1,000,000 CASES OF COVID-19 OUTSIDE OF CHINA: THE DATE PREDICTED BY A SIMPLE HEURISTIC

W.W. KOCZKODAJ, M.A. MANSOURNIA, W. PEDRYCZ, A. WOLNY-DOMINIAK, P.F. ZABRODSKII, D. STRZALKA, T. ARMSTRONG, A.H. ZOLFAGHARI, M. DEBSKI, AND J. MAZUREK

Abstract. Prediction of 1,000,000 COVID-19 cases outside of China based on WHO situation report has been computed by a heuristic for March 30, 2020. It is not modeling the COVID-19 pandemic but predicting a number of cases only.

The proposed heuristic is based on a simple observation that the plot of the given data is well approximated by an exponential curve. Parameters of the curves have been computed by R code. 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 the by 54 previous WHO situation reports. Preliminary findings will be posted at `WWK` and submitted to `arXiv` for expeditious posting.

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, 55 situation reports have been posted by WHO.

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], and [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 traveling when necessary, and keeping a distance from others all help to prevent the spread.

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.

2. Heuristic prediction

The presented heuristic is based on the exponential growth of the data collected by WHO situation reports for days 31 to 55. abductive reasoning in [5] describes reasoning as inference to the best explanation. As pointed out in [4]. In the 1990s, research in the fields of computer science, law, and artificial intelligence utilized abduction. Today, 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. 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.

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.

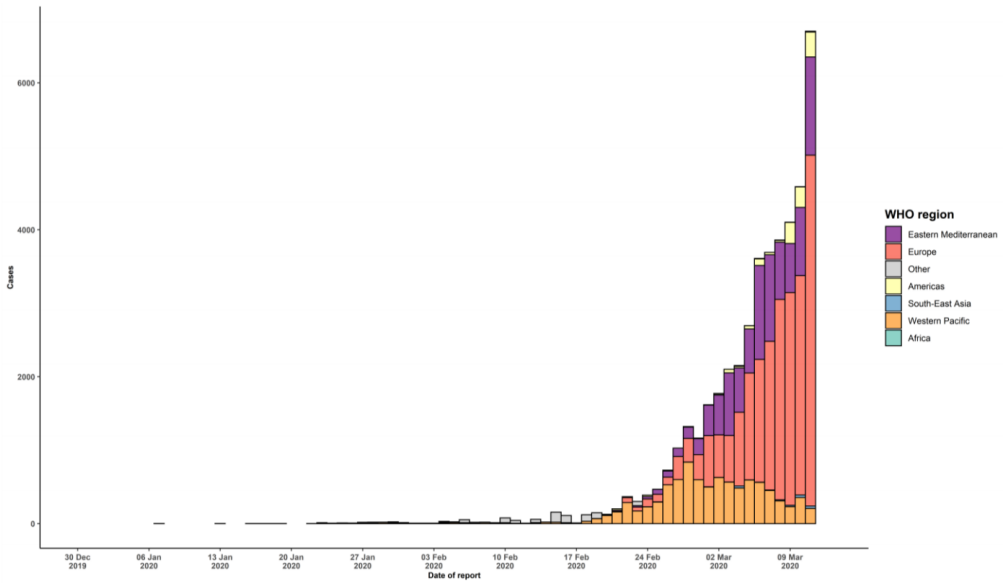

FIGURE 1. WHO report – Countries, territories or areas with reported confirmed cases of COVID-19, 07 March 2020

Due to the risk of the 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 situations seems to be under control as the Fig. 2 indicates.

