Journal of Turkish

Operations Management

Modelling and predicting the growth dynamics of covid-19 pandemic: A comparative study including Turkey

Nadi Serhan Aydın[1], Erfan Babaee Tirkolaee[2*]

[1] Department of Industrial Engineering, Istinye University, Maltepe Neighbourhood, Teyyareci Sami Street, Building No.3, Zeytinburnu, Istanbul, 34010, Turkey
e-mail: serhan.aydin@istinye.edu.tr, ORCID No: https://orcid.org/0000-0002-1453-0016
[2] Department of Industrial Engineering, Istinye University, Maltepe Neighbourhood, Teyyareci Sami Street, Building No.3, Zeytinburnu, Istanbul, 34010, Turkey
e-mail: erfan.babaee@istinye.edu.tr, ORCID No: https://orcid.org/0000-0003-1664-9210
*Corresponding Author

Article Info

Article History:
Received: 08.08.2021
Revised: 16.09.2021
Accepted: 17.09.2021

Keywords
Epidemics Modelling,
Exponential Model,
Logistic Model,
Gompertz Growth,
SIR/SEIR Model

Abstract

Estimating the growth dynamics of a pandemic is critical for policy makers to fine-tune emergency policies in health and other public sectors. The paper presents country-level calibration and prediction results on some well-known models in the literature, namely, the logistic, exponential, Gompertz, SIR and SEIR models. The models are implemented on real data from various countries, including Turkey, and their performance for different estimation windows have been analyzed using R^2 scores. The computational results are obtained using Python. The Gompertz model outperforms other models by consistently offering a better fit for the total number of infected. The exponential model is helpful in describing the growth dynamics in the early stages of the COVID-19 pandemic. Suspected-Infected-Recovered (SIR) and Susceptible-Exposed-Infectious-Removed (SEIR) models display a fair performance on the underlying active cases data in many circumstances. Quantitative models can offer policy makers in Turkey and elsewhere a better insight on the evolution of pandemic when everything else is held constant and the infections follow a typical path. The results can be highly sensitive to changes in policies. There is not a single model that can perfectly mimic all stages of pandemic. An ensemble model or multi-modal distributions can be used to capture the evolution of multi-wave pandemics.

1. Introduction

Since the discovery of the first official case in December 2019 in the city of Wuhan, China (which followed by a health emergency declaration by the World Health Organization, or WHO, on January 30, 2020, due to the obvious risk of a *pandemic*), the field of epidemiology have played an important role in helping policy makers estimate the future trajectory of the disease and better anticipate the potential overload on their health systems and other public services (Tomaskova and Tirkolaee, 2021; Tirkolaee and Aydın, 2021). At the time of writing (June 2021) there are 185 million confirmed cases and almost 4 million deaths ascribed to COVID-19. Turkey has registered a total 5.5 million confirmed cases on the same date, of which 5.3 million recovered and around 50 thousand lost their lives. The explorations on the transmission dynamics of the virus and its genetic structure have been high on the world agenda with a view to developing effective healthcare response and recovery mechanisms in the short- as well as long-run.

Mathematical models on transmissible diseases have widely proved to be helpful in not only gaining insights into but also making predictions about the growth dynamics of infectious diseases and the potential impacts of alternative intervention policies. The recent outbreak of SARS-CoV-2 has once again brought the accuracy and usefulness of different models under spotlight.

The objective of this work is to review some of the widely used growth models, namely the logistic, exponential, Gompertz, Suspected-Infected-Recovered (SIR) and Susceptible-Exposed-Infectious-Removed (SEIR) models for the COVID-19 dynamics and calibrate them to real epidemic data, including those of Turkey, Iran and some other countries. We then compare their goodness-of-fit. The final objective is to identify the most effective model(s) in predicting the dynamics and evolution of the COVID-19 so as to provide decision-makers with some more insight into the future possible trends. To this end, in Section 2, we survey some of the most relevant studies in the literature. The reviewed models and calibration results are presented in Section 3. Finally, Section 4 concludes with a brief summary and outlook for future research.

2. Selected Literature

Wu et al. (2020) employed the logistic model, the generalized logistic model, the generalized Richards model and the generalized growth model to investigate the dynamics of infections for 29 provinces in China and 33 other countries. A comparison was made by Liang (2020) for the novel transmission dynamics of the SARS-CoV-2 as compared to Severe Acute Respiratory Syndrome (SARS) and Middle-East Respiratory Syndrome (MERS). They developed a propagation growth model considering the infection inhibition constant and growth rate. According to the main finding of their research in Hubei province, the growth rate of COVID-19 is approximately twice that of the infections caused by SARS and MERS. Ma (2020) discussed the estimation of the growth rate using maximum likelihood method, as well as parametric and non-parametric approaches for exponential growth rate and basic reproduction number, and the least squares estimation model. Two differential equations models were proposed by Liu et al. (2020) to account for exposed or latency period of SARS-CoV-2 infections in China. They assessed the epidemiological parameters such as the transmission rate and the basic reproduction number and analyzed the effect of the exposed or latency period in the transmission dynamics of SARS-CoV-2.

Sarkar et al. (2020) suggested a mathematical model to predict the dynamics of COVID-19 in India. Their model was established based on the dynamics of six various components including asymptomatic, recovered, infected, isolated infected and quarantined susceptible cases. They found that reducing contact rate between uninfected, infected and quarantined people caused a reduction in the basic reproduction rate. Moreover, they claimed that performing social-distancing and effectively tracing contacts can significantly help eliminate the pandemic. Several forecasting models were employed by Sharma and Nigam (2020) to investigate the COVID-19 growth curve in India. They utilized exponential and polynomial regression analysis, Auto-Regressive Integrated Moving Averages (ARIMA) model, and exponential smoothing. An ARIMA(5, 2, 5) model was found to the most suitable model in predicting the number of cases in India. On the other hand, in Velasquez and Lara (2020), authors predicted and analyzed the COVID-19 incidence in the U.S. using a reduced-space Gaussian process regression model. The proposed model was related to chaotic dynamical systems. A modified mathematical model was proposed in Çakır and Savaş (2020) to simulate the spread of COVID-19 in Iran.

A binary classification model based on neural network was suggested in Pirouz et al. (2020) where authors concluded that the relative humidity and maximum daily temperature had the greatest influence on the number of confirmed cases. Authors of Rath et al. (2020) employed multiple linear regression model to predict the new active COVID-19 cases in India. They compared their proposed model with a simple linear regression model and, using Analysis of Variance (ANOVA), demonstrated the superiority of their model. In Duhon et al. (2021), researchers developed a multiple regression model to identify the main factors affecting the initial growth rate of the COVID-19 pandemic. It was revealed that socio-demographic and climatic variables are highly connected to the initial growth rate while others represent a weak connection. In Carcione et al. (2020), authors carried out a simulation of the COVID-19 epidemic using a deterministic SEIR model to estimate the numbers of infected people and casualties. They found out that the isolation measures, knowledge of the transmission conditions and social distancing were critical factors that affected the growth pace of the outbreak. Authors in Sun et al. (2020) conducted a research for predicting the long-term trend of COVID-19 epidemic in China by extending the classical SEIR model to Dynamic-Susceptible-Exposed-Infective-Quarantined (D-SEIQ).

