract

On March 11th, 2020, the WHO (World Health Organization) announced that the novel coronavirus 2019, a pandemic and at that time, pointing to over 1118,000 cases in over 110 countries and territories around the world. But, now this virus has penetrated into 6 continents which include 210 countries and territories as of April 20, 2020. Since the number of confirmed cases around the world has been surging speedily past the 3 million marks, and death crossed 2lakhs, strengthen the sustained jeopardy of further global spread when this manuscript was writing. Governments around the world are tempting numerous containment measures while the healthcare system is connecting itself for a tidal wave of infected individuals that will seek treatment. So, it is very vital to understand what is expect to in terms of the growth of the number of cases, and to realize what is needed to arrest the very worrying trends. To explain that effect, we here display forecasts obtained with a simple repetitive method that needs only the daily values of confirmed cases as input. The method takes into account anticipated recoveries and deaths, and it determines maximally approved daily growth rates that lead away from exponential increase toward stable and declining numbers. Forecasts expose that daily growth rates should be kept at least below 5% if we wish to see plateaus any time soon. But, unfortunately, most countries are far away from this fact. Here, we provide an executable as well as the source code for a frank application of the method on data from South Asian countries.

Keywords: COVID-19 infection, pandemic, South Asia, exponential growth, Forecasting.

AMS Classification: 92D30.

1. Introduction

A local outbreak of pneumonia initially unrevealed matter was epidemiologically linked to a seafood market at the Hubei province of Wuhan (China) during late December 2019[1]. After that, it was quickly determined to be caused by a novel coronavirus [2]. Finally, it was declared as severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) by the International Committee on Taxonomy Viruses (ICTV) and WHO also named it is Covid-19 [3]. This virus belongs to β coronavirus, a wide class of single-stranded RNA viruses dominant in nature which can affect animals along with humans and can initiate respiratory, hepatic, gastrointestinal, and neurologic disorders [4]. Likely to other viral species, this virus has multiple potential natural hosts, intermediate hosts, and final hosts and in comparison with other viral strains like MERS and SARS, it has intense infectivity and transmissibility, in spite of minute mortality rate [5]. These features constraint the treatment and prevention of this virus and make the situation out of control in many countries of the world. In accordance with real-time data, confirmed cases of coronavirus disease are spreading exponentially throughout the world [6]. Meanwhile, the World Health Organization (WHO) on January 30, 2020, has announced that SARS-CoV-2 is a "public health emergency of international concern" [7]. After that devastating outbreak in Wuhan, the pandemic has also been surcharged the whole health care system of Italy, and Spain [8].

In countries of South Asia, the confirm cases along with deaths are rising rapidly, so it will not take a chronic moment to grab the health care system of the countries of this region. Countries like Bangladesh, India and Pakistan, are over-populated and their situation is becoming worse day by day. So forecasting the prior situation of COVID-19 and its dissemination exhibits a key role [5, 9]. This process of prediction will help to permeate government, professionals of different health care and medical care system what to look forward, and which extensions to appoint. Then, the next step is to motivate the mass people to adhere to the extensions that were appointed to decline the spreading lest a regrettable scenario will expand [10]. Statistical physics has a long and significant history in promoting the research on epidemic processes [11]. The essence of spreading an epidemic can be described by applying simple mathematical models which can use a manageable number of parameters and the values obtained from them can be employed to make informed predictions.

Besides, the research assemblage has proved their abilities in explaining the heterogeneous and complex patterns of communication in social networks by accompanying the frantic shreds of evidence [12, 13]. So generally the behavior of equilibrium and non-equilibrium systems is being determined by using these key evidences. These also provide information regarding the spreading of epidemics and help to make the decision for particular containment strategies. Due to the revolutions in using real-time, social, and mobile by the mass people, a new pattern of epidemiology is being raised. These introduce interdisciplinary investigations at the interface of network science, statistical physics, and epidemiology, directed by large-scale data recording our health and diseases which has given rise to digital epidemiology [14] and the theory of epidemic process on complex networks [11]. Different classical and modern models are employed to assume well-mixed populations, behavioral feedback, and the structure of our social complex networks. These provide us to come a far way in better understanding disease dynamics and transmission.

The knowledge derived from this is helping us to develop effective prevention strategies and in particular [15], we can treat the synergies between these multiple fields of research to ameliorate our lives and societies [16, 17]. However, even the simplest model faces complications during urgency. The little gaps between different fields of research can become can be evident like gaping holes. So, in this paper, we have collected corona virus data up to 15th May for South Asian countries from WHO websites and present a plain repetitive method to forecast the number of COVID-19 cases, under the consideration that governmental data is spontaneous and truthful. The aim is not to fight for discrete accuracy nor to present our method as the realm of the art but simply to provide premier insights and guidelines on elementary principles. It is our congestion if this work inspires further research to yield more detailed and accurate prediction methods.

