Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(4):376-391

Original Article

A New Family of Time Series to Model the Decreasing Relative Increment of Spreading of an Outbreak

Babak Jamshidi[1], Hakim Bekrizadeh[2], Shahriar Jamshidi Zargaran[3], Mansour Rezaei[4]

[1]Medical Statistician, KiTEC, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK.
[2]Department of Statistics, Payam-e-Noor University, Iran.
[3]Department of Neuroimaging, Isfahan University of Medical Sciences, Isfahan, Iran.
[4]Department of Biostatistics, Kermanshah University of Medical Sciences, Kermanshah, Iran.

ARTICLE INFO

Received 23.06.2022
Revised 11.07.2022
Accepted 12.10.2022
Published 15.12.2022

Key words:
Relative increment;
Epidemic;
COVID-19;
Model;
Time series;
Spreading;

ABSTRACT

Introduction: There are different mathematical models describing the propagation of an epidemic. These models can be divided into phenomenological, compartmental, deep learning, and individual-based methods. From other viewpoints, we can classify them into macroscopic or microscopic, stochastic or deterministic, homogeneous or heterogeneous, univariate or multivariate, parsimonious or complex, or forecasting or mechanistic.

This paper defines a novel univariate bi-partite time series model able to describe spreading a communicable infection in a population in terms of the relative increment of the cumulative number of confirmed cases. The introduced model can describe different stages of the first wave of the outbreak of a communicable disease from the start to the end.

Methods: The outcome of the model is relative increment, and it has five positive parameters: the length of the first days of spreading and the relative increment in these days, the potent of the mildly decreasing trend (after the significant decrease), and the adjusting coefficient to adapt this trend to the initial pattern, and the fixed ratio of the mean to the variance.

Results: We use it to describe the propagation of various disease outbreaks, including the SARS (2003), the MERS (2018), the Ebola (2014-2016), the HIV/AIDS (1990-2018), the Cholera (2008-2009), and the COVID-19 epidemic in Iran, Italy, the UK, the USA, China and four of its provinces; Beijing, Guangdong, Shanghai, and Hubei (2020). In all mentioned cases, the model has an acceptable performance. In addition, we compare the goodness of this model with the ARIMA models by fitting the propagation of COVID-19 in Iran, Italy, the UK, and the USA.

Conclusion: The introduced model is flexible enough to describe a broad range of epidemics. In comparison with ARIMA time series models, our model is more initiative and less complicated, it has fewer parameters, the estimation of its parameters is more straightforward, and its forecasts are narrower and more accurate. Due to its simplicity and accuracy, this model is a good tool for epidemiologists and biostatisticians to model the first wave of an epidemic.

*.Corresponding Author: babak.j6668@gmail.com

Introduction

Communicable diseases play a significant role in human life. They affect millions of people yearly causing a variable number of health problems including death. Even within the twenty-first century, we see that millions of people die of infectious diseases like measles and respiratory infections, and right now, the world is faced with a crisis; the spread of COVID-19. By the end of November 2022, this communicable disease has infected over 637 M confirmed cases and killed over 6.6 M worldwide.[1] Nowadays, new infectious diseases are spreading around the world faster than ever. This unprecedented rate is the result of factors such as the increasing ease of international travel, population growth, resistance to drugs, and degradation of the environment.[2] This high rate highlights the importance of research on the spreading of communicable diseases.

The essence of these diseases is their transmission that is the main subject of the studies addressing this issue in the various sciences. An approach to query this subject is the mathematical one.[3] Generally, Mathematical modeling is the transformation from a real-world situation to a mathematical problem achieved using a mathematical language, which, is an idealized and simplified representation of the basic characteristics of the real situation by using a suitable set of mathematical symbols, concepts, relations, functions, and structures. Specifically, mathematical modeling in epidemiology provides an understanding of the underlying mechanisms that influence the spread of disease, and in the process, it suggests control strategies.[4]

The history of the modeling spread of a communicable disease dates back to the beginning of the 20th century. In 1906, Hamer argued that the spread of infection should depend on both the number of susceptible and infective individuals.[5] Thereafter, some approaches were introduced to model the spread of the infection among them.[6-10] Generally, mathematical models describing the propagation of an epidemic can be divided into phenomenological, compartmental, deep learning, and individual-based methods. From other viewpoints, we can classify these models into macroscopic or microscopic, stochastic or deterministic, homogeneous or heterogeneous, univariate or multivariate, parsimonious or complex, or forecasting or mechanistic. However, these models include agent-based models[11], random and non-random network models (for example, branching processes)[12], stochastic and deterministic differential equation models (including compartmental models)[13], multi-state models[14], time series models (especially, ARIMA models)[15-18], artificial intelligence models (notably, neural network models[19], growth function models (exponential, sub-exponential, or logistic growth functions)[20], and combinations of them.[21] Principally, the parsimonious models are suitable for understanding the short-term behavior of the pandemic, while complex models -requiring more data and estimation- are much more accurate and able to cover the different scenarios.

Most of the parsimonious univariate models (growth function, time series, and some differential equation models) concern the beginning of the outbreak among them stochastic branching models and exponential growth rate models, which based on them, we can calculate the critics like the reproductive number (R) and the exponential rate of growth

(λ) to represent the intensity of the infections. The absence of a simple model to describe the propagation of the disease after passing the first stage of the spread motivated us to write this paper.

It is worth saying that although the time series models have long been of interest in the literature, to the best of our knowledge, these applied time series models are restricted to the ARIMA family. The present paper attempts to address this shortage as well.

Methods

Statistical models can be divided into two groups: static (stationary) and dynamic (non-stationary). Static models do not change over time, then they are fixed and independent of time. Random variables are the most frequently used static statistical models. A stochastic process is a set of random variables defined on the basis of time. It is a time-dependent random phenomenon; therefore, it is classified as a dynamic model. The set of time points that the stochastic processes are defined on can be discrete or continuous. Similarly, the values that the random variables of a stochastic process take can be discrete or continuous. A time series is a sample path of a discrete-time stochastic process.[22]

We present a novel model to reproduce the time series of the spreading of an epidemic in terms of relative increments of the number of infected individuals. We adopt a probabilistic approach because firstly, randomness is a major player, especially at the beginning of an outbreak. Also, there are several sources of uncertainty regarding predicting the course of a disease.[23] Finally, the stochastic approach promotes the flexibility of the model. The stochastic approach to modelling communicable diseases mainly appears in the formats of random graphs, agent-based models, and stochastic differential equations. The time series models are not so commonly used in the modeling of spreading.