In Nikolopoulos et al. (2021), authors explored the COVID-19 growth rate using statistical, epidemiological, machine- and deep-learning models, and a hybrid forecasting approach. A machine learning tool was employed in Li et al. (2021) to determine new factors related to COVID-19 transmission and fatality, such as high temperature, economic inequality and blood types. In Tuli et al. (2020), machine learning and cloud computing tools were utilized to predict the growth and conjecture of the COVID-19 pandemic. They found that the method of iterative weighting can enhance the fitting performance of the Generalized Inverse Weibull distribution so as to develop a more accurate and real-time prediction framework via cloud computing. The evolution of COVID-19 pandemic in Turkey was modelled in Acar et al. (2021) using a probabilistic approach that employs a Bayesian negative binomial multilevel model with mixed effects. The proposed method predicted the daily confirmed COVID-19 cases and cumulative numbers for 20 days with different prediction intervals. There are a handful of other studies that focus on Turkey (Eroğlu, 2020); Baldemir et al., 2020; Önder, 2020).

Against this backdrop, this study aims to contribute to the line of literature that aims to offer more insight into real-world parameter estimations of epidemic models, with particular emphasis on Turkey. We present detailed estimation results on some of the well-known mathematical models, namely, the logistic, exponential, Gompertz, SIR and SEIR models, for a considerable number of countries, including Turkey. However, our end-to-end algorithm is able to extract the necessary data for any other country and perform calibrations on real-world data.

3. Models And Calibration Results

This study conforms to the research and publication ethics. We first discretize the model training period $t \in [0, T]$ as $t_i \in \{t_0, t_1, \dots, t_{n_0}\}$ where $t_0 = 0$ and $t_{n_0} = T$. The off-sample forecasting window will cover $n_1 = 20$ days. The model performances will then be compared using both in- and off-sample R^2 values. The logistic model is one of the most widely-used models in the literature that has proved to be effective in describing health phenomena like epidemics. Under this model, the total number of infected is given by

$$I_t^{(1)} = \frac{c}{1+e^{-\frac{t-b}{a}}}, \tag{1}$$

where a, b, c are the infection speed, inflection point (i.e., the point at which the maximum increase in the number of infected occurs) and the estimated number of infected once the pandemic ends, respectively. Setting $\gamma(t) = e^{-(t-b)/a}$, this can be easily seen from the second derivative I'' which is equal to $\frac{c\gamma(1-\gamma^2)}{a^2(1+\gamma)^4}$. Setting $I'' = 0$ gives $\gamma = 1$ which is possible only when $t = b$ or $t \to \infty$. This verifies b as the inflection point. The meaning of parameter c, on the other hand, can be validated by looking at $\lim_{t\to\infty} I_t$ which is apparently equal to c, rendering the latter as a horizontal asymptote. This is also the level at which I_t flattens out. In other words, $I_t' = \frac{c\gamma}{a(1+\gamma)^2} = 0$ implies $t \to \infty$.

While the logistic model converges to a constant over the long-run, the exponential model helps mimic periods where pandemics seem to be out of control (at least for a while). In its most general form, the total number of infected is modelled through the function

$$I_t^{(2)} = a\, e^{b(t-c)}, \tag{2}$$

where a, b, c are the initial number of infected, growth rate and infection start date, respectively. For the growth rate, observe that $I_t' = bI$ or $\mathrm{d}I_t = b\mathrm{d}t$. The peculiarity of the exponential model is that it doesn't converge in the long-run, i.e., $\lim_{t\to\infty} I_t^{(2)} = \infty$. Yet, there are periods in epidemic outbreaks that the exponential model explains the data well at least for some time. As an outlook, future research can harness a model selection tool which switches between different models at different phases of the pandemic (such as, second or third waves) to form an ensemble model that better represents the underlying data. A comparison of exponential versus sub-exponential growth is presented in Figure 1.

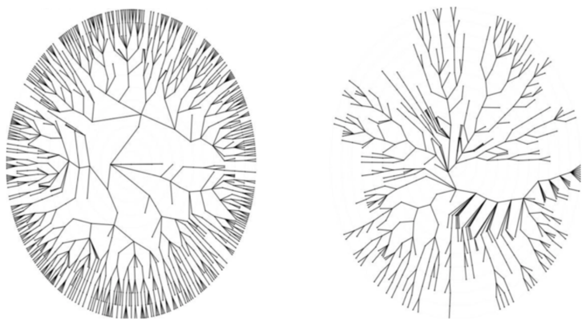

Fig. 1. Epidemic growth dynamics by Chowell et al. (2016).

Another model that is considered in this paper is the `Gompertz growth' which is in fact a special case of the logistic growth and also analogous to the sigmoid function used as an activation function in neural nets. Unlike the logistic growth, which is perfectly symmetrical around the inflection point, the Gompertz function presents a non-symmetrical behaviour characterized with a slower growth and convergence during the slowing-down phase of the pandemics. Due to the drawbacks in effectively calibrating the location parameter b, we consider a slightly modified version of the Gompertz model. Under this model, the total number of infected is given by

$$I_t^{(3)} = a\exp\{-e^{-b(t-c)}\}, \tag{3}$$

where a, b, c are the upper bound for the total number of infected (i.e., $\lim_{t\to\infty} I_G = a$), the growth rate and the location parameter, respectively. The maximum rate of increase in the number of infected occurs at $t = \frac{\ln b}{c}$ which can be seen by setting $I'' = Ib^2c^2e^{-ct}\left(e^{-ct} - \frac{1}{b}\right)$ equal to 0.

We perform the calibrations mainly using Python's *scipy* as well as *scikit* library. Our code is able to extract country data from related data sources and perform curve-fitting with stream data. The estimation results of all three models above for some selected countries, including Turkey, and different values of n_0 are presented in Figures 2-5. Green, black and blue lines represent the exponential, Gompertz and logistic models, respectively. Also, red and orange dots represent the real data, the latter being the test sample which is not used in the model training. Table 1 includes detailed calibration results, estimated parameter values, and model performances in terms of their R^2 values, both in- and off-sample.