2. Method

The present method requires only the readily accessible routine values of confirmed cases, as input. We denote these values as xi where $i \in [0, n)$ is the index of days. Assume we have n values available in total, we adopt the last m values of the x_i series and determine the average growth rate during this time according to,

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

We also record the minimal and maximal growth rate during the last m days as $G\downarrow$ and $G\uparrow$, respectively. The simple repetitive process already provides a polished forecast beyond, i = n - 1, pretending the original values are narrated well by exponential growth.

$$x_{i+1} = x_i(1+G_\Delta) \tag{2}$$

Still, this does not tell that after $h \approx 14$ days majority of infected will recover and that after $d \approx 21$ days a fraction $p \approx 0.04$ will die [6, 18, 19]. So, now we acquire a better forecast by analyzing this knowledge of the case-recovery and fatality rate.

$$x_{i+1}^* = x_{i+1} - px_{i-d} - (1-p)x_{i-h_i}$$
(3)

Where the asterisk accents that x_{i+1}^* is not the same value that composes for the equation (2) at the coming iteration. The forecasted laps of cases would drop very fast if we considered that was our case. That might be a feasible assumption if the lap of infected would whet access the population size, and if recovering from the COVID-19, would indicate becoming immune to the disease [20]. But the previous one is not yet the case, while the last one is also questionable provide that there are reports of individuals being re-infected. The fact behind that there are the presence of more multiple strains of n-CoV-2 identified and it is because the viral genome is evolving very rapidly [21, 22]. Nevertheless, the values h, d, and p are designed for COVID-19 discord significantly in the existing literature [6, 18, 19, 23, 24]. But the purpose of the present study is not to conclude them accurately rather we practice what be evident to be reasonable estimates to clarify our point. However, most importantly, this rational variations in h, d, and p do not sense the forecast that significantly. The prime factor is the average growth rate G_{\triangle} , fixed in equation (1). We have established $7 \le m \le 14$ to yield pretty results, by which the lower bound assures a legitimate statistics on G_{\triangle} while the upper bound should still content $n - 1 - m \ge d$ in case we run out of data(i < 0) in x_{i-d} in equation (3). Now, here m = 14 is used to forecast as shown in figure 1. Lastly, if we intend to rely on proper data in equation (3) beyond i = n - 1, and taking into consideration h < d, we have to ascribe a forecasting purview no longer than n - 1 + h.

We give an executable as well as the source code in C for a frank application of the above method on any data. The executable traces for the file data.txt in the directory and interprets the daily values of confirmed cases, which should be given one number per line. The executable further inquiries for the year, month, and day of the first entry in the data.txt file, and for the value of m. The maiden output file is actual.txt, which holds three space segregate columns, being the date, the number of cases on said date (returns what is in data.txt minus those recovered and dead up to them), and growth rate during the previous days. The second output file is forecast.txt, which also holds three space segregate 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 \uparrow (as refuted while calculating G_{\triangle} via equation 1) and decreasing in equal intervals with regard to growth rate zero. Forecasts executed with multiple growth rates are separated with an empty line. To estimate the forecasting error rate for SAARC countries, we considered the mean absolute percentage error (MAPE), where we have actual total cases and forecasting cases from 15th May to 29th May for SAARC countries. The MAPE denoted by M is defined as

$$M = \frac{1}{n} \sum_{t=1}^{n} |\frac{A_t - F_t}{F_t}| \times 100$$

where A_t is the actual value and F_t is the forecast value.