Construction of the model

In order to describe a mathematical model for the spread of a communicable disease, it is necessary to make some assumptions based on reality or evidence. Here, the model is based on the following assumptions:

A1. The trend of the time series of the relative increment is decreasing. It faces a significant sudden decrease, thereafter it falls gradually.[24]

A 2. In each time point, the distribution of the relative increment is normal[(1)].

A 3. Over time, the ratio of variance to mean remains fixed[(2)].

To fulfil the first assumption, we restrict ourselves to the first wave of an epidemic.

Let Y_t is a time series of the cumulative number of confirmed cases at time *t*. We want to study the relative increment ($X_t = \frac{Y_{t+1}}{Y_t} - 1$). It is well known that during a wave of an epidemic, like reproductive number and exponential growth rate, the relative increment decreases as time passes.[24] The relative decrement has studied by[25] but under a different name: the rate of growth.

The model has five positive parameters (b, IR,

[(1)]Normality is the frequently made assumption. This distribution is absolutely applicable to normal and natural characteristics. Since the relative increment is a positive continuous variable, we need to modify simulated normal increments to describe the intended variable.

[(2)]This simplifier assumption helps us to define our model with less parameters.

K, θ, a);

b: the length of the first days of spreading (before the significant decrease),

IR: the relative increment in the first days of spreading,

θ: the potent of smoothly decreasing trend of the relative increments after the significant decrease (the acceleration of falling during the period of gradual decreasing) ($X_t \propto t^{-\theta}$),

a: the fixed ratio of the mean to the variance[1], and

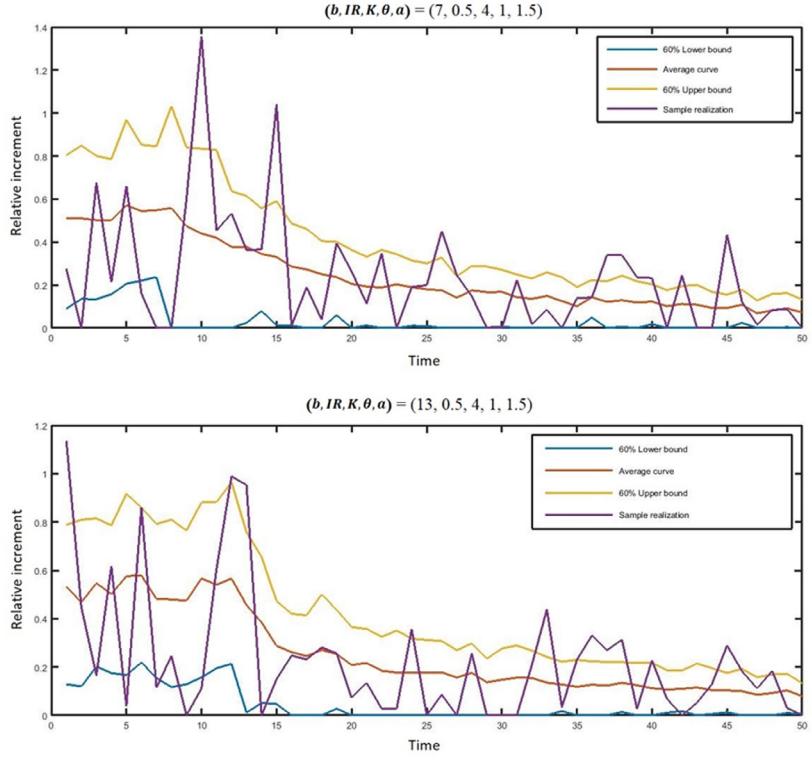
K: the adjusting coefficient for the curve $t^{-\theta}$ to adapt the two parts of the model.

Therefore, the first part of the model (X_t~*Normal(IR,* IR^2/a*), t=1,...,b-1*) is stationary and the second part (X_t~*Normal* (K/t^{θ}, $K^2/at^{2\theta}$) , *t=b, b+1, ...*) is non-stationary.