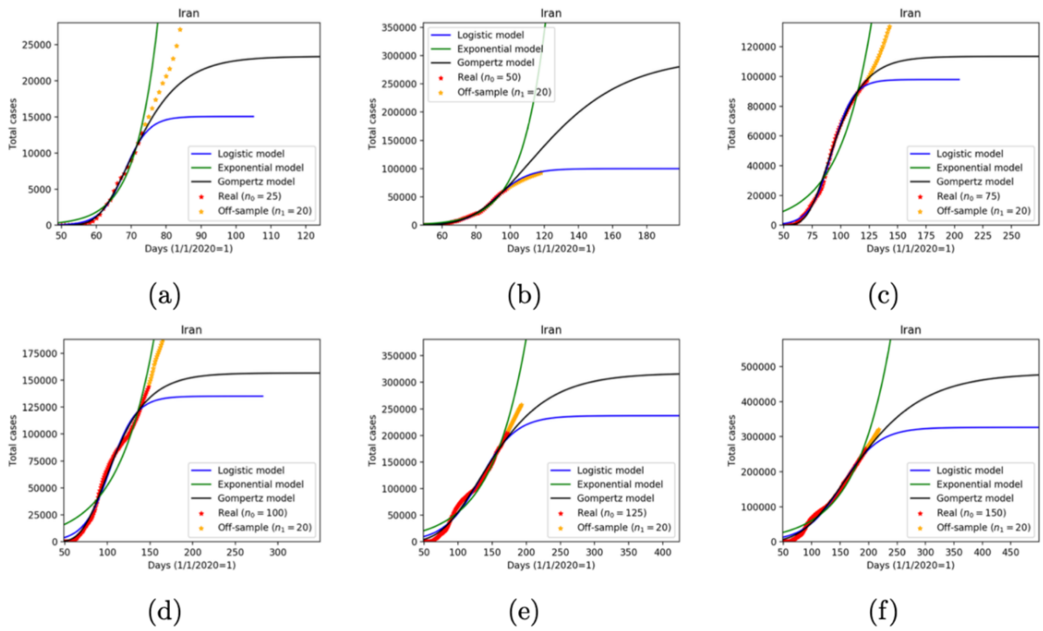

Fig. 2. $I_t^{(1)}$, $I_t^{(2)}$, $I_t^{(3)}$ calibrated to data from Turkey for $n_0 = \{25,50,75,100,125,150\}$ and $n_1 = 20$.

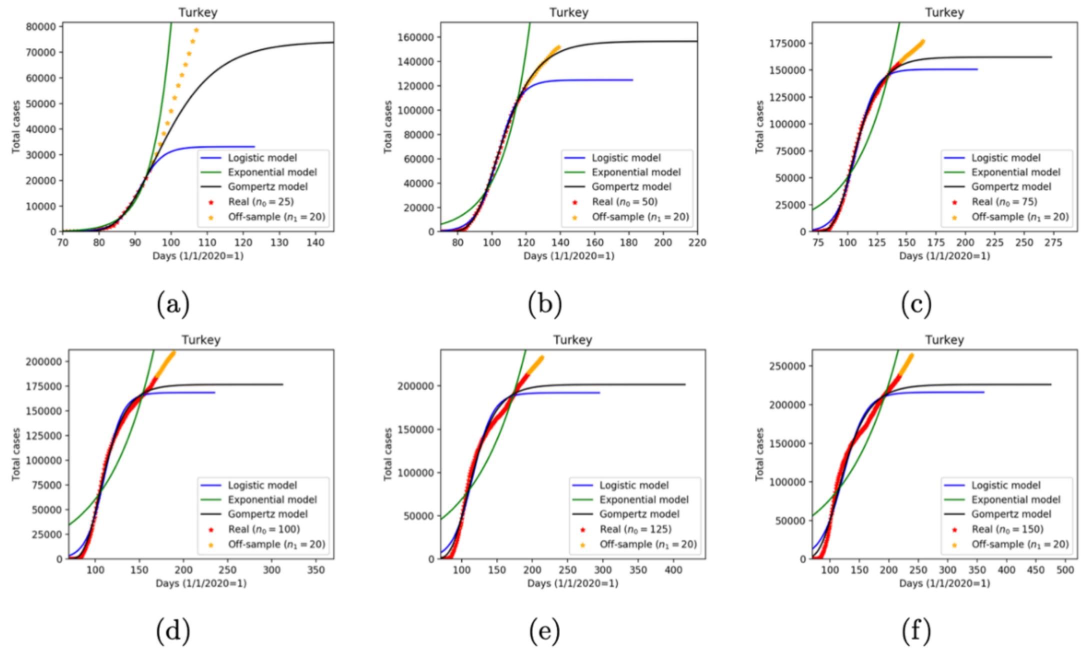

Fig. 3. $I_t^{(1)}$, $I_t^{(2)}$, $I_t^{(3)}$ calibrated to data from Iran for $n_0 = \{25,50,75,100,125,150\}$ and $n_1 = 20$.

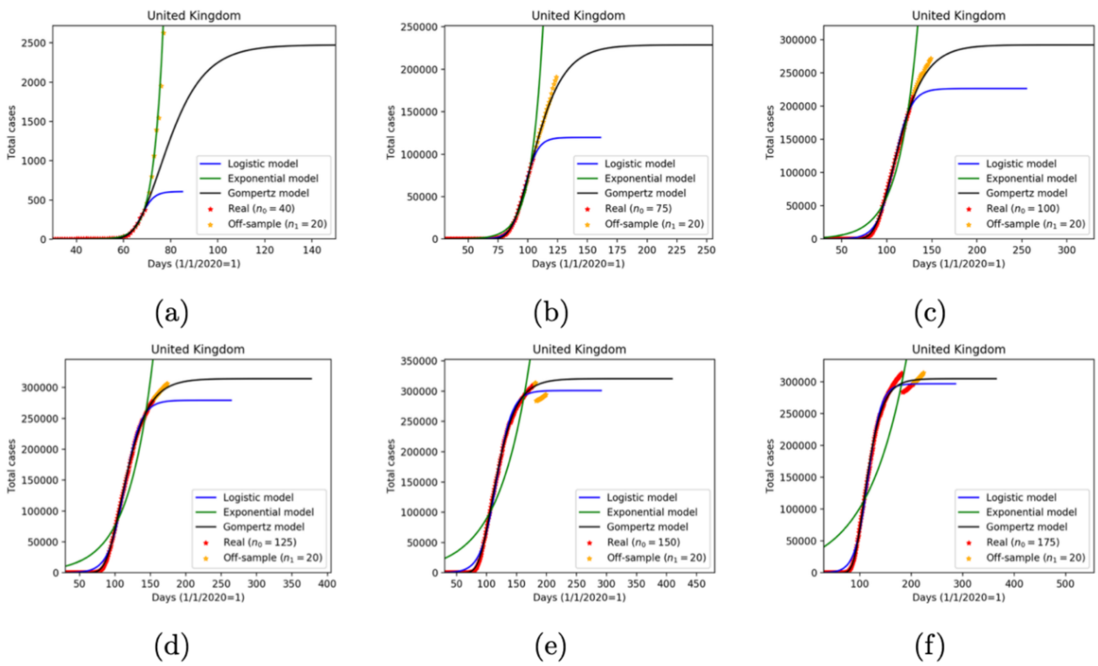

Fig. 4. $I_t^{(1)}$, $I_t^{(2)}$, $I_t^{(3)}$ calibrated to data from UK for $n_0 = \{40,50,75,100,125,150,175\}$ and $n_1 = 20$.