3. Forecast

The results of the proposed method are displayed in Figure 1 for India, Pakistan, Bangladesh, Srilanka, Afghanistan, Nepal, Maldives, and Bhutan for 2 weeks onwards from 15th May.If the average growth rates during the past 14 days, corresponding to $\approx 9.2\%$ for India, $\approx 6.6\%$ for Pakistan, $\approx 20\%$ for Bangladesh, $\approx 4.9\%$ for Srilanka, $\approx 7.5\%$ for Afghanistan, $\approx 14.4\%$ for Nepal, $\approx 14\%$ for the Maldives, $\approx 2.6\%$ for Bhutan, persist, we will be looking at ≈ 54000 cases in India, ≈ 16000 cases in Pakistan, ≈ 49000 cases in Bangladesh, ≈ 350 cases in Srilanka, ≈ 2400 cases in Afghanistan, ≈ 2400 cases in Nepal, ≈ 710 cases in the Maldives, and ≈ 3 cases in Bhutan by May 29th, as shown by the solid blue lines in each graph. If the daily growth rates luckily dropped to zero overnight, we would see what is exhibited with the solid green lines. That is of course radically unrealistic but serves to illustrate what would be the best-case scenario. The solid red lines demonstrate the forecast obtained if the maximal daily growth recorded during the past 14 days, corresponding to $\approx 14\%$ for India, $\approx 9.7\%$ for Pakistan, $\approx 29\%$ for Bangladesh, $\approx 12\%$ for Srilanka, $\approx 33\%$ for Afghanistan, $\approx 87\%$ for Nepal, $\approx 48\%$ for the Maldives and $\approx 20\%$ for Bhutan, would progress by 20\%. This is not the worst-case scenario, but it is reasonably bad enough. According to this, Bangladesh would have ≈ 148000 cases by May 29th, for instance.

Provided that the exponential growth yet persists in all four examples enumerated in this work- a mark that the vertical scale in all graphs is logarithmic, and that straight lines in this manner correspond to exponential growth- the prime goal is to arrest this very worrying trend.

Interim the green and the blue line we display forecasts obtained for daily growth rates between zero and the average of the past 14 days with dashed olive lines. By accompanying the lines from the bottom upwards, starting with the solid green line, we can identify one that flattens out by May 29th.

For India, for instance, it is the 8th line, which corresponds to the $\approx 3.9\%$ daily growth rate from May 15th onwards. This would be the target if we wished to see a plateau in the next 2 weeks there. For Pakistan, the same target is $\approx 4.2\%$ (12th line from the bottom), for Bangladesh, it is $\approx 5.8\%$ (6th line from the bottom), for SriLanka $\approx 3.4\%$ (8th line from the bottom), for Afghanistan $\approx 4.0\%$ (4th line from the bottom), for Nepal $\approx 3.5\%$ (2th line from the bottom), for the Maldives $\approx 7.7\%$ (5th line from the bottom) and finally for Bhutan $\approx 1.6\%$ (3th line from the bottom).

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

Finally, we evaluated the forecasting error rate by using the mean absolute percentage error (MAPE). We observed that the forecasting error rates are 32.49%, 51.97%, 39.75%, 60.06%, 24.60%, 86.64%, 45.34% and 91.13%. For Pakistan, Bangladesh, SriLanka, Afghanistan, Nepal, Maldives, Bhutan and Bhutan, respectively. In the case of Afghanistan, the method gave the lowest forecasting error rate compared to the other seven countries.



Figure 1: Forecasts of COVID-19 infection cases for India, Pakistan, Bangladesh, Sri Lanka, Afghanistan, Nepal, Maldives and Bhutan. Black solid line denotes the actual data, those were last updated May 15th. From this date onward, we thus have the predicted values, where the solid blue line denotes the continuation of the trend of the past 14 days, i.e., if nothing would change. The uppermost solid red line denotes the prediction obtained if the maximal daily growth rate recorded during the past 14 days G \uparrow would increase by 20%, while the lowermost green line denotes the prediction if the daily growth rate would drop to zero from May 15th onward. Plateaus in the next 14 days would be reached if: India target daily growth rate $\approx 3.9\%$ (8th line from the bottom),

Pakistan target daily growth rate $\approx 4.2\%$ (12th line from the bottom), Bangladesh target daily growth rate $\approx 5.8\%$ (6th linefrom the bottom), Sri Lanka target daily growth rate $\approx 3.4\%$ (8th line from the bottom), Afghanistan target daily growth rate $\approx 4\%$ (4th line from the bottom), Nepal target daily growth rate $\approx 3.5\%$ (2th line from the bottom), Maldives target daily growth rate $\approx 7.7\%$ (5th line from the bottom) and Bhutan target daily growth rate $\approx 1.6\%$ (3th line from the bottom).

4. Viewpoint or Outlook

So far we expect the presented forecasts clearly display, epidemic growth is a highly non-linear process, and where every day lost to lassitude is a day too much. Even just a few days down the road not luster today can indicate the intervening between a manageable situation and a hopelessly overburdened health care system. The attention very much relies on whatever we take these facts to heart and act accordingly, or not. The government can impose close down restaurants and shops, public transport, most importantly traveling bans from the foreign and encourage us to stay at home. However, altogether, it is on each one of us to awe these vital restrictions, and to do all that we can minimize the chances for further infections.

