

## Forecasting Confirmed Cases and Mortalities of COVID-19 in the US

**Babak Jamshidi<sup>1</sup>, Mohsen Kakavandi<sup>2</sup>, Shahriar Jamshidi Zargaran<sup>3</sup>, and Amir Talaei-Khoei<sup>4</sup>**

- 5 1. Postdoctoral Researcher in Biostatistics, Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran, [babak.j6668@gmail.com](mailto:babak.j6668@gmail.com)
2. Master of Science in Mechanic Engineering, Poznan University of Technology, [mohsen.kakavandy@yahoo.com](mailto:mohsen.kakavandy@yahoo.com)
- 10 3. Master of Science in Medical engineering, Tehran University of Medical Sciences, Tehran, Iran [shahriarjamshidy@gmail.com](mailto:shahriarjamshidy@gmail.com)
4. Associate Professor of Information Systems, [Department of Accounting and Information Systems](#), University of Nevada, Reno, Nevada, USA, [atalaeikhoei@unr.edu](mailto:atalaeikhoei@unr.edu)

15

---

### Abstract

*Background* - The wide spread of COVID-19 in the US has placed the country as the most infected population worldwide. This paper aims to forecast the number of confirmed cases and mortalities from 12 April  
20 to 21 May, 2020. There has been a large body of literature in forecasting epidemic outbreaks such as C algorithms with shortfall of predicting for long periods and autoregressive integrated moving average models with the limited flexibility. However, the US COVID-19 data shows great variety in the relative increments of confirmed  
25 cases. This requires a reproductive time series.

*Method* – This paper suggests a time series based on the relative increments of confirmed cases. The proposed time series assumes the changes in the time series and provides flexibility. The suggested model was applied on the data observed from 27 February to 11 April 2020  
30 and its objective is forecasting 40 days from 12 April to 21 May 2020.

*Results* - It is expected that by May 21, 2020, the accumulative number of confirmed cases of COVID-19 in the US rises to 1,464,729, with 80% confidence interval. Our analysis also shows that by the 21<sup>st</sup> of May, the cumulative number of mortalities caused by COVID-19 in  
35 the US from 18747 cases on 11 April increases to around 73250 cases on 21 May, 2020.

*Conclusion* - Our results highlight the value of reproductive strategies in time series modelling of COVID-19. Our model benefits from a reproductive strategy from a time point in which the US  
40 COVID-19 data demonstrates a sudden fall.

*Keywords:* COVID-19; Relative Increment; Forecast

Page Numbers: 32 Pages

Number of lines: 511 lines

## 1. Introduction

An epidemic disease of unknown cause, now called COVID-19, found in Wuhan, China was first reported to the World Health Organization Office in China on 31 December 2019. Shortly after, it was found that the ongoing pandemic is an infectious disease caused by the coronaviruses, a large family of viruses that cause illness ranging from the common cold to more severe diseases such as MERS and SARS. The outbreak was declared as a Public Health Emergency worldwide on 30 January 2020. Based on its geographic spread, severity of illnesses it causes, and also its impact on society, on March 11, 2020, COVID-19 was declared as a pandemic outbreak (WHO, 2020a). As of the 8<sup>th</sup> of April, 2020, the disease has spread to over 210 countries and infected 1,353,361 cases when causes the total Mortalities of 79,235 worldwide (WHO, 2020b).

The first confirmed case in the US was reported on 21 January, 2020 and on the 22nd February the first death was reported (WHO, 2020c). On the 17th of March, the US reported the arrival of the disease to all the states (CDC, 2020). In one month, this number is multiplied by over 10000. As of 11 April 2020, about one-third of all confirmed cases

65 infected by COVID-19 in the US belongs to New York. New York by  
almost 160 K confirmed cases, has reported the number of confirmed  
cases more than all countries (other than the US) (CDC, 2020). It is  
noticeable that up to now, USA has faced the greatest new mortalities  
(2035) and new confirmed cases (34196) in one day on 10 and 4 April,  
70 respectively. In addition, Active cases in USA on 11 April were 481849  
which is approximately equal to active cases of the other eight countries  
with the most confirmed cases (Spain, Italy, France, Germany, China,  
United Kingdom, Iran, Turkey) (WHO, 2020c)

We use the US recent COVID-19 data from 27 February to 11 April  
75 2020 to model and forecast the development of the outbreak in the US  
for the period of April 12th – May 21st, 2020 predicting the relative  
increment of the confirmed cases and Mortalities using two-part time  
series. The first part represents the propagation of the disease in the first  
period in which the rate of transition is high. The second part models an  
80 irreversible fall in the relative increment and continues the extinction of  
the disease.