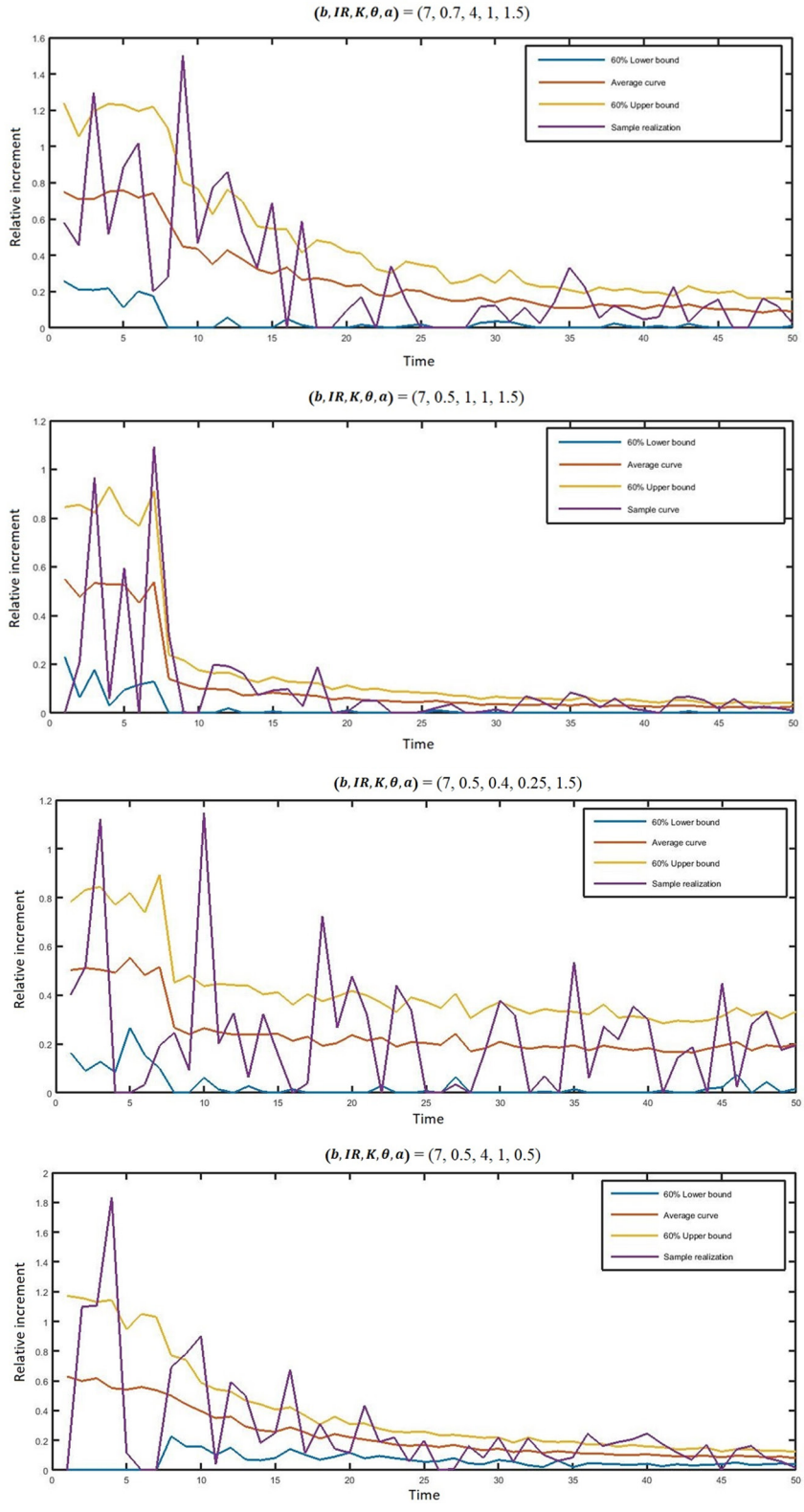
It is noticeable that by omitting the two first parameters, we obtain a simplified model suitable to describe the gradual decrease of the relative increments (*t=b,b+1,...*). It is possible to generalize the model if a in the first part differs from a in the second part.

Due to the parameters used to define the present model and its structure, it is flexible enough to cover the different stages of the first wave of a communicable disease. This flexibility is well illustrated by Plot 1, wherein the effect of changing each parameter of the model on its pattern is represented.

As it is clear from Plot 1, the formulation of our model indicates that increasing *b* (graph 2 vs graph 1) leads to increasing the probability of acute growth and, therefore more expected cumulative number of confirmed cases. Since *IR* determines the rate of growth in the first days, the more parameter *IR* is, the more intense the outbreak of the disease happens in the first days



[1]Notice that a is different from CV (coefficient of variation); $a = \frac{Expectation}{(Standard\ deviation)^2}$ while $\frac{1}{CV} = \frac{Expectation}{Standard\ deviation}$.



Plot 1. A sample realization, the average curve, and the 60% upper and lower bounds of 100 simulations of the model with different parameters for 50 steps

Table 1. The estimated parameters for the model to fit datasets regarding some epidemics

Epidemic	b	IR	K	θ	a
Cholera (Zimbabwe)*	10	0.5683	0.4961	0.8073	0.5664
Ebola (Worldwide)	3	0.7470	5.5946	1.4613	7.0739
SARS (China)	0	-	2.0561	0.9280	15.6498
MERS (South Korea)	9	0.3298	115.5675	2.4711	1.0624
HIV/AIDS (Worldwide)*	6	0.2053	0.8477	0.9976	0.0048
COVID-19 (Beijing)	9	0.4248	167.7165	2.8943	1.1424
COVID-19 (Guangdong)	11	0.3221	286.7553	3.4334	1.1209
COVID-19 (Shanghai)	7	0.8783	105.5936	2.8218	0.7418
COVID-19 (Hubei)	9	0.2760	10.0758	2.1307	22.7603
COVID-19 (China)	7	0.4661	754.1667	3.6763	16.5409
COVID-19 (Iran)	13	0.7217	45.2078	1.9042	0.0929
COVID-19 (UK)**	13	0.3272	88.2050	1.9283	0.3382
COVID-19 (UK)**	13	0.3272	5514.7	2.9317	0.0630
COVID-19 (Italy)	8	0.4753	48.0500	1.9672	0.2125
COVID-19 (USA)	24	0.3755	1539.972	2.6609	0.0532

*Cholera (Zimbabwe) and HIV/AIDS (Worldwide) data were collected weekly and annual, respectively, while other data were gathered daily.
**The first row of COVID-19 (UK) was obtained to predict 15 Apr – 30 May 2020, while the second row was calculated to predict the interval 31 May – 1 July 2020.

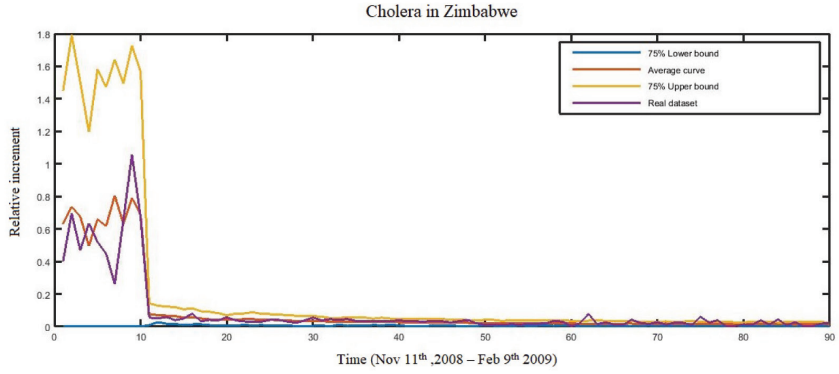
and thereafter (graph 3 vs graph 1). Although it is not so interpretable to manipulate K due to its adjusting role, by decreasing K (graph 4 vs graph 1), we have a lower range of relative. Since θ represents the acceleration of falling during the gradual decrease, by declining θ (graph 5 vs graph 1), the graph falls later and we have a shallower trend. Finally, the change of the parameter a (graph 6 vs graph 1), as it is somehow similar to the inverse of the coefficient variation causes an opposite effect on the range of oscillations.