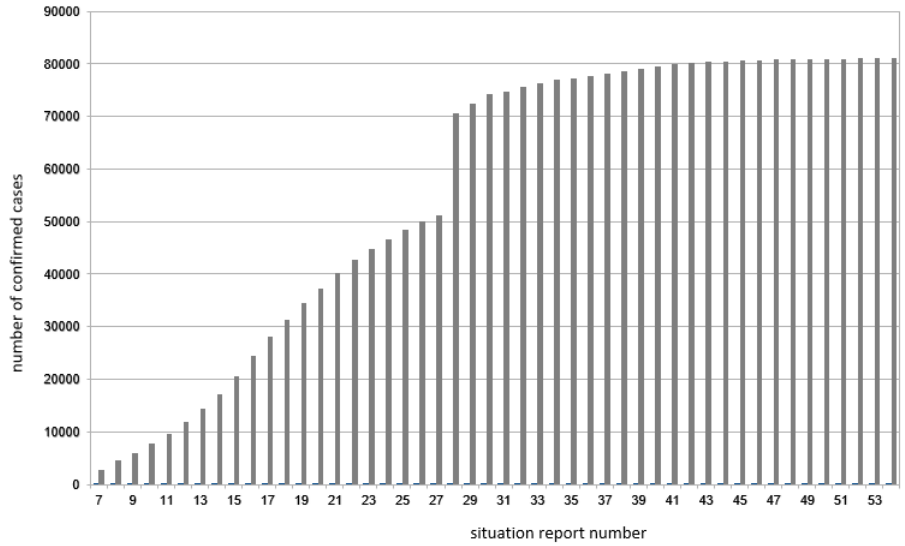

FIGURE 2. WHO situation report for China

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, logLik, 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:

$$y_i = f(x_i; a, b) + \epsilon_i, i = 1, ..., n \tag{2.1}$$

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

$$f(x) = a \exp^{bx} \tag{2.2}$$

The WHO situation report #31 (see [7]) has been assumed as the starting data point since it shows, for the first time, over 1,000 cases outside China, see [7]. 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]:

$$S_n(a, b) = \sum_{i=1}^{n} [y_i - f(x_i; a, b)], \tag{2.3}$$

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	10.45	1.11	0.00
b	0.16	0.002	0.00

The residual standard error is $S_u = 741$. According to this results, we predict 1,000,0000 COVID-19 cases outside of China by the WHO situation report day 70 which is March 30.

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

1. the line connecting WHO data,
2. the exponential curve computed by R to be as close as possible to the line in point 1.

The vertical bar 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, it 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. The more people know about our finding, the better chance that they may regard the self-care as a major contribution to prevention spreading COVID-19. Our assumption do not consider the complexity of a pandemic. In particular, it does not consider *flattening* of the approximating exponential curve. Simply, it is a short term prediction model but it is very simple and very accurate.

