Global Epidemiology 2 (2020) 100023

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: https://www.journals.elsevier.com/global-epidemiology

Commentary

1,000,000 cases of COVID-19 outside of China: The date predicted by a simple heuristic

W.W. Koczkodaj[a,*], M.A. Mansournia[b], W. Pedrycz[c], A. Wolny-Dominiak[d], P.F. Zabrodskii[e], D. Strzałka[f], T. Armstrong[a], A.H. Zolfaghari[a], M. Dębski[g], J. Mazurek[h]

[a] Computer Science, Laurentian University, Sudbury, Ontario P3E 2C6, Canada
[b] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[c] Department of Electrical & Computer Engineering, Alberta University, Edmonton, Alberta, Canada
[d] University of Economics in Katowice, 1 Maja 47, 40-287 Katowice, Poland
[e] Saratov Medical University, Reaviz, Saratov, Russian Federation
[f] Rzeszów University of Technology, Powstańców Warszawy 12, 35-959 Rzeszów, Poland
[g] Department of Sociology, University of Gdańsk, Bażyńskiego 8, 80-309 Gdańsk, Poland
[h] Silesian University in Opava, Opava, Czech Republic.

ARTICLE INFO

Available online 23 March 2020

ABSTRACT

We forecast 1,000,000 COVID-19 cases outside of China by March 31st, 2020 based on a heuristic and WHO situation reports. We do not model the COVID-19 pandemic; we model only the number of cases.
The proposed heuristic is based on a simple observation that the plot of the given data is well approximated by an exponential curve. The exponential curve is used for forecasting the growth of new cases. It has been tested for the last situation report of the last day. Its accuracy has been 1.29% for the last day added and predicted by the 57 previous WHO situation reports (the date 18 March 2020).
Prediction, forecast, pandemic, COVID-19, coronavirus, exponential growth curve parameter, heuristic, epidemiology, extrapolation, abductive reasoning, WHO situation report.

1. Introduction

Using WHO situation reports for Coronavirus disease 2019 (COVID-19), this study forecasts 1,000,000 confirmed cases outside of China in approximately two weeks. So far, #59 situation reports have been posted by WHO (the date 20 March 2020). In this study we refer to reports #31- #57.

Due to potentially overwhelming numbers of severe COVID-19 patients, medical resources need to be allocated wisely. With hospital beds and life-saving machinery, such as ventilators in limited supply, preparations should be made ahead of time on how to allocate these finite resources. More information about COVID-19 can be found in [2,3,7]. The best course of action to “flatten the curve” is to follow WHO guidelines. The best way to keep hospitals under capacity is social distancing: limiting or cancelling large gatherings, only travelling when necessary, and keeping a distance from others all help to prevent the spread.

* Corresponding author.
URL's: wkoczkodaj@cs.laurentian.ca (W.W. Koczkodaj), mansournia_ma@yahoo.com (M.A. Mansournia), pedrycz@ee.ualberta.ca (W. Pedrycz), alicja.wolny-dominiak@ue.katowice.pl (A. Wolny-Dominiak), pfzabrodsky@gmail.com (P.F. Zabrodskii), strzalka@prz.edu.pl (D. Strzałka), tarmstrong@laurentian.ca (T. Armstrong), azolfaghari@laurentian.ca (A.H. Zolfaghari), maciej.debski@ug.edu.pl (M. Dębski), mazurek@opf.slu.cz (J. Mazurek).

https://doi.org/10.1016/j.gloepi.2020.100023

2. Heuristic prediction

Heuristic

(1) Load WHO situation report data into data frame.

(2) Use R package `nls` to fit the non-linear model:

$$y = f(x) = a * e^{b*x}$$

to data by computing a and b.

(3) Set x to the day number of the WHO situation report data.

(4) In a look, use the exponential curve $f(x) = a * e^{b*x}$ to compute the predictions.

The presented heuristic is based on the exponential growth of the data collected by WHO situation reports for days 31 to 57.

As pointed out in [4] the predictability could be improved by pairwise comparisons based on abductive reasoning [5]. Abduction is frequently used in diagnostic expert systems. The abductive reasoning (or inference) process was used for this study. It is a type of logical inference which starts with a set of observations and then searches for the simplest and most likely explanation for the observations. In our case, the most likely explanation is exponential growth. This process yields a plausible conclusion but may not always positively verify it. The abductive conclusions are heuristics (see [1]), hence involve uncertainty, which is expressed by the bounded rationality as satisficing. Satisficing is a decision making process which takes into account the costs of optimization into the optimization process, thereby producing an efficient but suboptimal result. This can be compared with maximizing, which produces an optimal result at the expense of suboptimal costs.