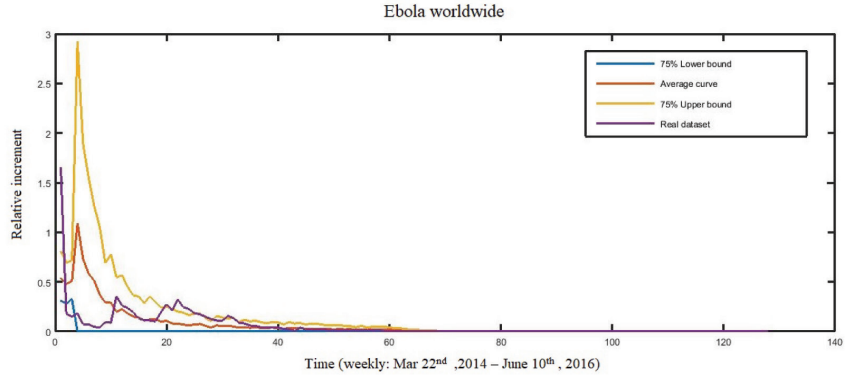
Generally, formulating a model is accompanied by a trade-off between three important and often conflicting elements: accuracy, transparency, and flexibility. These three elements refer to the ability to reproduce the observed data and predict future dynamics reliably, being understandable analytically (how the components of the model influence the dynamics and interact), and adaptability to new or different situations, respectively.[26] Undoubtedly, no model is perfect generally, particularly in epidemiology. However, our model is simultaneously accurate, transparent, and sufficiently flexible. Our model is competent enough to describe a wide range of recent outbreaks (See Table 1 and Plot 2).

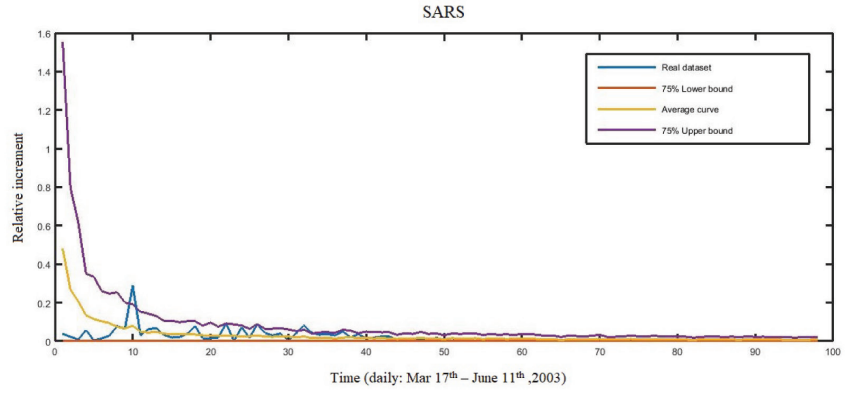
From the complexity standpoint, the models can be classified into two groups;

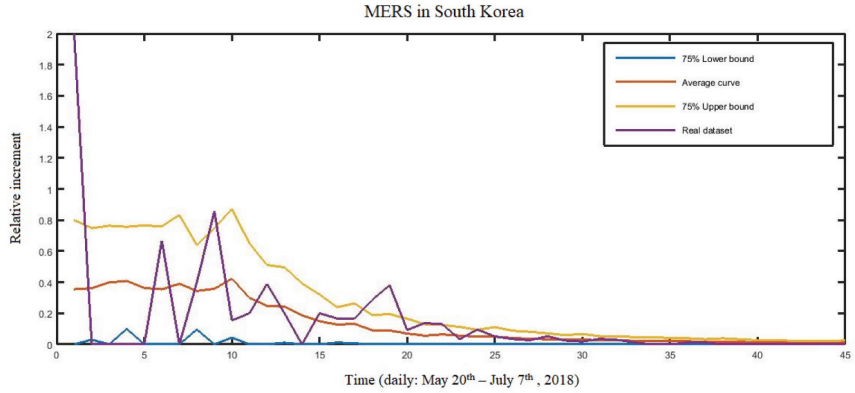
- simple or strategic models which omit most details and are designed only to highlight general qualitative behavior, and
- detailed or tactical models, which are usually designed for specific situations, including short-term quantitative predictions.

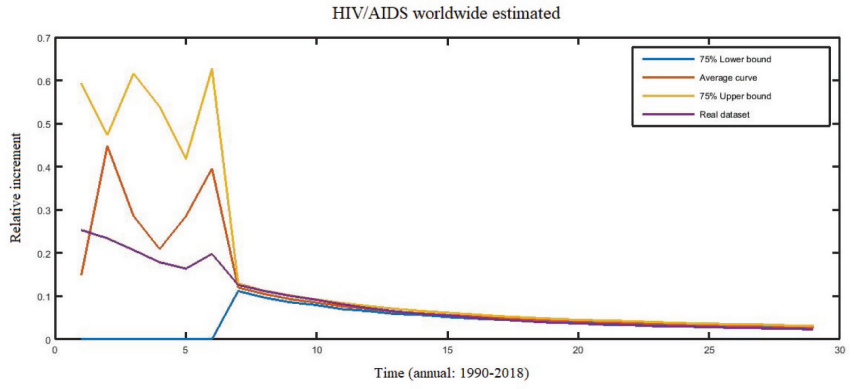
Detailed models are generally difficult or impossible to solve analytically and hence, their usefulness for theoretical purposes is limited, although their strategic value may be high.[26]