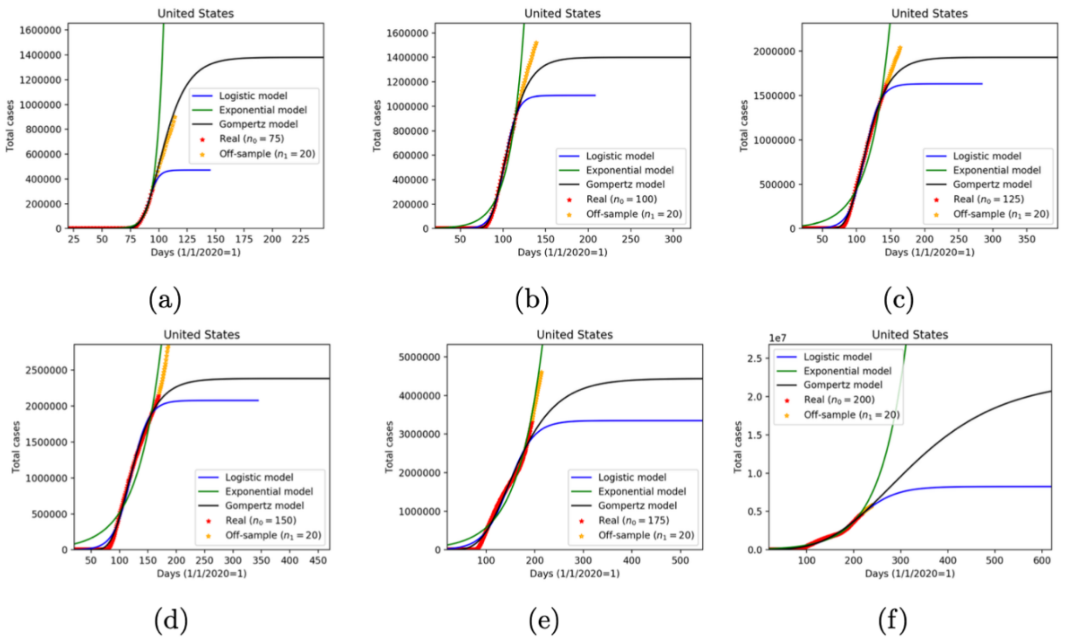

Fig. 5. $I_t^{(1)}$, $I_t^{(2)}$, $I_t^{(3)}$ calibrated to data from US for $n_0 = \{75,100,125,150,175,200\}$ and $n_1 = 20$.

Table 1. Model calibration results for selected countries.

		Logistic model					Exponential model					Gompertz model				
	n	a	b	c (10^3)	R^2	R^2 (off)	a	b	c	R^2	R^2 (off)	a (10^3)	b	c	R^2	R^2 (off)
Turkey	25	3.13	91.23	33.14	0.9984	0.3608	0	0.19	12.29	0.9843	NaN	74.25	0.1	95.24	0.9989	0.8175
	50	6.28	103.41	124.69	0.9988	0.9733	27.18	0.06	-14.41	0.9426	NaN	156.57	0.08	102.22	0.9997	0.9995
	75	8.26	107.2	150.64	0.9957	0.9811	81.49	0.03	-110.7	0.8665	0.259	161.99	0.07	102.83	0.9995	0.9945
	100	10.43	110.49	168.58	0.9898	0.9622	89.81	0.02	-245.77	0.8193	0.6608	176.67	0.06	104.53	0.9971	0.9787
	125	14.3	115.83	192.04	0.9755	0.9555	208.1	0.01	-320.87	0.8186	0.7363	201.5	0.04	107.93	0.9884	0.9752
	150	19.15	122.43	216.01	0.967	0.9495	134.54	0.01	-489.04	0.819	0.7851	225.92	0.03	111.94	0.9829	0.97
Iran	25	3.43	68.16	15.06	0.9941	0.3343	0.89	0.16	10.79	0.9697	NaN	23.37	0.12	69.26	0.9975	0.6071
	50	9.08	92.16	99.93	0.9962	0.9958	42.25	0.07	-6.77	0.9829	NaN	298.46	0.03	110.97	0.9969	0.8476
	75	9.42	92.16	97.94	0.9983	0.9515	62.17	0.03	-94.95	0.9117	0.4967	113.52	0.06	88.66	0.9979	0.9794
	100	15.22	103.2	135.06	0.9828	0.906	124.87	0.02	-157.47	0.9146	0.8854	156.56	0.04	97.71	0.9915	0.9397
	125	26.36	132.39	237.04	0.9786	0.9695	74.98	0.02	-241.8	0.9479	0.9164	317.41	0.02	131.13	0.9871	0.9826
	150	32.59	153.26	326.52	0.9865	0.986	183.59	0.02	-257.64	0.9582	0.9233	481.77	0.01	159.31	0.992	0.9936
UK	40	2.62	67.72	0.61	0.9953	NaN	0	0.24	0.85	0.9889	NaN	2.47	0.1	75.51	0.9932	0.0557
	75	5.56	97.21	119.73	0.9997	0.9061	12.1	0.1	13.79	0.984	NaN	228.68	0.06	102.03	0.9997	0.9984
	100	9.34	108.06	226.27	0.9968	0.9806	42.21	0.05	-55.01	0.9592	0.2427	291.84	0.05	106.88	0.9995	0.9985
	125	11.72	113.77	279.33	0.9972	0.9942	83.53	0.03	-142.08	0.9201	0.58	314.26	0.05	108.61	0.9997	0.9997
	150	13.2	116.47	300.85	0.9968	0.9964	126.35	0.02	-246.83	0.8743	0.5863	320.45	0.05	109.15	0.9998	0.9961
	175	12.85	115.92	296.83	0.9967	0.9966	232.95	0.01	-362.5	0.7927	0.667	304.81	0.05	107.81	0.9978	0.998
Germany	75	5.3	90.13	127.83	0.9993	0.9752	15.83	0.09	-0.64	0.9725	NaN	182.85	0.08	90.81	0.9996	0.9978
	100	7.04	93.77	160.15	0.9981	0.9929	53.86	0.04	-92.32	0.9027	0.1174	172.01	0.08	90.02	0.9998	0.999
	125	8.25	95.6	172.87	0.9968	0.9942	110.06	0.02	-193.14	0.8372	0.5338	178.78	0.08	90.63	0.9996	0.9986
	150	9.15	96.79	180.11	0.9953	0.9916	156.34	0.02	-308.55	0.7843	0.6386	184.2	0.07	91.13	0.9989	0.9965
	175	10.14	98.02	186.79	0.9924	0.9866	235.76	0.01	-414.48	0.7504	0.6716	190.44	0.06	91.73	0.997	0.9927
Russia	100	8.29	125.37	320.85	0.9986	0.9744	11.18	0.08	9.46	0.9918	NaN	1120.23	0.03	145.97	0.9996	0.986
	125	10.27	132.58	468.44	0.999	0.9729	41.74	0.04	-58.05	0.9668	0.5549	642.92	0.04	132.86	0.9998	0.996
	150	14.21	142.63	651.38	0.9968	0.9867	89.8	0.03	-134.96	0.9504	0.7572	800.48	0.03	139.24	0.9995	0.9976
	175	17.7	150.67	787.73	0.9956	0.9889	150.89	0.02	-218.43	0.9331	0.8227	916.07	0.03	144.21	0.9993	0.9974
	200	21.15	157.97	902.19	0.9942	0.9893	155.15	0.02	-327.87	0.9179	0.8456	1019.07	0.03	148.92	0.9989	0.9972
Japan	50	7.38	62.32	0.56	0.9948	0.6889	1.06	0.09	3.39	0.9889	0.5792	8.38	0.02	114.64	0.9928	0.976
	75	12.16	93.77	4.21	0.9921	0.5292	11.53	0.07	12.99	0.9898	0.8523	43.63	0.02	162.12	0.9937	0.6262
	100	8.85	108.15	19.2	0.9968	0.987	11.45	0.07	14.59	0.9886	NaN	102.45	0.03	142.53	0.9939	0.7
	125	8.05	105.48	16.55	0.9991	0.999	52.99	0.03	-36.88	0.9171	0.2958	17.79	0.07	101.32	0.9964	0.9971
	150	8.42	105.96	16.89	0.9992	0.996	103.97	0.02	-96.9	0.8463	0.6303	17.33	0.08	100.93	0.9977	0.9965
Indonesia	50	8.37	106.16	11.15	0.9983	0.9605	1.72	0.08	6.2	0.9885	NaN	42.17	0.03	129.04	0.999	0.9775
	75	11.8	118.01	19.27	0.9966	0.8927	23.41	0.05	-11.88	0.9709	0.8028	31.75	0.03	123.62	0.9992	0.9566
	100	18.37	144.65	45.51	0.996	0.9453	41.86	0.03	-38.96	0.9807	0.9477	102.63	0.02	168.24	0.9985	0.9782
	125	25.01	179.75	109.47	0.9969	0.9838	56.56	0.03	-60.42	0.9898	0.9701	488.91	0.01	259.02	0.9987	0.9959
	150	29	205.37	191.01	0.9985	0.9974	89.96	0.03	-74.69	0.9919	0.9571	1110.03	0.01	319.99	0.9994	0.9994
Brazil	50	6.01	100.88	41.16	0.9986	0.6017	4.73	0.11	22.58	0.9879	NaN	128.41	0.05	114.15	0.9987	0.8738
	75	11.71	139.47	526.52	0.9986	0.9673	24.48	0.07	8.66	0.9972	0.8495	43104.26	0.01	283.24	0.9992	0.9991
	100	13.88	158.34	1378.25	0.9995	0.9954	20.4	0.06	-28.92	0.9964	0.7635	26485.71	0.01	264.22	0.9997	0.9837
	125	15.91	168.41	1987.42	0.9992	0.9841	62.01	0.04	-70.69	0.9884	0.8435	5779.17	0.02	200.16	0.9997	0.9998
	150	19.03	181.71	2915.15	0.999	0.9849	79.64	0.03	-143.25	0.9801	0.8958	5616.35	0.02	198.57	0.9998	0.9985