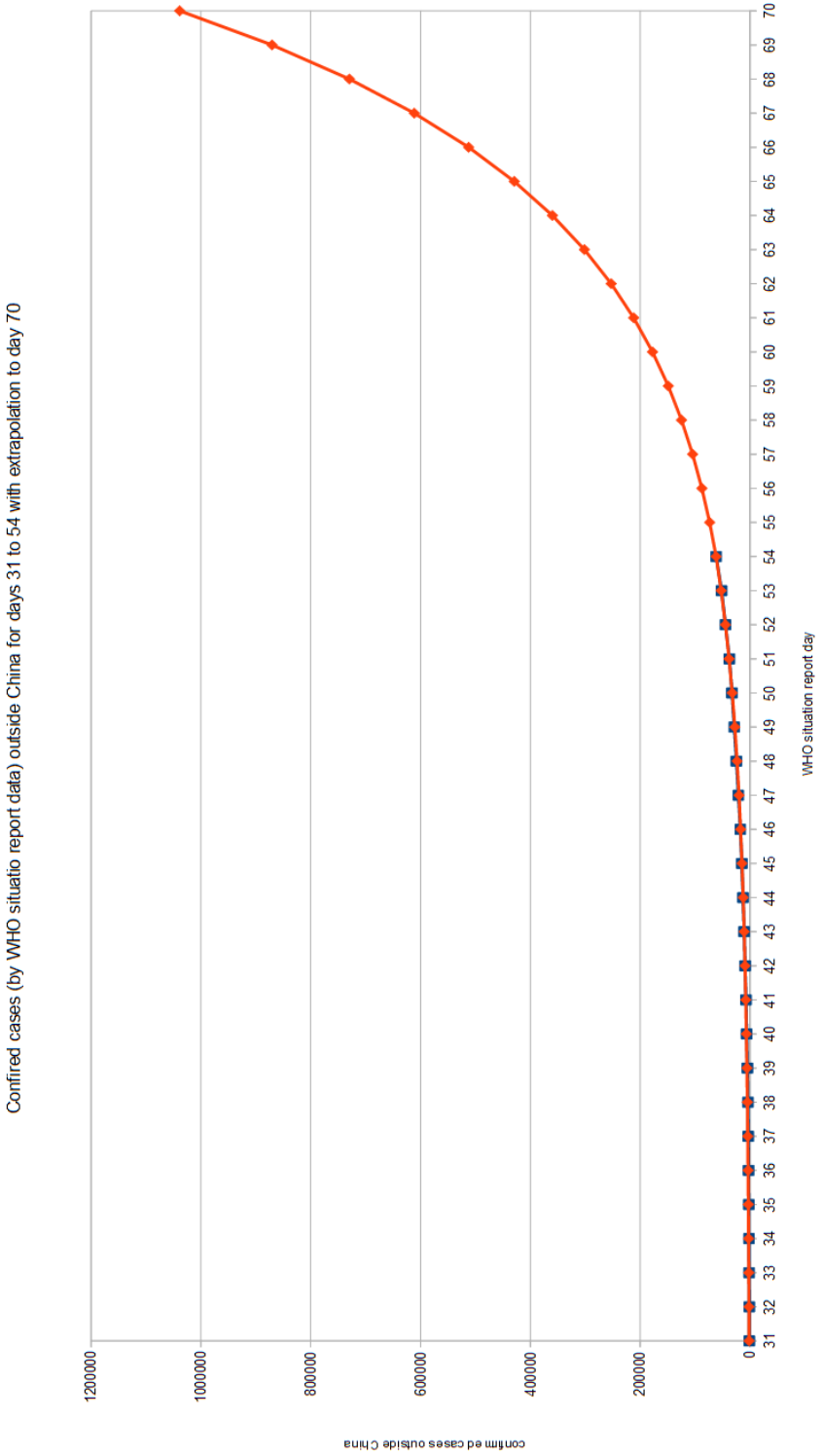

FIGURE 3. Prediction of 1,000,000 cases of COVID-19 by WHO situation report data for outside China

As for the prediction standards, 1.29% error is more than acceptable for the short term predictions.

We regard the WHO situation report #31 as the starting data point since it shows over 1,000 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. Evidently, it is a mathematical abstract 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), 42(10): 1240–1256, 2004.

[2] Heymann, D.L.;Shindo, N., COVID-19: What is next for public health?, 395, no. 10224, pp. 542–545, 2020.

[3] Murthy, S.; Gomersall, C.D.; Fowler, R.A. , Care for Critically Ill Patients with COVID–19, JAMA online, 2020.

[4] Kakiashvili, T.; Koczkodaj, W.W.; Woodbury-Smith, M., Improving the medical scale predictability by the pairwise comparisons method: Evidence from a clinical data study, Computer Methods and Programs in Biomedicine, 105(3): 210–216, 2012.

[5] Paul, G., Approaches to Abductive Reasoning - an Overview, Artificial Intelligence Review, 7(2): 109–152, 1993.

[6] Venables, W.N.; Smith, D.M. and the R Core Team, An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics Version 3.6.3 (2020–02=-29)

[7] WHO, Coronavirus disease (COVID-19), 2020 vol.

[8] Wu, C.-F., Asymptotic Theory of Nonlinear Least Squares Estimation, 9(3): 501–513, 1981.

99

Computer Science, Laurentian University, Sudbury, Ontario P3E 2C6, Canada
Email address: wkoczkodaj@cs.laurentian.ca

Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

Dept of Dept Elect & Comp Engn., Alberta University, Edmonton, Alberta, Canada

University of Economics in Katowice, 1 Maja 47, 40-287 Katowice, Poland
Email address: alicja.wolny-dominiak@ue.katowice.pl

Saratov Medical University, Reaviz, Saratov, Russian Federation

Rzeszow University of Technology, Rzeszow, Poland
Email address: strzalka@prz.edu.pl

Computer Science, Laurentian University, Sudbury, Ontario, Canada
Email address: tarmstrong@laurentian.ca

Computer Science, Laurentian University, Sudbury, Ontario, Canada
Email address: azolfaghari@laurentian.ca

Department of Sociology, University of Gdansk, Gdansk, Poland

Silesian University in Opava, Opava, Czech Republic