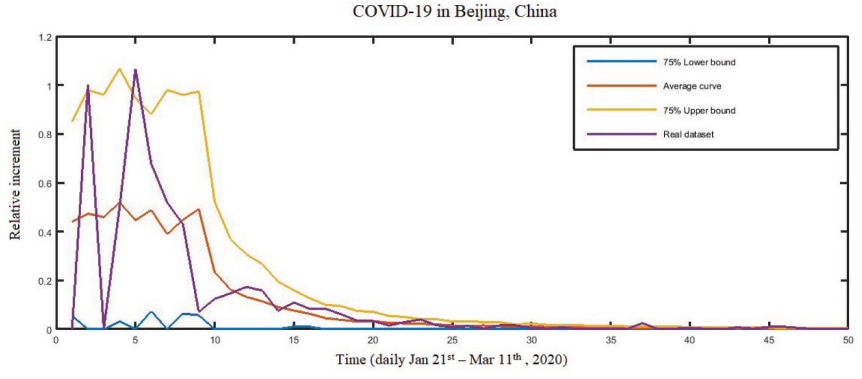

Cholera in Zimbabwe
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (Nov 11th ,2008 – Feb 9th 2009)

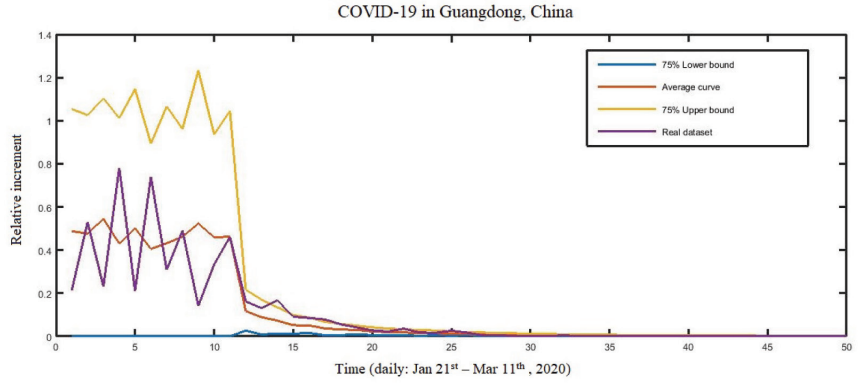

Ebola worldwide
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (weekly: Mar 22nd ,2014 – June 10th , 2016)

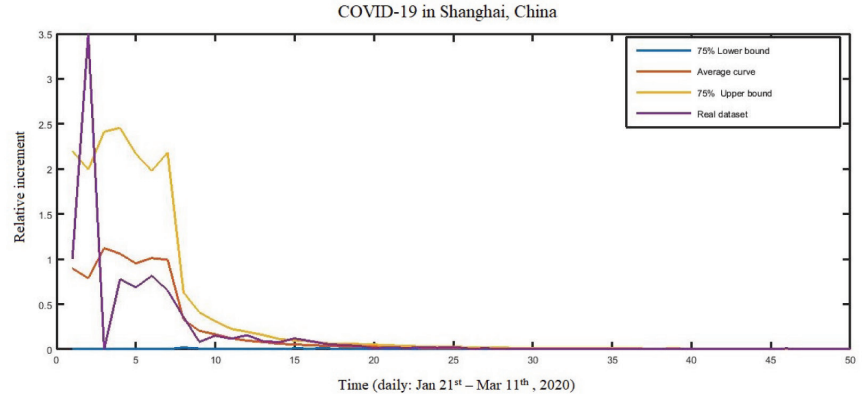

SARS
Real dataset
75% Lower bound
Average curve
75% Upper bound
Relative increment
Time (daily: Mar 17th – June 11th ,2003)

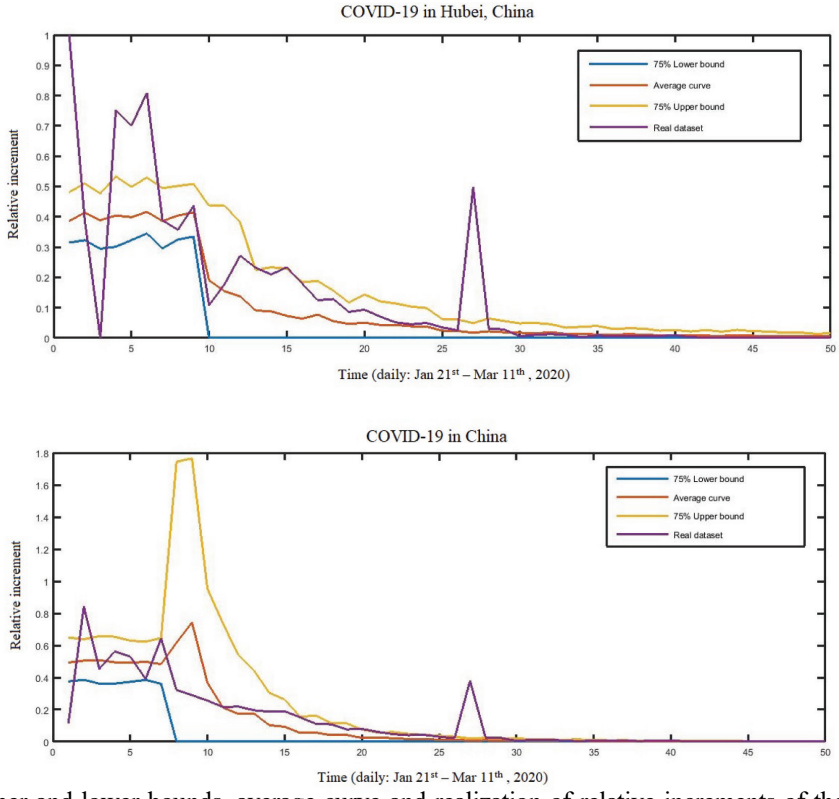

MERS in South Korea
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (daily: May 20th – July 7th , 2018)


HIV/AIDS worldwide estimated
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (annual: 1990-2018)


COVID-19 in Beijing, China
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (daily Jan 21st – Mar 11th , 2020)


COVID-19 in Guangdong, China
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (daily: Jan 21st – Mar 11th , 2020)


COVID-19 in Shanghai, China
75% Lower bound
Average curve
75% Upper bound
Real dataset
Relative increment
Time (daily: Jan 21st – Mar 11th , 2020)



Plot 2. The 75% upper and lower bounds, average curve and realization of relative increments of the outbreaks regarding Cholera, Ebola, SARS, MERS, AIDS, and COVID-19

We can classify our model as a simple model due to ignoring the steps of the disease (as the compartmental models) and lack of focus on the transmission dynamics from individual to individual in a population (as the random graphs), and as a detailed model, it deals with the procedure of the spreading step by step.