Since all 3 models compared in the table are univariate, reporting the adjusted R^2 values would not affect the results. A quick observation from the table is that the exponential model is generally useful in modelling the early stages of pandemic, but this doesn't extend well into the later stages when the speed of infection tends to diverges from that implied by the exponential model. It also offers a relatively weak performance off the sample in most of the cases. Instead, logistic and Gompertz models are able to reproduce the overall pattern of infections and provide consistently more realistic estimates even for periods extending beyond the early stages of the pandemic. Overall, it can be concluded that the Gompertz offers a better performance off the sample, however. The model predicts not only a significantly higher value for the upper bound of the total number of infected in almost all cases, but also a longer period for the pandemic. Again, the exponential model performance seems to diverge from those of 2 other models after the early stages of the pandemic, both in- and off-sample.

Another popular model is the SIR Model which borrows its name from its underlying approach of partitioning the population into three sub-groups, namely, susceptible (S) but not yet infected, infected and infectious (I), and recovered and immune (R). The SIR model, invented by Ross (1911) and Hamer (1906), does not have an explicit solution, however. There are various assumptions behind the model that restrict the model's ability to fit different types of pandemic outbreaks. These include the assumptions that: (*i*) the population is large, closed and distributed homogeneously; (*ii*) the outbreak always dies out; (*iii*) no natural births or natural deaths occur; (*iv*) the infection has zero latent period, that is, an individual becomes infectious as soon as she becomes infected; and (v) infection leads to lifetime immunity for individuals. The SIR model is represented by the following system of quadratic ODEs:

$$\mathrm{d}S_t = -\beta S_t\, I_t^{(4)}\mathrm{d}t, \tag{4}$$

$$\mathrm{d}I_t^{(4)} = \left(\beta S_t I_t^{(4)} - \gamma I_t^{(4)}\right)\mathrm{d}t, \tag{5}$$

$$\mathrm{d}R_t = \gamma I_t^{(4)}\,\mathrm{d}t, \tag{6}$$

where β and γ are the transmission and recovery rates, respectively. All of the limits $\lim_{t\to\infty} S_t$, $\lim_{t\to\infty} I_t$ and $\lim_{t\to\infty} R_t$ exist, supporting (ii) above. Here, $\beta = \frac{b}{N}$ where $b = \kappa\tau$ with κ and τ being the number of contacts by each infected individual and the proportion of contacts that result transmission (known as the *transmissibility* parameter), respectively. Defining $S_t^{\%} = \frac{S_t}{N}$ as the susceptible proportion (and recalling the homogeneity assumption), each infected individual infects βS individuals. The effective reproductive number is given by $r_0 = \frac{S_0}{N}\frac{b}{\gamma} = \frac{S_0\beta}{\gamma}$. The threshold value for r_0 that determines whether an outbreak is predicted to die out soon or evolve into an epidemic is 1. If transmission rate per unit of time exceeds the recovery rate, then $r_0 > 1$ and an epidemic can be expected before I_t goes to 0 as $t \to \infty$. Otherwise, I_t decreases monotonically and reaches 0. If $r_0 > 1$, I_t can increase up to S_t before it starts to decrease. Control of epidemics is mostly about keeping r_0 below 1. This requires public interventions to (*i*) reduce $\frac{1}{\gamma}$ (duration of infection), (*ii*) reduce κ (contact rate of infected), (*iii*) reduce S_0 and (*iv*) reduce τ (transmissibility). The type of interventions (such as "mask-distance-hygiene" or antivirals) is beyond the scope of this study.

While calibrating the SIR model, one thing which is critical is to work with active infections (i.e., $\text{Total Cases} - \text{Deaths} - \text{Recoveries}$), rather than the cumulative cases directly. We again design our code to extract all of these ingredients from the relevant resources and then calculate the active cases before using them for estimating parameters of the SIR model. Then Python's *odeint* class under the *scipy* library is used to integrate SIR differential equations and solve for optimal parameter values for β and γ. In Figures 6-9, we present the parameter estimation results for $I_t^{(4)}$ using data from Turkey, Russia, Germany and Iran for different values of n, although our code is able to calibrate the model to many other country data of our choice. Estimated parameter values are provided in figure keys. In most cases, the SIR curve offers a fairly good representation of the underlying data and can be used to predict the number of infected. This is particularly true in the case of Germany (see Figure 8) where the pattern of infected closely follows an SIR curve characterized by the estimated parameters.