The extrapolation is a mathematical estimation, predicting unknown future values based on existing values. Compared to interpolation, which determines unknown values between existing values, extrapolation is less accurate. The best method for extrapolation is dependent on which method was used to initially acquire the data.

The WHO situation report #31 (see [7]) has been assumed as the starting data point since it shows, for the first time, over 1000 cases outside China (see Fig. 1).

Due to the risk of data from any individual country being biased or politically motivated to misreport data, we decided to use data from many countries; as such, any doctored data becomes statistically insignificant. In China, where COVID-19 originated, the situation seems to be under control as the Fig. 2 indicates.

For this reason, including data about China would deviate the results or at least make them difficult to obtain.

The visual inspection suggested the exponential growth, but could not be assumed. As such, R code was needed to be used for it with its nls function. According to [6]:

Nonlinear Least Squares (nls) determines the nonlinear (weighted) least-squares estimates of the parameters of a nonlinear model. An nls object is a type of fitted model object. It has methods for the generic functions anova, coef, confint, deviance, df.residual, fitted, formula, log- Lik, predict, print, profile, residuals, summary, vcov and weights.

Variables in formula (and weights if not missing) are looked for first in data, then the environment of formula and finally along the search path. Functions in formula are searched for first in the environment of formula and then along the search path.

For more details see [8]. We consider a non-linear model of the form:

$$y_i = f\ (x_i; a, b) + \varepsilon_i, i = 1, \ldots, n \tag{1}$$

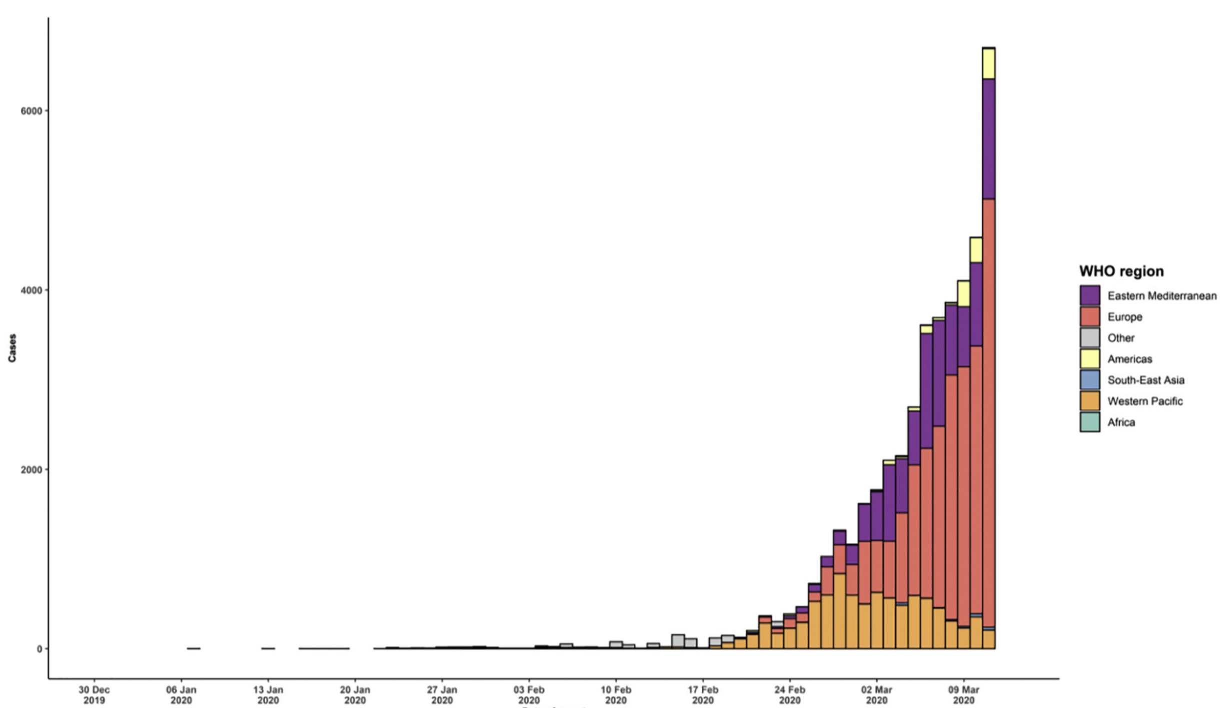

Fig. 1. WHO #31 report countries, territories or areas with reported confirmed cases of COVID-19, 07 March 2020.