There has been a body of research presenting forecasting trends in  
fields such as epidemiology (Chretien et al., 2014; Paul et al., 2014;  
Talaie-Khoei et al., 2019), economics (Baltagi and Baltagi, 2001; Scott

85 and Varian, 2015), weather (Kumar and Jha, 2013; Radzuan et al.,  
2013; Voyant et al., 2013). Forecasting epidemic outbreaks plays a  
significant role in the success of epidemiology and public health  
systems to effectively respond to the epidemic situations. As such, there  
are several attempts suggesting methods for forecasting the confirmed  
90 cases in outbreaks. Chretien et al. (2014) reviews the methods in  
forecasting influenzas in human populations. From theoretical  
standpoint, forecasting methods are provided with time series data from  
previous stages of an outbreak and predicts surveillance attributes of a  
disease such as the confirmed cases in future. In order to do so, there  
95 have been several statistical methods (Reis and Mandl, 2003; Rounds et  
al., 2017; Zeger et al., 2006; Zhang et al., 2014) suggested by literature.  
Accordingly, there have been several attempts by professional bodies.  
For example, The Centres for Disease Control and Prevention initiated  
a challenge to forecast the 2013-2014 United States influenza season  
100 (Biggerstaff et al., 2016). Similar challenges was also organized for  
Ebola (Viboud et al., 2018), and Dengue (NOAA, 2016). The literature  
in time series forecasting has used aberrancy-detection algorithms to  
identify temporal changes in the data. These temporal changes may be  
the indicator of epidemic outbreaks (Murphy and Burkom, 2008). The

105 Early Aberration Reporting Systems from Centres for Disease Control  
and Prevention utilizes C algorithms. However, C algorithms suffer  
from short-term prediction capabilities (Tokars et al., 2015). General  
speaking, aberrancy-detection algorithms focuses on a single time point  
in a given time when the outbreaks occurs and they are limited to be  
110 used to forecast the number of confirmed cases in the future of outbreak  
(Kass-Hout et al., 2012). Zhang (2003) suggests the use of  
Autoregressive Integrated Moving Average (ARIMA) to address the  
shortfall of aberrancy-detection algorithms. ARIMA models have been  
largely used in forecasting infectious diseases such as dengue  
115 (Wongkoon et al., 2012) and tuberculosis (Rios et al., 2000). ARIMA  
models are committed to the assumption that in the Autoregressive  
model under the study, the present value of the time series is a linear  
function of the past values and random noise (Akaike, 1969). Second,  
the present value of the time series is a linear function of present and  
120 past residuals in the moving average model (Haining, 1978). Third,  
ARIMA is based on both Autoregressive and moving average as well  
as the past values and residuals (Rojas et al., 2008). These assumptions  
limit the flexibility of ARIMA models to reproduce the different stages  
of time series.

125                    **Looking at data from COVID-19, it appears that the**  
**outbreak has demonstrated decreasing relative increment**  
**in the number of new cases. Since ARIMA models focuses**  
**on the linear relationships between the present values**  
**with past data and results, the forecasting ability of**  
130                    **ARIMA models for COVID-19 would be limited.**  
**Addressing the shortfall of ARIMA models, this paper,**  
**using a novel model to reproduce the time series of**  
**COVID-19, aims to forecast the number of new cases in**  
**the US. The present paper adopts a stochastic approach**  
135                    **due to two main reasons. Firstly, there are several sources**  
**of uncertainty in a pandemic outbreak (Chowell et al.,**  
**2020) such as COVID-19. Randomness is a major player**  
**particularly at the beginning of an outbreak. Secondly,**  
**the stochastic approach used in our model promotes the**  
140                    **flexibility of the model. The proposed method commits us**  
**to three assumptions. Firstly, the relative increase in the**  
**number of confirmed cases are increasing at the**  
**beginning, then it experiences a sudden fall and it**  
**decreases thereafter. Looking at the data from COVID-19**

145 **in the US, we can confirm that this is the case, See Day 0:**

27 February 2020, Day 45: 12 April 2020.

Figure 1. Secondly, at any time, the time series of relative increase in the confirmed cases holds a normal distribution. The COVID-19 in the US demonstrates a normal distribution for the relative increase of the confirmed cases. Third, over time, the ratio of variance to mean  
150 remains constant, which is the case for COVID-19 in the US.

The present paper relates to the recent articles attempting to model the spread of the COVID-19 19 (Berger et al., 2020; Kucharski et al., 2020; Read et al., 2020). The present study differs from these papers  
155 from Data and Method perspectives. Berger et al. (2020) study the data from the United States and it investigates the role of testing and quarantine as two effective practices to contain the disease. The article does not use time series analysis. Kucharski et al. (2020) look into cases in Wuhan and internationally exported cases from Wuhan. The  
160 article addresses uncertainty of COVID-19 by using some outside databases, while in this study we only utilize data from the US. Read et al. (2020) explores the early data from Wuhan and focuses on epidemiological parameter estimates. However, the present paper

mainly looks at time series analysis to forecast the newly confirmed  
165 cases.

The rest of this paper is organized in the following way, Section 2  
presents the proposed method. Section 3 presents the results of our  
forecast. Section 0 discusses the paper and provides some limitations as  
170 well as directions for future research.

## 2. Methods

In this section, we will discuss the proposed model, and the  
estimations. We will also simulate the model and the estimation on the  
data from the US to forecast for the next 40 days. The calculations,  
175 simulations have been conducted in MatLab R2015b. In addition,  
we have opted five output forecasting variables to compare, namely  
the average curve and 80% upper bound, 80% lower lower bound  
and two realizations. All these outputs are computed based on 100  
times simulations of the fitted models.