Maintaining the daily growth rates at least below 5% is the main target for a promising outlook.

COVID-19 pandemic that attacked China very devastatingly, now this seems to be coming to end as per data from china and confirming this prediction. The daily growth rates in china dropped to around 4% and then to 3% and lower during the mid-February [25]. This indicted the starting of the plateau of confirmed cases, which simultaneous with recoveries and deaths led to declining numbers of infected individuals. South Korea, Vietnam, Singapore, and Hong Kong, have also tremendously turned their epidemics around by applying strict tactics used in China. But, unfortunately, this has not been the case in many other countries [26].

We have actually two options. The maiden one is to show collective intelligence and restrict our movements so that new COVID-19 cases will not augment as rapidly as they do now. The next is that we continue to let it side until the situation will become so grim that draconian government karma will force us to restrict our behavior [26]. There still enough time to act very smartly, but a vivid outlook is moving away from us exponentially fast.

References

- Zhu, N., et al., (2020). A Novel Corona virus from Patients with Pneumonia in China, 2019. New England Journal of Medicine, 2020. 382(8): p. 727-733.
- [2] Li, Q., et al., (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus– Infected Pneumonia. New England Journal of Medicine, 2020. 382(13): p. 1199-1207.
- [3] Wang, L.-s., et al., (2020). A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence. International Journal of Antimicrobial Agents, 2020: p. 105948.
- [4] Weiss, S. R. and J. L. Leibowitz, (2011). Coronavirus pathogenesis. Adv Virus Res, 2011. 81: p. 85-164.
- [5] Liu, Y., et al., (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med, 2020. 27(2).
- [6] Dong, E., H. Du, and L. Gardner, (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, 2020.

- [7] Li, X., et al., (2020). Transmission dynamics and evolutionary history of 2019-nCoV. Journal of Medical Virology, 2020. 92(5): p. 501-511.
- [8] Remuzzi, A. and G. Remuzzi, (2020). COVID-19 and Italy: what next? Lancet, 2020. 395(10231): p. 1225-1228.
- [9] Lai, A., et al., (2020). Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. Journal of Medical Virology, 2020. 92(6): p. 675-679.
- [10] Ippolito, G., et al., (2020). Toning down the 2019-nCoV media hype—and restoring hope. The Lancet Respiratory Medicine, 2020. 8(3): p. 230-231.
- [11] Pastor-Satorras, R., et al., (2015). Epidemic processes in complex networks. Reviews of Modern Physics, 2015. 87(3): p. 925-979.
- [12] Lü, L., et al., (2016). Vital nodes identification in complex networks. Physics Reports, 2016. 650: p. 1-63.
- [13] Kivelä, M., et al., (2014). Multilayer networks. Journal of Complex Networks, 2014. 2(3): p. 203-271.
- [14] Salathe, M., et al., (2012). Digital epidemiology. PLoS Comput Biol, 2012. 8(7): p. e1002616.
- [15] Wang, Z., et al., (2016). Statistical physics of vaccination. Physics Reports, 2016. 664: p. 1-113.
- [16] Helbing, D., et al., (2015). Saving Human Lives: What Complexity Science and Information Systems can Contribute. Journal of Statistical Physics, 2015. 158(3): p. 735-781.
- [17] Perc, M., (2019). The social physics collective. Scientific Reports, 2019. 9(1): p. 16549.
- [18]Huang, C., et al., (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 2020. 395(10223): p. 497-506.
- [19] Lazzerini, M. and G. Putoto, (2020). COVID-19 in Italy: momentous decisions and many uncertainties. Lancet Glob Health, 2020.
- [20] Chen, D., et al., (2020). Recurrence of positive SARS-CoV-2 RNA in COVID-19: A case report. International Journal of Infectious Diseases, 2020. 93: p. 297-299.
- [21] Zhou, P., et al., (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 2020. 579(7798): p. 270-273.
- [22] Ceraolo, C. and F. M. Giorgi, (2020). Genomic variance of the 2019-nCoV coronavirus. Journal of medical virology, 2020.
- [23] Lauer, S. A. and K. H. Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler (2019). The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine, 2020. 3.
- [24] Jung, S.-m., et al., (2020). Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. Journal of clinical medicine, 2020. 9(2): p. 523.
- [25] Perc M, Gorišek Miksi' c N, Slavinec M. and Stožer A (2020). Forecasting COVID-19. Front. Phys. 8:127.
- [26] Cohen J, Kupferschmidt K. Countries test tactics in 'war' against COVID-19. Science. (2020). 367:1287–8.