Estimation of the parameters

The procedure of calculating estimations of the parameters is straightforward. To estimate the parameters of the model, we can

1. 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 or the time point that the relative increments fall irreversibly.
2. Calculate the geometric mean of the ratio cumulative numbers in the previous points $(1+X_t)$ from $t=1$ to $t=b-1$ as the estimation of the parameter IR.
3. Estimate the parameters θ and K due to the following linear relation

$$X_t \cong {}^{K}/_{t^{\theta}} \Rightarrow {}^{1}/_{X_t} \cong {}^{t^{\theta}}/_{K} \Rightarrow \ln\left({}^{1}/_{X_t}\right) \cong \theta \ln(t) - \ln(K) \Rightarrow \ln(X_t) \cong -\theta \ln(t) + \ln(K) \Rightarrow \ln(X_t) \cong \theta \ln\left(\frac{1}{t}\right) + \ln(K),$$

and

4. Multiply all the observations after

$$t = b - 1 \text{ by } \frac{t^{\theta}}{K}$$

to have an identical mean and variance for all of the newly obtained data (W_t):
$X_t \sim Normal\left({}^{K}/_{t\theta}, {}^{K^2}/_{at_{2\theta}}\right) Wt = {}^{t\theta}/_{K} X_t \sim Normal\left(1, {}^{1}/_{a}\right)$.

Therefore, the variance of the newly obtained data is a good candidate for ${}^{1}/_{a}$. Accordingly,

$\hat{a}=1/S_W^2$.

The most significant feature of a data-driven mathematical model is its competence in describing the data. The next step is devoted to representing the flexibility and ability of the model.

This model allows us to provide analysis for the propagation of

- the SARS epidemic of 2003[27]
- the MERS epidemic in South Korea May 20th – July 7th, 2018[28]
- the Ebola outbreak of 2014-2016[29-30]
- the propagation of HIV/AIDS from 1990 to 2018[31]
- the spreading of the Cholera of 2008-2009 in Zimbabwe[32],
- the COVID-19 epidemic in China and four of its provinces, Beijing, Guangdong, Shanghai, and Hubei in 2020[1],
- the propagation of the COVID-19 pandemic in Iran, 2020[23],
- the propagation of the COVID-19 pandemic in the UK, 2020[33-34],
- the spreading of COVID-19 in Italy[35], and
- the COVID-19 epidemic in the USA[36].

At first, according to our calculation, the estimations listed in Table 1 are obtained. Also, to illustrate the ability of our model to fit the pattern of different time series, we present the average curve and 75% upper and lower bounds(1) of 100 simulations conducted in the software MatLab (Plot 2).

Results

Fitting the pattern

This model enables us to compare the trend of propagation of different epidemics or diseases in different populations. For example, based on the calculations to fit this model to describe the spreading of some epidemics (Table 1), during the periods of study,

From the viewpoint of the outbreak in the first days (b), SARS has the least length of extreme behavior, and cholera possesses the most intense propagation. Regarding COVID-19, the USA possesses the longest period of extreme behavior, and this point is well reflected by its first rank among the involved countries.

Although the calculated reproduction number for COVID-19 is higher, Ebola showes the highest relative increment in the first days. This point can be due to the high percentage of asymptomatic infected cases in COVID-19. Considering the spreading of COVID-19, the highest initial relative increment belongs to Iran. For this reason, the least distance between the first case and becoming an epicenter belongs to Iran.[37]

COVID-19 has the most acute falling of the relative increments (θ).

In terms of fluctuation, HIV/AIDS has the most oscillating pattern (a). Concerning COVID-19, the USA, the UK, and Iran are the three countries with the highest fluctuation.

Note that the unstable pattern of the data during the first days justifies its relatively poor performance in comparison with the following days (illustrated by the exit from bounds in Plot 2). However, this problem is common, and other models suffer this drawback as well.

Prediction

To assess the performance of our model, we use it to predict over a period and then we

(1) Since 90% or 95% intervals mainly lead to zero as the lower bound, we use 75% interval to better highlight the lower bound.

investigate the accuracy of point predictions by using the following criterion:

$$Relative\ error = |\ predicted\ value - real\ value\ |\ /\ real\ value \times 100\%.$$

The number of cases infected by the first wave of COVID-19 in the USA (12 April – 21 May), the UK (15 April – 30 May & 31 May – 1 July), Italy (9 April – 18 May), and Iran (15 March – 15 April) was predicted by our model.[23 & 33-36] We apply two following criteria to assess the goodness of these predictions:

C1. The confidence interval contains the real value, and

C2. The relative error is less than a cut-off point (25%).

Since ARIMA is the most frequently used time series model, we used it to predict the number of cases over the mentioned periods in the mentioned countries. Then, we compare the performance of the two models using the criteria above. Also, we added the results of the other studies using ARIMA to predict the propagation of COVID-19 to further highlight the results. These studies include KSA (21 April – 21 May), Iran (5 – 24 May & 24 April – 7 July), Italy (31 March – 31 May & 24 April – 7 July), the USA (24 April – 7 July), and the UK (24 April – 7 July).[38-41]

Table 2 shows that our model has a great performance in terms of the second criterion, and it is superior to ARIMA while from the viewpoint of the first criterion, ARIMA is a little better than it. It is noticeable that the relatively narrower intervals of our model