180 The data in this section is obtained from WHO situations reports  
(WHO, 2020c) where we collected the data for the US from 27

February to 11 April 2020 and the paper forecasts from 12 April to 21 May, 2020.

## 2.1 Definition of Relative Increment of Confirmed Cases and Mortality

185 *Rate*

*Relative Increment of Confirmed Cases:* Let  $Y_t$  is a time series of the number of confirmed cases up to time  $t$ . We aim at studying  $X_t = \frac{Y_{t+1}}{Y_t} - 1$ , called the daily relative increment of confirmed cases.

*Mortality Rate:* In order to define the mortality rate of COVID-19 in the US, we define

$$Z_t = \frac{\text{The cumulative number of death in the day } i}{\text{The cumulative number of confirmed cases in day } (i-5)}.$$

## 2.2 Proposed Time Series

The model we applied has five positive parameters  $(b, IR, K, \theta, a)$ ;

For  $t = 1, \dots, b$   $X_t \sim \text{Normal}(IR, \frac{2*IR^2}{a})$  which confirms the

195 second assumption of our time series, See Section 1.

1. For  $t = b + 1, b + 2, \dots$

$X_t \sim \text{Normal}(K/t^\theta, K^2/at^{2\theta})$ , where;

- $b$ : The length of the first periods of the spread of the disease,

200 2.  $IR$ : The geometric mean of the relative increment in the first periods of the spread of the disease,

3.  $\theta$ : The acceleration of falling of the relative increments after the first days of spreading ( $X_t \propto t^{-\theta}$  for  $t = b + 1, b + 2, \dots$ ),

205 4.  $a$ : The fixed ratio of the mean to the variance ( $\sqrt{a} = \frac{\text{Expectation}}{\text{Standard deviation}}$ ), which confirms the third assumption of our time series, See Section 1.

210 5.  $K$ : The adjusting coefficient for the curve  $t^{-\theta}$  to fit the time series of the relative increment after the first period of the spread.

The estimation of the proposed time series for the US Data of COVID-19 would be discussed in the next subsection.

### 2.3 Estimation for Increment of Confirmed Cases of COVID-19 in the US

In this section, we discuss on how the time series explained above will fit to the COVID-19 data from the US. To estimate the parameters of the model, the following practices must be taken:

- take  $b$  as the first point that the geometric mean of the relative increments in the previous points exceeds  $3/2$  times the geometric mean of the next three points.

$$\hat{b} = \min \{n | \text{Geo. mean} (X_{n+1}, X_{n+2}, X_{n+3}) < \frac{2}{3} \text{Geo. mean} (X_1, X_2, \dots, X_n)\}$$

which can be computed as

$$\hat{b} = \min \{n | \sqrt[3]{(1 + X_{n+1})(1 + X_{n+2})(X_{n+3})} - 1 < \frac{2}{3} \sqrt[n]{(1 + X_1)(1 + X_2) \dots (X_n)} - 1\}$$

**Graphically, this time can be identified as the time when the plot of relative increments falls irreversibly, see Day 0:**

27 February 2020, Day 45: 12 April 2020.

Figure 1.

225

Day 0: 27 February 2020, Day 45: 12 April 2020.

230

Figure 1 presents the relative increment in time series of the COVID-19 confirmed cases for the US. The figure demonstrates that the time series has a time point in which the first stationary period ends, and the second decreasing period starts. In this case, it is resulted that  $\hat{b} = 24$ . This confirms our first assumption for the proposed time series, See Section 1.

<Insert Figure 1, here>

235

- calculate the geometric mean of the ratio cumulative numbers in the previous points  $(1 + X_t)$  from  $t = 1$  to  $t = b$  as the estimation of the parameter  $IR$  as  $\widehat{IR} = \sqrt[b]{(1 + X_1)(1 + X_2) \dots (X_b)} - 1$ . For the US COVID-19 data,  $\widehat{IR} = 0.3755$ .

240

- estimate the parameters  $\theta$  and  $K$  due to the following linear relation

$$X_t \cong K/t^\theta \Rightarrow 1/X_t \cong t^\theta/K \Rightarrow \ln(1/X_t) \cong \theta \ln(t) - \ln(K) \Rightarrow$$

$$\ln(X_t) \cong -\theta \ln(t) + \ln(K) \Rightarrow \ln(X_t) \cong \theta \ln\left(\frac{1}{t}\right) + \ln(K).$$

For the US COVID-19 data, the calculations lead in

245  $\hat{\theta} = 2.6609$  and  $\hat{K} = 1539.972$ , See Figure 2.