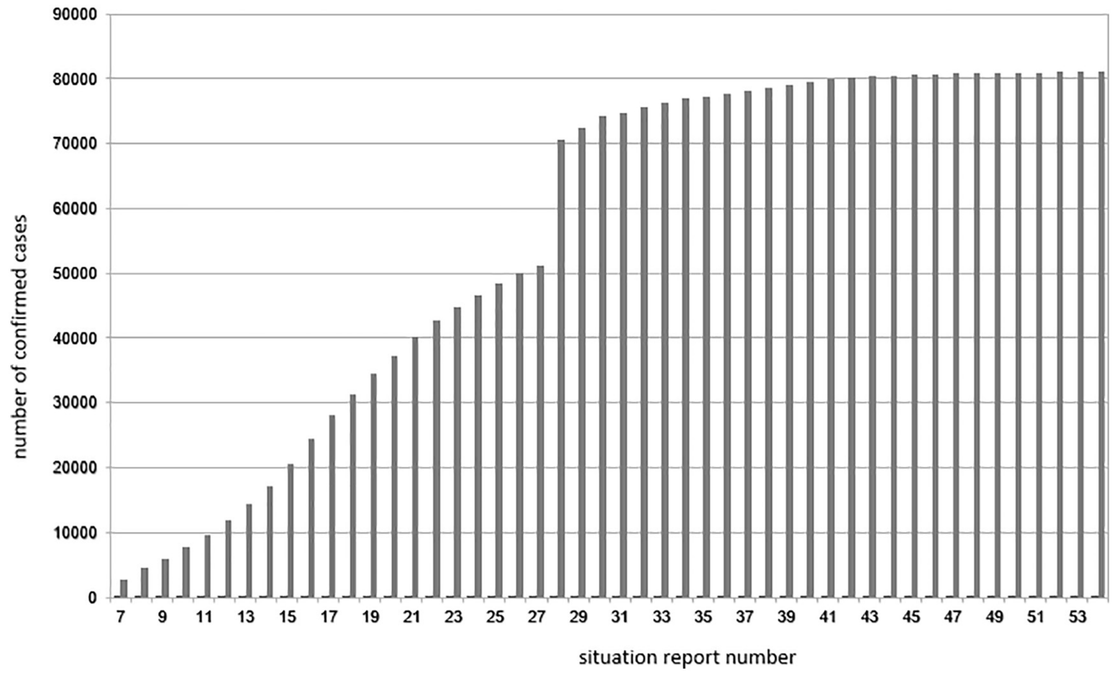

Fig. 2. WHO situation report for China.

with type exponential function $f(.)$ of the form:

$$f(x) = a \ \exp(bx) \tag{2}$$

In order to estimate the parameters a, b, we apply the non-linear least squares method, in which the residual sum of squares is minimized, see [8]:

$$S_n(a,b) = \sum_{\{i=1\}}^{n} [y_i - f(x_i;a,b)]^2 \tag{3}$$

where y_i is the number of total infected by COVID-19 outside China. In a, b parameters estimation we use well-known nls function from R program receiving:

par.	Estimated	Std. Error	p-value
a	10.4791	1.75	0.00
b	0.161	0.003	0.00

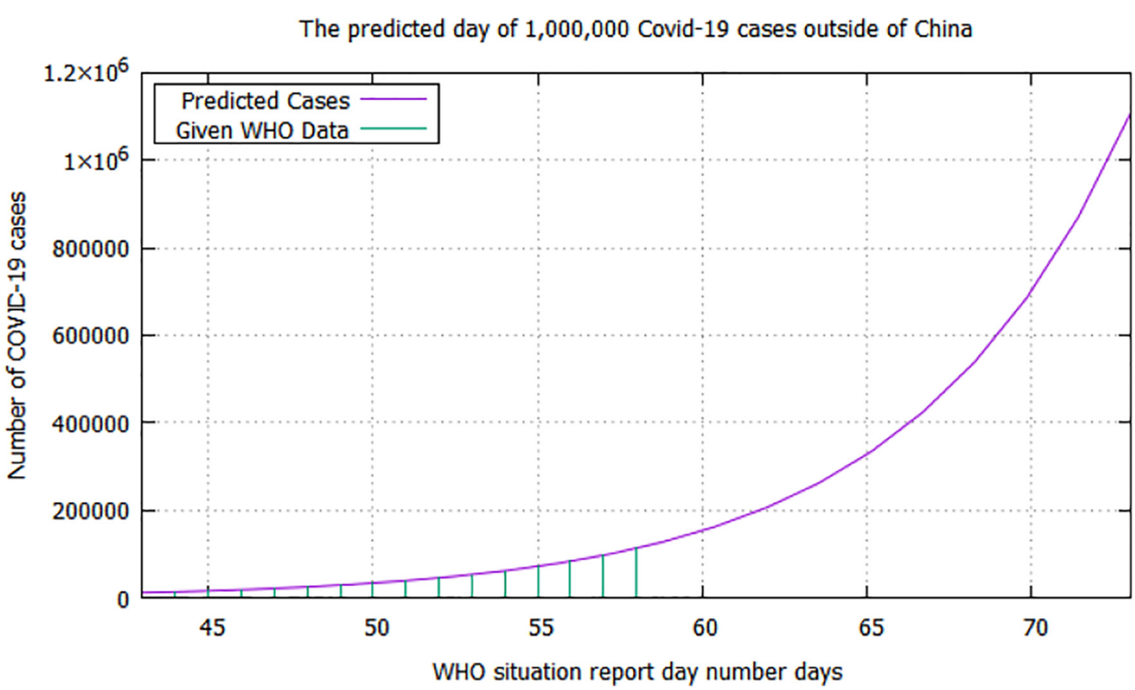

Fig. 3. Prediction of 1,000,000 cases of COVID-19 by WHO situation report data for outside China.

The residual standard error is $S_n = 1827$. According to these results, we predict 1,000,000 COVID-19 cases outside of China by the WHO situation report day 70/71 which is 31 March/01 April (see Fig. 3).

The lines of the plot, up to the last day of WHO situation report, are:

(1) the blue line connecting 18 March WHO data,
(2) the red line standing for 1,000,000 cases,
(3) the exponential curve computed by R to be as close as possible to the real data up to 18 March.

The vertical blue bar (Fig. 3) shows where the WHO data ends and where the predicted results start. For this reason, on the right hand side of the vertical bar there is only one line which is the computed exponential curve.

Evidently, we do not have knowledge of how long (in terms of days) such an exponential curve will be an acceptable extrapolation; a million cases in 16 days, however, seems to have a high likeliness. Such a finding has considerable importance and should not be ignored.

3. Conclusions

To the best of our knowledge, this may be the first study proposing a heuristic for computing parameters a and b for the approximating exponential curve $a * exp(b * x)$ and for using x as the day number for the COVID-19 situation. The more people know about our finding, the better chance that they may regard self-care as a major contribution to preventing the spread of COVID-19. Our assumptions do not consider the complexity of a pandemic. In particular, we do not consider flattening of the approximating exponential curve.

Simply, it is a short term prediction model, but it is very simple and we believe it is very accurate. As for the prediction standards, 1.29% error is more than acceptable for short term predictions.

We regard the WHO situation report #31 as the starting data point since it shows over 1000 cases outside China for the first time. The presented approach is based on a heuristic solution and makes a realistic assumption that the current trend can continue for the next 17 days. Obviously, it is an abstract, mathematical model; the reality may be different and COVID-19 situation may change in just a few days.

Acknowledgment

There is no conflict of interest and this study has been conducted pro bono publico.

References

[1] Brown R. Consideration of the origin of Herbert Simon's theory of satisficing (1933-1947). Man Decision 2004;42:1240–56.
[2] Heymann DL, Shindo N. COVID-19: what is next for public health? The Lancet 2020; 395:542–5. https://doi.org/10.1016/S0140-6736(20)30374-3.
[3] Murthy S, Gomersall CD, Fowler RA. Care for critically ill patients with COVID-19. JAMA 2020. https://doi.org/10.1001/jama.2020.3633 In press.
[4] Kakiashvili T, Koczkodaj WW, Woodbury-Smith M. Improving the medical scale predictability by the pairwise comparisons method: evidence from a clinical data study. Comput Methods Programs Biomed 2012;105:210–6.
[5] Paul G. Approaches to abductive reasoning: an overview. Artificial Intelligence Review 1993;7:109–52.
[6] Venables WN, Smith DM, The R Core Team. An introduction to R, Notes on R: A programming environment for data analysis and graphics, version 3.6.3. https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf; 2020.
[7] WHO. Coronavirus disease (COVID-19). https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200226-sitrep-37-covid-19.pdf?sfvrsn=2146841e_2; 2020. (Accessed Date: 20 March, 2020).
[8] Wu C-F. Asymptotic theory of nonlinear least squares estimation. Annal Stat 1981;9: 501–13.