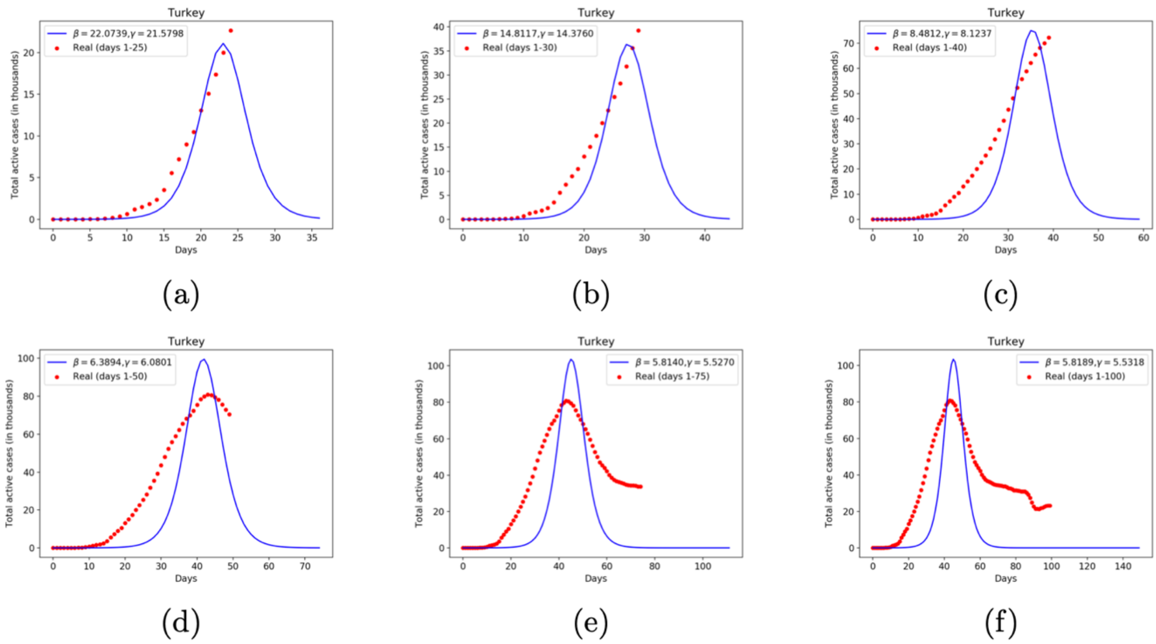

Fig. 6. SIR model $I_t^{(4)}$ calibrated to data from Turkey for $n \in \{25,30,40,50,75,100\}$.

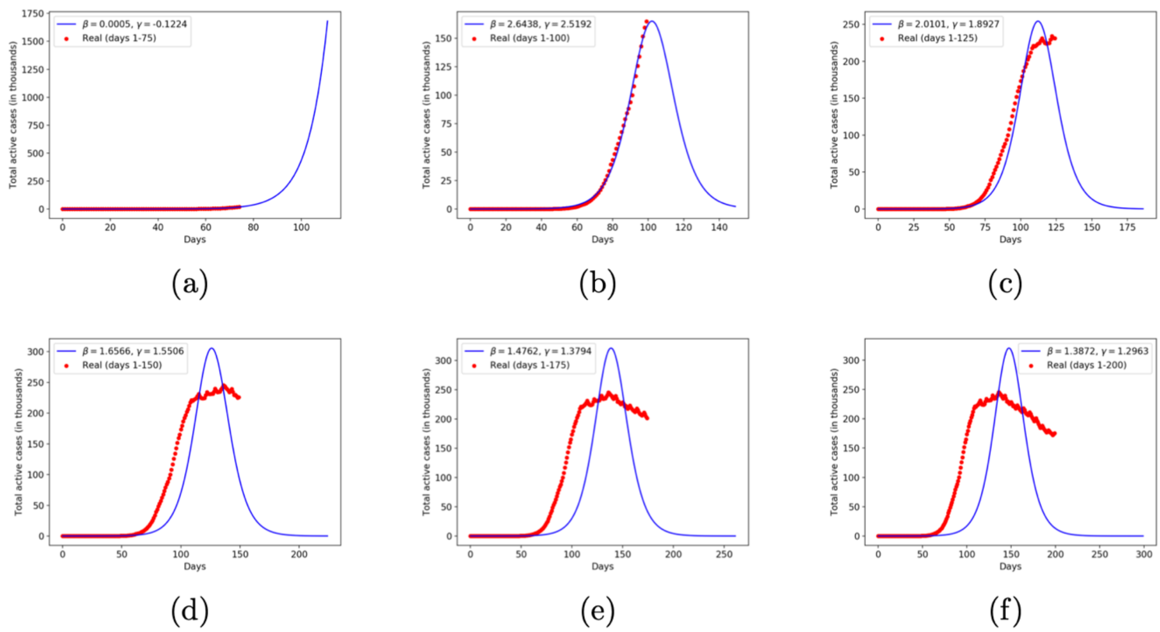

Fig. 7. SIR model $I_t^{(4)}$ calibrated to data from Russia for $n \in \{75,100,125,150,175,200\}$.

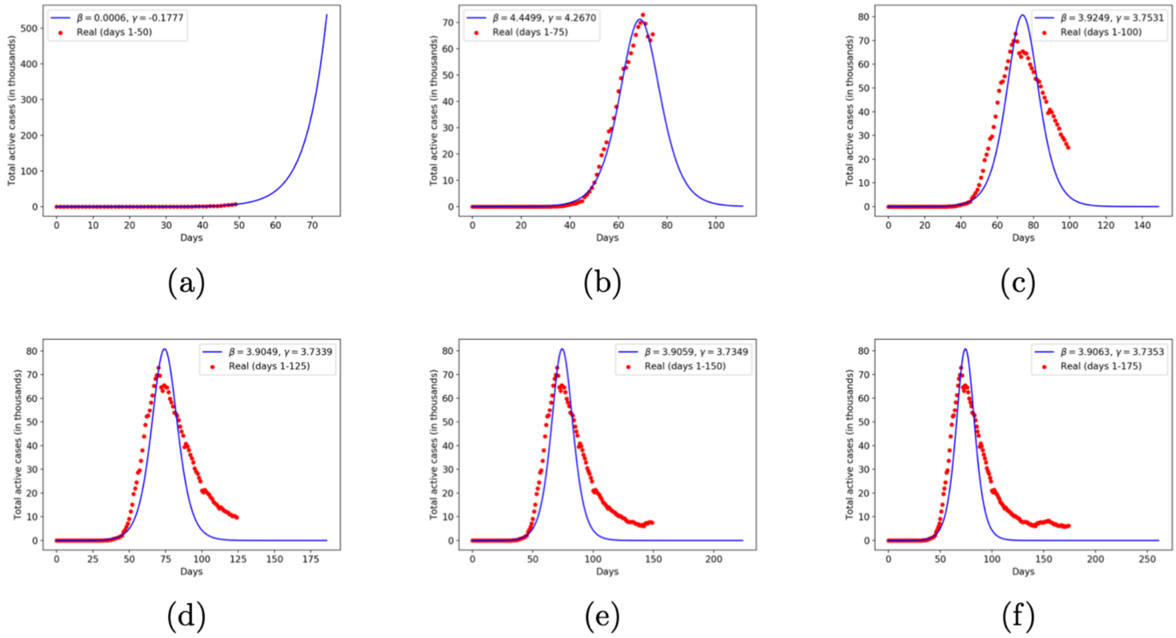

Fig. 8. SIR model $I_t^{(4)}$ calibrated to data from Germany for $n \in \{50,75,100,125,150,175\}$.