Table 2. Some examples of performance of our model and ARIMA model

Country	Intended Period	Count Before Period	Real Count During		Predicted Count During (80% CI)	Relative Error
KSA	21 Apr – 21 May	11631	53 K	ARIMA [38]	115 K	117%
Iran	5 May – 24 May	99970	37 K	ARIMA [39]	18 K	51%
Italy	31 Mar – 31 May	105776	127 K	ARIMA [40]	77 K	39%
Iran	24 Apr – 7 Jul	88194	168 K	ARIMA [41]	92 K	45%
Italy	24 Apr – 7 Jul	192979	149 K	ARIMA [41]	207 K	39%
USA	24 Apr – 7 Jul	933833	2167 K	ARIMA [41]	2700 K	25%
UK	24 Apr – 7 Jul	129975	157 K	ARIMA [41]	370 K	136%
UK	15 Apr – 30 May	93400	159 K	Model [33]	189 (149-223 K)	19% *
				Model [34]	180 K (150-207 K)	13% *
				ARIMA	513 K (198 – 828 K)	223%
UK	31 May – 1 July	232664	52 K	Model [34]	59 K (47-72 K)	13% *
				ARIMA	66 K (8 – 104 K)	27% *
Iran	15 Mar – 15 Apr	13938	62 K	Model [23]	57 K (21 – 119 K)	8% *
				ARIMA	59 K (41 – 85 K)	5 % *
USA	12 Apr – 21 May	546508	1091 K	Model [36]	919 K (830 – 995 K)	16%
				ARIMA	899 K (679 – 1119 K)	18 % *
Italy	9 Apr – 18 May	139408	87 K	Model [35]	98 K (87-109 K)	13%
				ARIMA	142 K (44 – 235 K)	63 % *

Asterisk shows that the CI of prediction includes the real count of interest.

is the main reason for its inferiority in terms of interval predictions. However, if we used 90% intervals, our model would be superior to ARIMA in both aspects.

In terms of interpretability, while the ARIMA models fitted to the countries Saudi Arabia, Iran, Italy, the UK, and the USA say nothing about their comparison, the new model enables us to compare the trend of propagation in different countries. Based on the estimated parameters of the model to describe the spreading of COVID-19 in different countries, we realize some valuable points -which are comparable- including:

- the length of un-controlled behavior of the epidemic in different regions,
- the extreme intensity of the epidemic in different regions, and
- the rate of falling relative increments in different regions.

Discussion

There are ways to restrict or generalize the model; We can omit the two first parameters to obtain a simplified model suitable to describe the gradual decrease of the relative increments ($t=b,b+1,...$), or generalize the model by replacing a of the first or second part with a new parameter.

Although the ability of our model is limited to the first wave, it is possible to adopt this model for the other waves. If we fit the model for each wave separately and ignore a number of days at the beginning of each wave, it is possible to apply it to other waves as well.

Our model needs a transformation to change the numbers to relative increments and therefore an inverse transformation to change the increments back to numbers. Although these two transformations are straightforward, it may seem a drawback.

The parameters in this model are interpretable and this is an advantage of our model in comparison with the ARIMA model. This model is much less complicated than ARIMA models; It has only five parameters, therefore, its complexity is comparable with an ARIMA (p,d,q) model, which $p + q \leq 4$, while the volume of calculation of this model is much less than ARIMA models. Since our model is trend-based, the estimation of its parameters is easier than ARIMA models, which are correlation-based.

Both the ARIMA model and our model are based on some presumptions. ARIMA applies to the dataset that after some differentiating (once, twice, or d times), a stationary time series will be obtained. If this presumption is not met, we must look for some suitable transformation to get a stationary time series. Box-Cox is one of these transformation methods.[22] It is noticeable that sometimes the transformation does not lead to a suitable time series. About our model, we presume that the relative increment is distributed normally. However, neither ARIMA nor our model fits well more than a wave.

Finally, the length of the forecast intervals by ARIMA models is significantly greater than the analogous confidence intervals in our model.[38-41] In some cases, the interval predicted by ARIMA models is so wide that it is trivial, non-informative, and worthless. On the other hand, our narrower interval predictions have less likelihood of including the real value of interest.

Owing to the unstable pattern of the data during the first days of an outbreak, over these days, the model has a relatively poor performance in comparison with the following days. However,

this problem is common, and other models suffer this drawback as well.

Conclusion

In the present paper, we defined a novel bi-partite time series model able to describe the spreading of a communicable infection in a population from the start of its emerging to the end of the first wave. This model has five parameters and we explained how to estimate them. Because of these parameters and the structure of the model, it is flexible enough to reproduce the different stages of a great variety of communicable diseases, among them COVID-19 in Iran, Italy, the UK, the USA, China, and four of its provinces; Beijing, Guangdong, Shanghai, and Hubei. In comparison with ARIMA time series models, our model is more initiative and less complicated; it has fewer parameters; the estimation of its parameters is more straightforward; and its point and interval forecasts are more accurate and narrower, respectively. The model has some limitations to consider, including three presumptions, the unstable pattern of data in the initial days of the wave and therefore its relatively poor performance, the ability to model the first wave and adapt it to the other waves, and required transformations.

References

1. WHO. Coronavirus disease 2019 (COVID-19) situation reports: Jan 2020 to Nov 2022, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

2. Church DL (2004) Major factors affecting the emergence and re-emergence of infectious diseases, Clin Lab Med. 24(3): 559-86. doi: 10.1016/j.cll.2004.05.008. PMID: 15325056; PMCID: PMC7119055.

3. Jamshidi B, Alavi SMR, & Parham GA (2019) The distribution of the number of the infected individuals in a stochastic SIR model on regular rooted trees, Communications in Statistics: Computation and Simulation, DOI: 10.1080.03610918.1584299.

4. Brauer F, Castillo-Chavez C (2010) Mathematical Models in Population Biology and Epidemiology, 2nd edition, Springer.

5. Hamer WH (1906) The Milroy Lectures on Epidemic Disease in England: The Evidence of Variability and of Persistency of Type, Bedford Press.

6. Ross R (1911) Some quantitative studies in epidemiology, Nature 87: 466–467.

7. Kermack M & McKendrick AG (1927) Contributions to the mathematical theory of epidemics. Part I, Proceedings of the Royal Society. Series A, Containing Papers of a Mathematical and Physical Character 115(5): 700–721.