<Insert Figure 2, here>

- Multiply all the observations after  $t = b$  by  $\frac{t^\theta}{K}$  to have an identical mean and variance for all the newly obtained data ( $W_t$ )

$$X_t \sim \text{Normal}\left(K/t^\theta, K^2/at^{2\theta}\right) \Rightarrow W_t = \frac{t^\theta}{K} X_t \sim \text{Normal}(1, 1/a)$$

250 Therefore, the variance of the newly recorded data is a good candidate for estimating  $1/a$ . Accordingly,

$$\hat{a} = \frac{1}{s_w^2} = 0.0532.$$

**All in all, the US COVID-19 data from 27 February to 11 April 2020**

**generates the estimation of**

255  $(\hat{b}, \widehat{IR}, \hat{K}, \hat{\theta}, \hat{a}) = (24, 0.3755, 1539.972, 2.6609, 0.0532)$  to simulate

**and predict the US confirmed cases using the time series introduced**

**in Section 2.2. Day 0: 22 March 2020, Day 20: 12 April 2020.**

Figure 3 demonstrates that our model has an appropriate fitting power to the relative increments of the of the confirmed COVID-19 cases in the US. Based on the model, the relative increment is decreasing from around 7% to about 1% (its 80% confidence interval is 0.88% - 1.56%).

<Insert Figure 3, here>

#### 2.4 *Estimation for Mortality Rate of COVID-19 in the US*

We utilize the mortality rate of 5% to estimate the number of deaths in the US from 28 February to 11 April.

### 3. **Forecasting Results**

In this section, we first forecast the relative increments of confirmed COVID-19 cases in the US from 12 April to 21 May, 2020. Accordingly, we will forecast the accumulative number of conformed cases in the next 40 days in the US. Second, we forecast the COVID-19 mortality rate and accordingly the accumulative number of Mortalities from 12 April to 21 May, 2020. In our forecast, as mentioned earlier,

we use five different forecasting output variables, namely 80% lower  
275 bound, 80% upper bound, average and two possible realizations.

### 3.1 *Forecasting relative increment and accumulative number of confirmed cases*

Day 0: 11 April 2020, Day 40: 21 May 2020.

Figure presents our five forecasting variables for the relative  
280 increment of confirmed cases of COVID-19 in the US from 12 April to 21 May, 2020, presented on the model as day 0 to day 40.

<Insert Figure 4, here>

Figure presents the forecast of the newly confirmed cases of  
COVID-19 in the US from 12 April to 21 May, 2020. As demonstrated  
285 by Figure , based on the decreasing trend of daily relative increments, despite the increasing cumulative number of confirmed cases, the new confirmed cases infected by the pandemic follows a decreasing pattern. In the first 10 days of April, the variable is fluctuate around 30000 while our prediction says that this variable decreases to [12801 22578]  
290 with the probability of 80% (the point prediction is equal to 17551) on 21 May 2020.

<Insert Figure 5, here>

Figure presents the forecast of cumulative confirmed cases for the studied period. The forecast predicts that over the period of 12 April to 21 May, 2020, the cumulative number of cases rises from 502876 on 11 April to 1464729 cases on 21 May 2020, with 80% confidence interval equal to [1375362 1540424].

<Insert Figure 6, here>

### *3.1 Forecasting accumulative number mortalities*

According to Figure , the number of Mortalities caused by COVID-19 in the US increases from 18747 cases on 11 April to around 73000 cases on 21 May, with 80% confidence interval equal to [69 77] K. It means we will encounter with about 1350 daily Mortalities on average over the studied period. The slope of the increasing graph for the cumulative number of Mortalities is falling as well as the graph of the cumulative number of confirmed cases.

<Insert Figure 7, here>

#### 4. Conclusions, Discussion and Limitations

The relative increments in confirmed cases of COVID-19 in the US  
310 has shown in the beginning an increase in the rate of its growth,  
then it has experienced a sudden fall and continued to decrease in a  
relaxer rate. This demonstrates a great variety and presents  
different stages of COVID-19 in the US. This paper aims to forecast  
the number of confirmed as well as the Mortalities of COVID-19 in  
315 the US from 12 April to 21 May, 2020. Given the shortfall of C  
algorithms to forecast in longer terms and limited flexibility of  
ARIMA to model different stages of time series in an outbreak like  
COVID-19, we suggested the use of relative increment measure for  
the confirmed cases and the mortality rate for the number of  
320 Mortalities caused by COVID-19. The model presented in Section 2  
is able to model COVID-19 data in the US because the proposed  
structure takes variety of different stages in the data.

Applying the model to the US COVID-19 data from (WHO, 2020c),  
the paper predicts 1464729 confirmed cases on the 21<sup>st</sup> of May,  
325 2020. The forecast illustrates that the US can be considered the first  
country to face the problem of the explosion of the pandemic for

COVID-19. Although the accumulative number of confirmed cases in the US will be increasing to the 21<sup>st</sup> of May, the forecasting exercise presented in this paper found some evidence that the relative increment of the cases is falling. This shows the growth rate of confirmed cases of COVID-19 in the US will be dropping. The estimated drop in the relative increments is driven by the changes in behaviours of population or specific social distancing measures that the US has implemented.

**Our results highlight the value of reproductive strategies in time series modelling of COVID-19. For instance, the rapid growth of COVID-19 confirmed cases in the US at the beginning of the outbreak would result into inaccurate forecasts, while Day 0: 27 February 2020, Day 45: 12 April 2020.**

Figure 1 demonstrates a sudden fall followed by a relaxer decreasing rate of growth. This problem has been also reported by Kucharski et al. (2020). Our model benefits from a reproductive strategy from a time point in which the US COVID-19 data demonstrates a sudden fall.