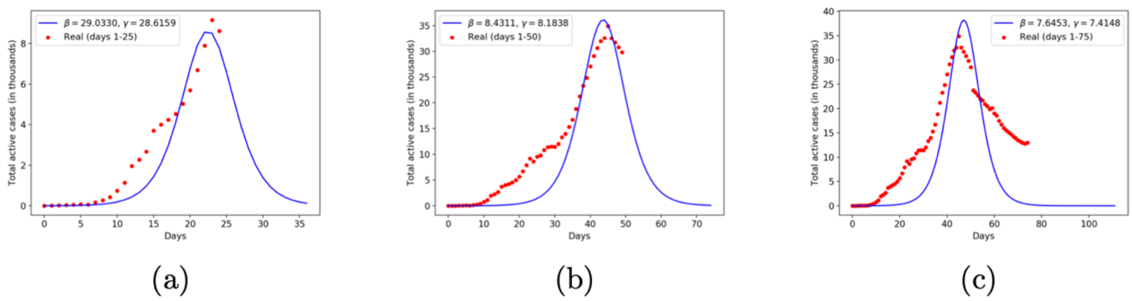

Fig. 9. SIR model $I_t^{(4)}$ calibrated to data from Iran for $n \in \{25,50,75\}$.

A well-known extension to the SIR model is the SEIR model which incorporates an additional population compartment, namely 'E' or 'Exposed', that represents individuals who are already infected by the virus but yet to be able to transmit it (the time between known as the *latent period*). However, one frequently stated challenge with SARS-CoV-2 is that it can transfer from one individual to another even before the incubation period ends, i.e., when symptoms are expected to appear (*negative latency*). Therefore, considering the fact that the incubation period for SARS-CoV-2 is unknown and varies between 1-14 days (with an average of 4-5 days), the transmission period can start even 1-2 days before the end of that unknown incubation period.

How this can affect the performance of the SEIR model, which introduces a distinction between the people who are infected by the virus and those who are able to transmit it, is beyond the scope of this work. The system of quadratic ODEs describing the SEIR model are given by:

$$\mathrm{d}S_t = -\beta S_t\, I_t^{(5)} \mathrm{d}t, \tag{7}$$

$$\mathrm{d}E_t = \left(\beta S_t I_t^{(5)} - \theta E_t\right)\mathrm{d}t, \tag{8}$$

$$\mathrm{d}I_t^{(5)} = \left(\theta E_t - \gamma I_t^{(5)}\right)\mathrm{d}t, \tag{9}$$

$$\mathrm{d}R_t = \gamma I_t^{(5)}\, \mathrm{d}t, \tag{10}$$

where β and γ are the transmission and recovery rates, respectively, and θ is the rate at which already exposed individuals become infected (incubation rate). So, $\frac{1}{\theta}$ can be seen as the average latency, i.e., the mean time between contracting the virus and starting to transmit it. Again, in Figure 10, we present the results for some selected countries. Like SIR, the SEIR model is also able to successfully represent the underlying active cases data unless there is a multi-modal structure, such as a second or third wave of infection.

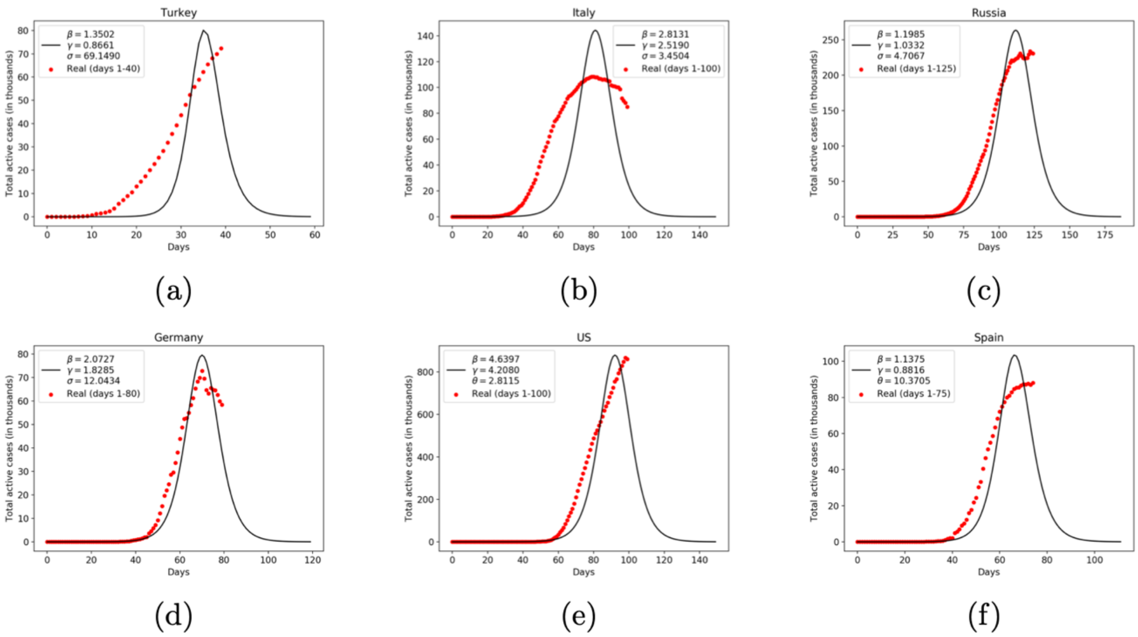

Fig. 10. SEIR model $I_t^{(5)}$ calibrated to various country data for arbitrary values of n.

4. Conclusion

The present study reviewed some of the most widely used epidemic growth models in the literature, namely the logistic, exponential, Gompertz, SIR and SEIR models, and presented their calibration results using real-world epidemic data from selected countries, including Turkey. Among the first three models, which aimed to predict the total number of infected, the Gompertz model seems to have outperformed the logistic and exponential models. SIR and SEIR models, which rather work with active infection data (contrary to cumulative), have also offered a fair representation of the underlying data. All in all, our calibration algorithm was able to return the best possible fit in almost all cases when provided with a target model and underlying data its input.

As an outlook, we'd like to extend the current study into a one which takes into account multiple waves of infection (due to variants, seasonal changes, loosening of measures, etc.). This could be achieved through, *inter alia*, impulsive control model, robust-entropic optimal control, robust-entropic optimal control, bi-modal, or, in general, multi-modal distributions (see, e.g., Pearce et al., 2005; Baltas et al., 2018; Valvo, 2020; Khalilpourazari et al., 2021; Gürbüz et al., 2021). Furthermore, the next trend in epidemic data science could be the prediction models that focus on early detection of pandemics or epidemics. Machine learning (ML) tools can be used to accommodate more complex relationships between many interacting factors (policies, social measures, travel restrictions, vaccine development, mutations, etc.) and the growth dynamics of outbreaks. In this regard, one extension could be to consider a model which integrates future mutation scenarios into predictions on the growth pace of epidemic. Future research can also consider an ensemble of models and an optimal switching rule to obtain better predictive performance than could not be achieved with any of the constituent models per se.

Contribution of Researchers

Nadi Serhan Aydın carried out model calibrations and data analysis. Erfan Babaee Tirkolaee reviewed the literature and contributed to the interpretation of model results.

Conflicts of Interest

The authors declared that there is no conflict of interest.