8. Bartlett M (1949) Some evolutionary stochastic processes, Journal of the Royal Statistical Society. Series B (Methodological) 11 (2): 211–229.

9. En'ko PD (1989) On the course of epidemics of some infectious diseases, International Journal of Epidemiology 18 (4): 749–755.

10. Diekmann O, Heesterbeek J, & Metz JA (1990) On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations, Journal of Mathematical Biology 28 (4): 365–382.

11. Silva, Petrônio CL et al. (2020) COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos, solitons, and fractals vol. 139: 110088. doi:10.1016/j.chaos.2020.110088

12. Firth JA, Hellewell J, Klepac P, et al. (2020) Using a real-world network to model localized COVID-19 control strategies. Nat Med 26, 1616–1622. https://doi.org/10.1038/s41591-020-1036-8

13. Mesa: http://mechatronics.ucmerced.edu/covid19

14. Deschepper M, Eeckloo K, Malfait S, Benoit GD, Callens S, and Vansteelandt S (2020) Prediction of hospital bed capacity during the COVID-19 pandemic: DOI: https://doi.org/10.21203/rs.3.rs-60176/v1

15. Zhang X, Liu Y, Yang M, Zhang T, Young AA, et al. (2013) Comparative study of four time series methods in forecasting Typhoid fever incidence in China. PloS one 8: e63116.

16. Li Q, Guo N-N, Han Z-Y, Zhang Y-B, Qi S-X, et al. (2012) Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic Fever with renal syndrome. The American Journal of Tropical Medicine and Hygiene 87: 364–370.

17. Nobre FF, Monteiro ABS, Telles PR, & Williamson GD (2001) Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. Statistics in Medicine 20: 3051–3069.

18. Zhang X, Zhang T, Young AA, & Li X (2014) Applications and comparisons of four time series models in epidemiological surveillance data. PLoS ONE 9(2): e88075. https://doi.org/10.1371/journal.pone.0088075.

19. Gu Y, COVID-19 Projections Using Machine Learning. https://covid19-projections.com/

20. EpiGro: https://jocelinelega.github.io/EpiGro/

21. Jamshidi B, Jamshidi Zargaran S, and Rezaei M (2020) Modelling Spreading of an Infection Using Time series by a Novel Family of Models; Fitting the Time Series of the confirmed cases of COVID-19 in China, Epidemiologic Methods, 9(s1).

22. Box G & Jenkins G (1990) Time Series Analysis: Forecasting and Control, Holden-Day.

23. Jamshidi B, Rezaei M, Najafi F, & Sheikhi A (2020) Forecasting the spreading of COVID-19 in Iran from March 15th to April 15th, 2020, Iranian Red Crescent Medical Journal 22(5).

24. Britton T, House T, Lloyd AL, Mollison D, Riley S, & Trapman P (2014) Five challenges

for stochastic epidemic models involving global transmission, Epidemics 10: 54-57.

25. Jamshidi B, Rezaei M, & Rezaei K (2020) A new model for epidemic spreading (with focus on COVID-19), Health Scope 9(3).

26. Brauer F, Castillo-Chavez C, & Feng Z (2019) Mathematical Models in Epidemiology, Springer.

27. WHO. Severe Acute Respiratory Syndrome 2003 (SARS) situation reports: Mar - July 2003, https://www.who.int/csr/sars/country/en/

28. WHO. MERS-Cov 2018 situation reports: May - July 2018, https://www.who.int/csr/don/archive/disease/coronavirus_infections/en/

29. WHO. Ebola 2014-2015 situation reports: Mar 2014 - Nov 2015, https://www.who.int/csr/disease/ebola/situation-reports/archive/en/

30. Freiberger M (2015) Ebola in numbers: using mathematics to tackle epidemics, Science in School I (32): 14-19.

31. UN. Global HIV and AIDS statistics 2019, https://www.unaids.org/en/resources/fact-sheet

32. WHO. Cholera 2008-2009 situation reports: https://www.who.int/csr/don/archive/country/zwe/en/

33. Jamshidi B, Rezaei M, Kakavandi M, & Jamshidi Zargaran S (2020) Modeling the number of confirmed cases and deaths from COVID-19 pandemic in the UK and forecasting over April 15th – May 30th, Disaster Medicine and Public Health Preparedness, doi:10.1017/dmp.2020.312.

34. Jamshidi B, Bekrizadeh H, Jamshidi Zargaran S, & Rezaei M (2020) Modelling propagation of COVID-19 in the UK, Accepted, Disaster Medicine and Public Health Preparedness.

35. Jamshidi B, Talaei-khuei A, Jamshidi Zargaran S, & Rezaei M (2020) Forecasting confirmed cases of COVID-19 in Italy, Preprint.

36. Jamshidi B, Talaei-khuei A, Kakavandi M, & Jamshidi Zargaran S (2020) Forecasting confirmed cases and mortalities of COVID-19 in the US, Preprint: https://www.researchgate.net/publication/340606620_Modelling_and_Forecasting_The_Number_of_Confirmed_Cases_and_Deaths_from_COVID-19_Pandemic_in_USA.

37. Jamshidi B, Rezaei M, Jamshidi Zargaran, S & Najafi, F (2020) Mathematical modeling the epicenters of coronavirus disease-2019 (COVID-19) pandemic, Epidemiologic Methods 9 (s1). 20200009. https://doi.org/10.1515/em-2020-0009.

38. Alzahrani SI, Aljamaan IA, & Al-Fakih EA (2020) Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions, Journal of Infection and Public Health 13:914–919.

39. Tran TT, Pham LT, Ngo QX (2020) Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran), Global Journal of Environmental Science and Management 6(SI):1-10.

40. Chintalapudi N, Battineni G, & Amenta F (2020) COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach, Journal of Microbiology, Immunology and Infection 53 (3): 396-403; https://doi.org/10.1016/j.jmii.2020.04.004.

41. Singh RK, Rani M, Bhagavathula AS, Sah R, et al. (2020) Prediction of the COVID-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model, JMIR Public Health and Surveillance 6(2):e19115; doi: 10.2196/19115.