The model presented in this paper is limited to three assumptions on the relative increments. Firstly, similar to Italian COVID-19 it must rapidly grow with a dramatic increasing rate, experiences a sudden fall and then decreases in a smaller rate. 350 Second, it must follow a normal distribution and third, over time, the ratio of variance to mean remains constant. Although holding these assumptions is an advantage of the proposed method to model the US COVID-19, it may limit the generalizability of the model fitting to other outbreaks.

355 Researchers in this area are recommended to apply the method to COVID-19 data from other regions. Our initial analysis shows the model fits to Iranian, UK and Italian data. From operational perspective, in order to simplify the proposed model for quicker implementation of forecasts, the model can be simplified by 360 removing the two first parameters suitable to describe the gradual decrease of the relative increments ( $t=b, b+1, \dots$ ).

## References

- Akaike, H., 1969. Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics* 21, 243–247.
- 365 Baltagi, B.H., Baltagi, B.H., 2001. *A companion to theoretical econometrics*. Wiley Online Library.
- Berger, D.W., Herkenhoff, K.F., Mongey, S., 2020. An seir infectious disease model with testing and conditional quarantine. *National Bureau of Economic Research*.
- 370 Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I.C.-H., Hickmann, K.S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., Velardi, P., Vespignani, A., Finelli, L., Influenza Forecasting Contest Working Group, 2016. Results from the centers for disease control and prevention’s predict the 2013-2014 Influenza Season Challenge. *BMC Infect. Dis.* 16, 357.
- 375 <https://doi.org/10.1186/s12879-016-1669-x>
- CDC, 2020. Coronavirus Disease 2019 (COVID-19) [WWW Document]. Centers for Disease Control and Prevention. URL <https://www.cdc.gov/coronavirus/2019-ncov/index.html> (accessed 4.15.20).
- Chowell, G., Luo, R., Sun, K., Roosa, K., Tariq, A., Viboud, C., 2020. Real-time forecasting of epidemic trajectories using computational dynamic ensembles. *Epidemics* 30, 100379.
- 380 Chretien, J.-P., George, D., Shaman, J., Chitale, R.A., McKenzie, F.E., 2014. Influenza Forecasting in Human Populations: A Scoping Review. *PLOS ONE* 9, e94130. <https://doi.org/10.1371/journal.pone.0094130>

- 385 Haining, R.P., 1978. The moving average model for spatial interaction. *Transactions of the Institute of British Geographers* 202–225.
- Kass-Hout, T.A., Xu, Z., McMurray, P., Park, S., Buckeridge, D.L., Brownstein, J.S., Finelli, L., Groseclose, S.L., 2012. Application of change point analysis to daily influenza-like illness emergency department visits. *J Am Med Inform Assoc* 19, 1075–1081. <https://doi.org/10.1136/amiajnl-2011-000793>
- 390 Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., Munday, J.D., 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*.
- 395 Kumar, N., Jha, G.K., 2013. A time series approach for weather forecasting. *Int J Control Theory Comput Model (IJCTCM)* 3, 19–25.
- Murphy, S.P., Burkom, H., 2008. Recombinant Temporal Aberration Detection Algorithms for Enhanced Biosurveillance. *J Am Med Inform Assoc* 15, 77–86. <https://doi.org/10.1197/jamia.M2587>
- 400 NOAA, 2016. Forecasting the Next Dengue Outbreak | National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC) [WWW Document]. URL <https://www.ncdc.noaa.gov/news/forecasting-next-dengue-outbreak> (accessed 4.9.20).
- 405 Paul, M.J., Dredze, M., Broniatowski, D., 2014. Twitter Improves Influenza Forecasting. *PLoS Curr.* <https://doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9>

- Radzuan, N.F.M., Othman, Z., Bakar, A.A., 2013. Uncertain Time Series in Weather  
410 Prediction. *Procedia Technology*, 4th International Conference on Electrical  
Engineering and Informatics, ICEEI 2013 11, 557–564.  
<https://doi.org/10.1016/j.protcy.2013.12.228>
- Read, J.M., Bridgen, J.R., Cummings, D.A., Ho, A., Jewell, C.P., 2020. Novel  
coronavirus 2019-nCoV: early estimation of epidemiological parameters and  
415 epidemic predictions. *MedRxiv*.
- Reis, B.Y., Mandl, K.D., 2003. Time series modeling for syndromic surveillance.  
*BMC Medical Informatics and Decision Making* 3, 2.
- Rios, M., Garcia, J.M., Sanchez, J.A., Perez, D., 2000. A statistical analysis of the  
seasonality in pulmonary tuberculosis. *European Journal of Epidemiology*  
420 16, 483–488.
- Rojas, I., Valenzuela, O., Rojas, F., Guillén, A., Herrera, L.J., Pomares, H., Marquez,  
L., Pasadas, M., 2008. Soft-computing techniques and ARMA model for  
time series prediction. *Neurocomputing* 71, 519–537.
- Rounds, J., Charles-Smith, L., Corley, C.D., 2017. Soda Pop: A Time-Series  
425 Clustering, Alarming and Disease Forecasting Application. *Online journal of  
public health informatics* 9.
- Scott, S.L., Varian, H.R., 2015. Bayesian Variable Selection for Nowcasting  
Economic Time Series, in: *NBER Chapters*. National Bureau of Economic  
Research, Inc, pp. 119–135.
- 430 Talaei-Khoei, A., Wilson, J.M., Kazemi, S.-F., 2019. Period of Measurement in Time-  
Series Predictions of Disease Counts from 2007 to 2017 in Northern Nevada:

- Analytics Experiment. *JMIR Public Health Surveill* 5, e11357.  
<https://doi.org/10.2196/11357>
- 435 Tokars, J.I., Burkom, H., Xing, J., English, R., Bloom, S., Cox, K., Pavlin, J.A., 2015.  
Enhancing Time-Series Detection Algorithms for Automated  
Biosurveillance. *Emerging infectious diseases - CDC* 15.  
<https://doi.org/10.3201/eid1504.080616>
- Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q.,  
440 Chowell, G., Simonsen, L., Vespignani, A., 2018. The RAPIDD ebola  
forecasting challenge: Synthesis and lessons learnt. *Epidemics* 22, 13–21.
- Voyant, C., Paoli, C., Muselli, M., Nivet, M.-L., 2013. Multi-horizon solar radiation  
forecasting for Mediterranean locations using time series models. *Renewable  
and Sustainable Energy Reviews* 28, 44–52.  
<https://doi.org/10.1016/j.rser.2013.07.058>
- 445 WHO, 2020a. Coronavirus Disease (COVID-19) - events as they happen [WWW  
Document]. URL [https://www.who.int/emergencies/diseases/novel-  
coronavirus-2019/events-as-they-happen](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen) (accessed 4.6.20).
- WHO, 2020b. Coronavirus disease 2019 (COVID-19) Situation Report – 79  
(Situation Report No. 79).
- 450 WHO, 2020c. Novel Coronavirus (COVID-19) situation reports [WWW Document].  
URL [https://www.who.int/emergencies/diseases/novel-coronavirus-  
2019/situation-reports](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports) (accessed 4.6.20).
- Wongkoon, S., Jaroensutasinee, M., Jaroensutasinee, K., 2012. Development of  
temporal modeling for prediction of dengue infection in Northeastern  
455 Thailand. *Asian Pacific journal of tropical medicine* 5, 249–252.

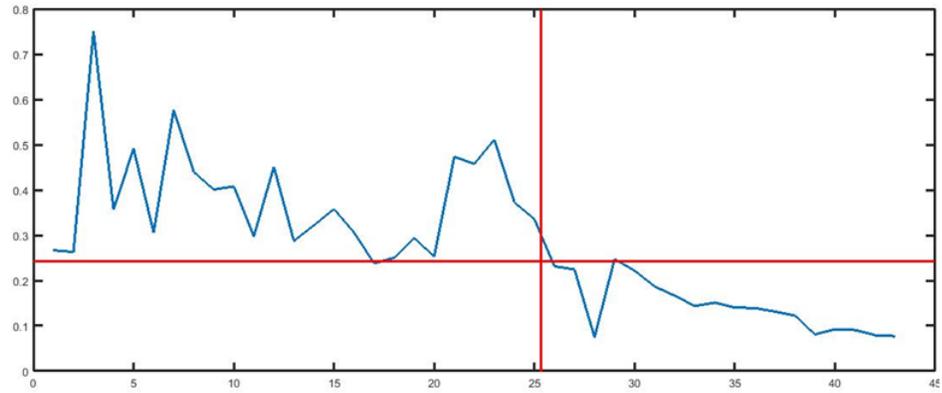
Zeger, S.L., Irizarry, R., Peng, R.D., 2006. On time series analysis of public health and biomedical data. *Annu. Rev. Public Health* 27, 57–79.

Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159–175.

460 Zhang, X., Zhang, T., Young, A.A., Li, X., 2014. Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. *PLOS ONE* 9, e88075. <https://doi.org/10.1371/journal.pone.0088075>