References

Acar, A.C., Er, A.G., Burduroğlu, H. C., Sülkü, S. N., Aydin Son, Y., Akin, L. and Ünal, S. (2021). Projecting the course of COVID-19 in Turkey: A probabilistic modeling approach. *Turkish Journal of Medical Sciences*, 51(1):16-27, DOI: https://doi.org/10.3906/sag-2005-378.

Baldemir, H., Akın, A. and Akın, Ö. (2020). Fuzzy modelling of COVID-19 in Turkey and some countries in the world. *Turkish Journal of Mathematics and Computer Science*, 12(2):136-150, DOI: https://doi.org/10.47000/tjmcs.751730.

Baltas, I., Xepapadeas, A. and Yannacopoulos, A. N. (2018). Robust portfolio decisions for financial institutions. *Journal of Dynamics and Games*, 5(2): 61-94, DOI: http://dx.doi.org/10.3934/jdg.2018006.

Carcione, J.M, Santos, J.E, Bagaini C. and Ba J. (2020). A simulation of a COVID-19 epidemic based on a deterministic SEIR model. *Frontiers in Public Health*, 8:230, DOI: https://doi.org/10.3389/fpubh.2020.00230.

Chowell, G., Sattenspiel, L., Bansal, S. and Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66-97, DOI: https://doi.org/10.1016/j.plrev.2016.07.005.

Çakır Z. and Savaş, H. (2020). A mathematical modelling for the COVID-19 pandemic in Iran. *Ortadoğu Tıp Dergisi*, 12(2):206-210, 18:66-97, DOI: https://doi.org/10.21601/ortadogutipdergisi.715612.

Duhon, J., Bragazzi, N. and Kong, J.D. (2021). The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Science of the Total Environment,* 760:144325, 18:66-97, DOI: https://doi.org/10.1016/j.scitotenv.2020.144325.

Eroğlu, Y. (2020). "Forecasting models for COVID-19 cases of Turkey using artificial neural networks and deep learning. *Endüstri Mühendisliği*, 31(3):353-372, 18:66-97, DOI: https://doi.org/10.46465/endustrimuhendisligi.771646.

Gürbüz, B., Mawengkang, H., Husein, I. and Weber, G. W. (2021). Rumour propagation: an operational research approach by computational and information theory. *Central European Journal of Operations Research*, 1-21, DOI: https://doi.org/10.1007/s10100-020-00727-0.

Hamer, W.H. (1906). The Milroy lectures on epidemic diseases in England: The evidence of variability and of persistency of type. *The Lancet*, 167(4305):569-574, DOI: https://doi.org/10.1016/S0140-6736(01)80264-6.

Khalilpourazari, S., Doulabi, H. H., Çiftçioğlu, A. Ö. and Weber, G. W. (2021). Gradient-based grey wolf optimizer with Gaussian walk: Application in modelling and prediction of the COVID-19 pandemic. *Expert Systems with Applications*, 177:114920, DOI: https://doi.org/10.1016/j.eswa.2021.114920.

Li, M., Zhang, Z., Cao, W., Liu, Y., Du, B., Chen, C., Liu, Q., Uddin, M.N., Jiang, S., Chen, C., Zhang, Y. and Wang, X., (2021). Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. *Science of the Total Environment,* 764:142810, 18:66-97, DOI: https://doi.org/10.1016/j.scitotenv.2020.142810.

Liang, K., (2020). Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases*, 82:104306, DOI: https://doi.org/10.1016/j.meegid.2020.104306.

Liu, Z., Magal, P., Seydi, O. and Webb, G., (2020). A COVID-19 epidemic model with latency period. *Infectious Disease Modelling*, 5:323-337, DOI: https://doi.org/10.1016/j.idm.2020.03.003.

Ma, J., (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5:129-141, DOI: https://doi.org/10.1016/j.idm.2019.12.009.

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C. and Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*, 290(1):99-115, DOI: https://doi.org/10.1016/j.ejor.2020.08.001.

Önder, H. (2020). Short-term forecasts of the COVID-19 epidemic in Turkey: March 16–28, 2020. *Black Sea Journal of Health Science*, 3(2):27-30, Available at: https://dergipark.org.tr/tr/pub/bshealthscience/issue/51721/710.

Pearce, C., Kaya, Y. and Belen, S. (2005). Impulsive control of a sequence of rumour processes. *In Continuous Optimization*. Springer, Boston, MA, 99:387-407, DOI: https://doi.org/10.1007/0-387-26771-9_14.

Pirouz, B., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S. and Piro, P. (2020). Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability*, 12(6):2427, DOI: https://doi.org/10.3390/su12062427.

Rath, S., Tripathy, A. and Tripathy, A.R. (2020). Prediction of new active cases of coronavirus disease, (COVID-19) pandemic using multiple linear regression model. *Diabetes and Metabolic Syndrome Clinical Research and Reviews*, 14(5):1467-1474, DOI: https://doi.org/10.1016/j.dsx.2020.07.045.

Ross, R. (1911). *The Prevention of Malaria*. London: John Murray. Available at: https://archive.org/details/pr00eventionofmalarossrich.

Sarkar, K., Khajanchi, S. and Nieto, J.J., (2020). Modeling and forecasting the covid-19 pandemic in India. *Chaos, Solitons and Fractals*, 139:110049, DOI: https://doi.org/10.1016/j.chaos.2020.110049.

Sharma, V.K. and Nigam, U., (2020). Modeling and forecasting of COVID-19 growth curve in India. *Transactions of the Indian National Academy of Engineering*, 5:697-710, DOI: https://doi.org/10.1007/s41403-020-00165-z.

Sun, J., Chen, X., Zhang, Z., Lai, S., Zhao, B., Liu, H., Wang, S., Huan, W., Zhao, R., Ng, M.T.A. and Zheng, Y. (2020). Forecasting the long-term trend of COVID-19 epidemic using a dynamic model", *Scientific Reports*, 10:21122, DOI: https://doi.org/10.1038/s41598-020-78084-w.

Tirkolaee, E. B. and Aydın, N. S. (2021). A sustainable medical waste collection and transportation model for pandemics. *Waste Management and Research*, 39(1_suppl): 34-44, DOI: https://doi.org/10.1177/0734242X211000437.

Tomaskova, H. and Tirkolaee, E. B. (2021). Using a process approach to pandemic planning: a case study. *Applied Sciences*, 11(9):4121, DOI: https://doi.org/10.3390/app11094121.

Tuli, S., Tuli, S., Tuli, R. and Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, 11:100222, DOI: https://doi.org/10.1016/j.iot.2020.100222.

Valvo, P. 2020. A Bimodal Lognormal Distribution Model for the Prediction of COVID-19 Deaths. *Applied Sciences*, 10:8500, DOI: https://doi.org/10.3390/app10238500.

Velasquez, R.M.A. and Lara, J.V.M. (2020). Forecast and evaluation of COVID -19 spreading in USA with reduced-space gaussian process regression. *Chaos, Solitons and Fractals*, 136:109924, DOI: https://doi.org/10.1016/j.chaos.2020.109924.

Wu, K., Darcet, D., Wang, Q. and Sornette, D., (2020). Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynamics*, 101(3):1561-1581, DOI: https://doi.org/10.1007/s11071-020-05862-6.