465 **Summary Points**

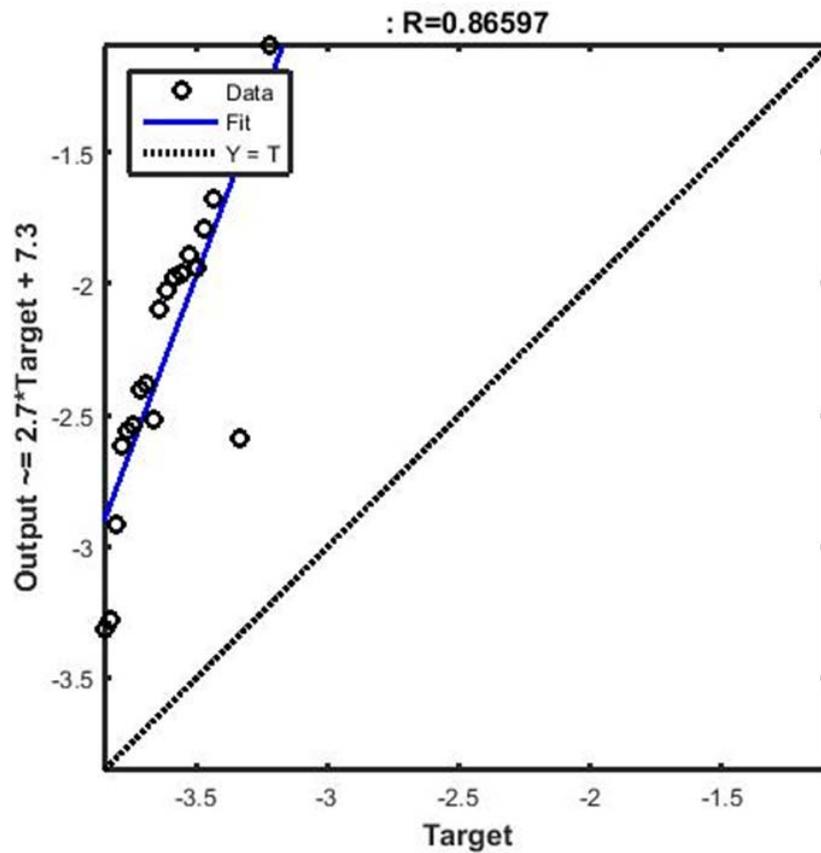
1. The study aims to forecast the number of confirmed cases of COVID-19 in the US.
2. The paper builds a time series model based on COVID-19 data from 27 Feb to 11 April 2020.
- 470 3. The study forecasts 1,464,729 confirmed cases of COVID-19 by the 21st of May, 2020.
4. The paper forecasts the accumulative number of mortalities of COVID-19 to around 47000 on the 21st of May, 2020.
5. The proposed method is based on relative increments of confirmed cases.
- 475



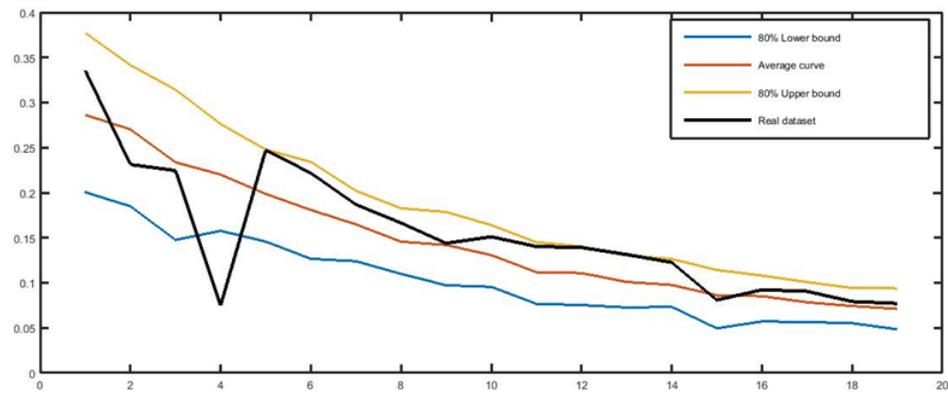
Day 0: 27 February 2020, Day 45: 12 April 2020.

480

**Figure 1** The time series of relative increments and the time of passing from the first stationary period to the new decreasing period.



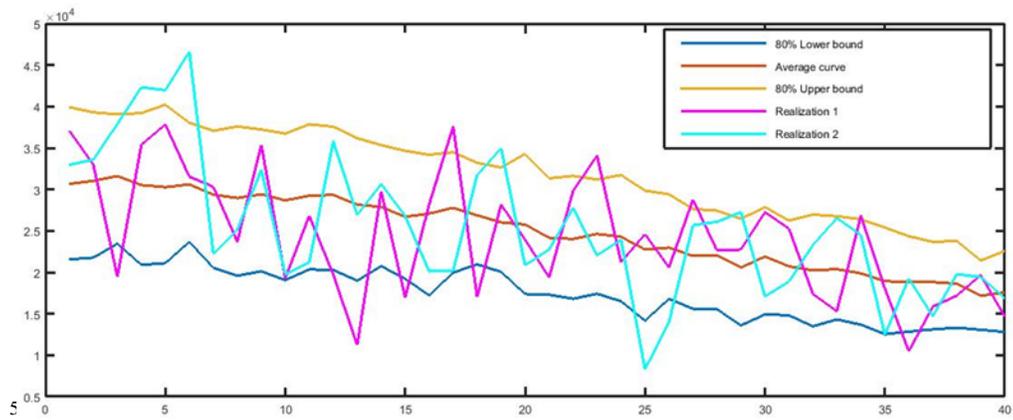
485 **Figure 2 Estimating the linear regression for fitting the association between relative increment and the inverse of time to obtain estimations of the parameters  $\theta$  and  $K$ .**



Day 0: 22 March 2020, Day 20: 12 April 2020.

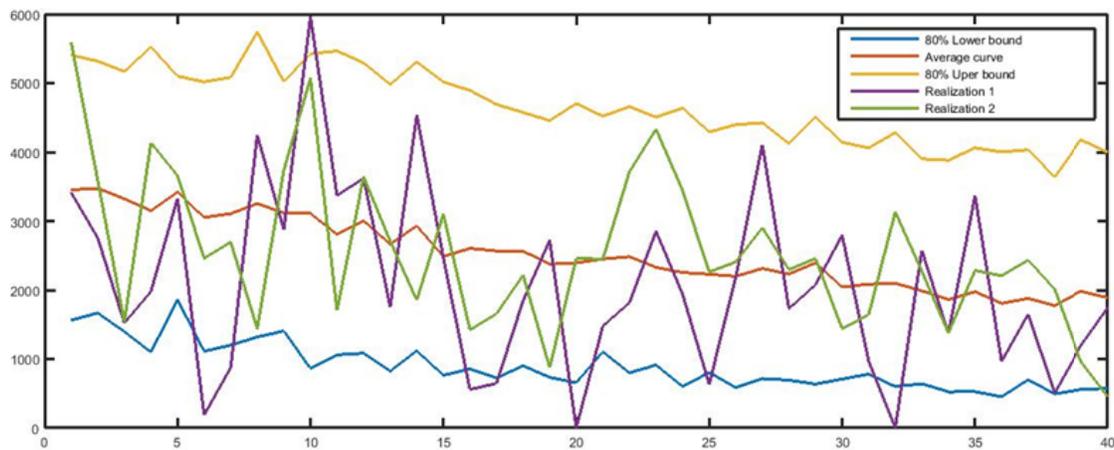
490 **Figure 3 Fitting Power of the model to the relative increments of**  
**confirmed COVID-19 cases in the US**

495



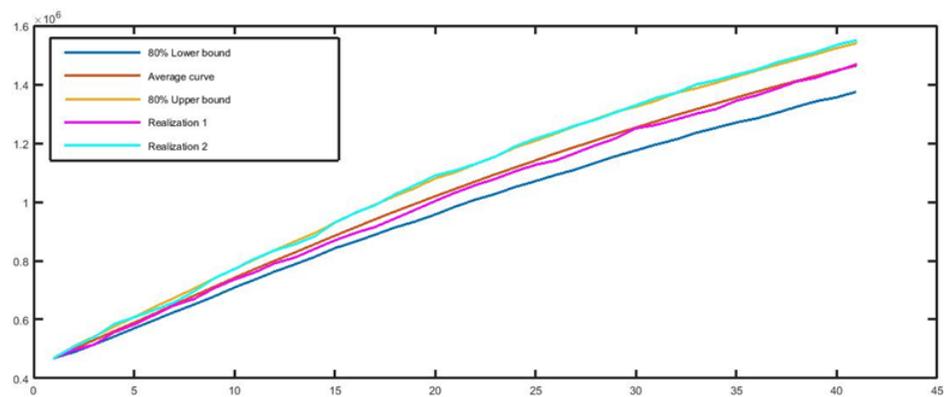
Day 0: 11 April 2020, Day 40: 21 May 2020.

**Figure 4 Forecast of the relative increments of the COVID-19 cases in the US, from 12 April to 21 May, 2020.**



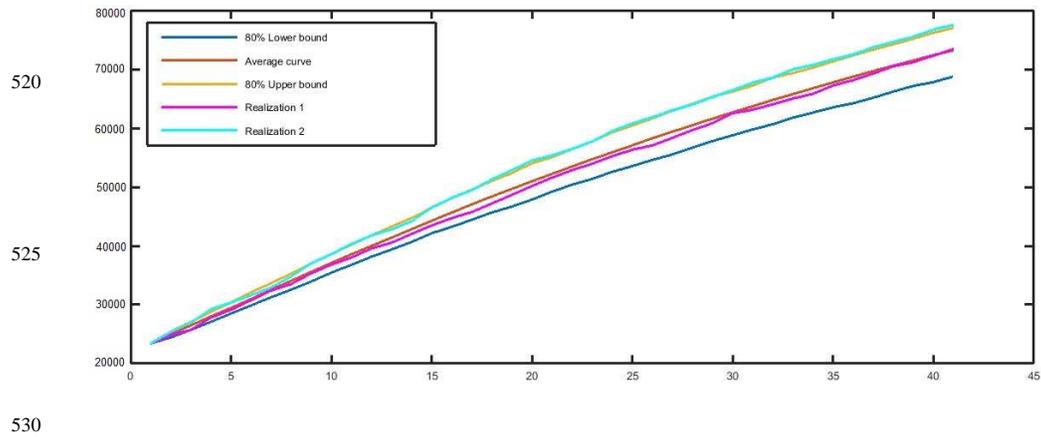
Day 0: 11 April 2020, Day 40: 21 May 2020.

**Figure 5** Forecast of the newly confirmed cases of COVID-19 in the US, from 12 April to 21 May, 2020.



Day 0: 11 April 2020, Day 40: 21 May 2020.

515 **Figure 6 Forecast of the cumulative number of confirmed cases of COVID-19 in the US, from 12 April to 21 May, 2020.**



Day 0: 11 April 2020, Day 40: 21 May 2020.

**Figure 7 Forecast of the cumulative number of mortalities caused by COVID-19 in the US, from 12 April to 21 May, 2020.**