Karaelmas Fen ve Müh. Derg. 11(1):33-43, 2021

Karaelmas Science and Engineering Journal
Journal home page: https://dergipark.org.tr/tr/pub/karaelmasfen
DOI: 10.7212/karaelmasfen.888269

Research Article
Received / Geliş tarihi : 27.09.2020
Accepted / Kabul tarihi : 19.11.2020

Using Mathematical Modeling to Evaluate the Restricted Course of Coronavirus Pandemic (COVID-19) in Turkey

Türkiye'deki Koronavirüs Pandemisinin (COVID-19) Kısıtlamalı Seyrinin Matematiksel Modelleme ile Değerlendirilmesi

Kaan Yetilmezsoy*, Majid Bahramian, Necdet Cem Ayla

Yildiz Technical University, Faculty of Civil Engineering, Department of Environmental Engineering, Istanbul, Turkey

Abstract

The new type coronavirus pandemic (COVID-19) was first identified in December 2019 in Wuhan, the administrative center and the largest city in the Hubei Province of China. The virus subsequently spread and is today affecting the entire world. As of Saturday April 25, 2020 (the date on which the present study was conducted), there were 2,908,206 recorded cases and 202,501 reported deaths attributable to COVID-19 around the world. After beginning to spread outside China in January 2020, COVID-19 was first identified in Turkey on March 10–11, 2020, and there had been 107,773 recorded cases and 2,706 deaths reported in the country as of April 25, 2020. The present study examines the restricted course of the coronavirus pandemic in Turkey from a mathematical perspective, and reports on a predictive modeling of the pandemic based on in-country data. Considering factors such as the peak period of the pandemic in Turkey, the infection and recovery rates, and the potential number of active cases at the end of the pandemic, logistic models were proposed for the cumulative numbers of positive cases and recovered cases. The model parameters were analyzed using the Nelder–Mead simplex method (direct search) and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) Quasi-Newton cubic line-search algorithms within the MATLAB® software. For the interpretation of the course of infections, the exponential models proposed for the numbers of daily cases and daily deaths were analyzed using the Richardson's extrapolation method and the double-precision Levenberg–Marquardt method via the DataFit® software. In case that the current restricted course of the data shows a similar progression without any extreme situations (violations of rules, a new wave of outbreak, etc.), the number of daily COVID-19 cases in Turkey could be predicted to drop below 1,000 by the end of the first week, and to below 100 by the last week of May 2020. The predicted results indicated that there would be more than 120,000 active cases in our country at the end of the outbreak, and that infections in numerical terms would continue until the last week of July 2020.

Keywords: Coronavirus pandemic (COVID-19), Mathematical modeling, Logistic function, Nonlinear regression analysis, Statistical analysis, DataFit®, MATLAB®.

Öz

Yeni koronavirüs pandemisi (COVID-19), ilk olarak Çin'in Hubei eyaletinin yönetim merkezi ve en büyük şehri olan Wuhan kentinde Aralık 2019'da tespit edilmiş olup, tüm dünyayı etkisi altına almış durumdadır. Bu çalışmanın yapıldığı 25 Nisan 2020 Cumartesi günü itibarıyla dünyada COVID-19 kaynaklı 2,908,206 vaka ve 202,501 ölüm rapor edilmiştir. Ocak 2020'den itibaren Çin dışında da görülmeye başlayan COVID-19 vakaları, ülkemizde (Türkiye'de) ilk kez 10–11 Mart 2020 tarihinde görülmüş ve 25 Nisan 2020 tarihi itibarıyla Türkiye'nin koronavirüs bilançosu 107,773 vaka ve 2,706 ölüm değerleriyle kayıtlara geçmiştir. Bu çalışma kapsamında, Türkiye'deki koronavirüs pandemisi seyrinin matematiksel açıdan irdelenmesi hedeflenmiş ve ülkemiz için elde edilen verilere göre salgının kısıtlamalı sürecinin gidişatı hakkında bir tahmin modellemesi yapılmıştır. Salgının şu ana kadar Türkiye'de pik düzeyde görüldüğü zaman, enfeksiyon ve iyileşme hızları ve salgın sonunda maruz kalınabilecek vaka sayısı gibi faktörler göz önüne alınarak birikimli pozitif vaka ve iyileşme sayıları için lojistik modeller önerilmiştir. Model parametreleri MATLAB® programı vasıtasıyla Nelder–Mead simpleks doğrudan arama ve Broyden–Fletcher–Goldfarb–Shanno (BFGS) Quasi-Newton kübik doğrultu belirleme algoritmaları kullanılarak çözümlenmiştir. Günlük vaka ve günlük ölüm sayısı verileri için önerilen ekponansiyel modeller ise DataFit®

*Corresponding author: yetilmez@yildiz.edu.tr

Kaan Yetilmezsoy orcid.org/0000-0003-1478-9957
Majid Bahramian orcid.org/0000-0002-7571-5567
Necdet Cem Ayla orcid.org/0000-0003-3391-5517

programı vasıtasıyla Richardson ekstrapolasyon metodu ve çift hassasiyetli Levenberg–Marquardt yöntemi kullanılarak çözümlenmiş ve enfeksiyonun gidişatı yorumlanmıştır. Mevcut kısıtlamalı veri seyrinin ekstrem durumlar (kural ihlalleri, yeni salgın dalgası, vb.) olmaksızın benzer şekilde ilerlemesi durumunda, Türkiye'deki COVID-19 kaynaklı günlük vakaların Mayıs 2020'nin ilk haftası itibarıyla 1000'in altına ve son haftası itibarıyla da 100'ün altına inebileceği öngörülmüştür. Tahmin sonuçları, kısıtlamalı sürecin devam etmesi halinde ülkemizdeki salgın sonunda 120,000'in üzerinde maruz kalınmış vaka olabileceğini ve enfeksiyonların sayısal anlamda sona ermesi sürecinin ise Temmuz 2020'nin son haftasını bulabileceğini göstermiştir.

Anahtar Kelimeler: Koronavirüs pandemisi (COVID-19), Matematiksel modelleme, Lojistik fonksiyon, Nonlineer regresyon analizi, İstatistiksel analiz, DataFit®, MATLAB®.

1. Introduction

The first case of the COVID-19 was identified in December 2019 at a seafood and animal market in the city of Wuhan in the Hubei Province of China, which with 1.4 billion residents is the most populous country in the world. COVID-19, dubbed the Wuhan coronavirus by local authorities, can be transmitted from person to person through respiratory droplets. This new viral respiratory disease is reported to manifest with symptoms of fever and shortness of breath, and to spread through droplets and direct contact. It has been established that respiratory droplets landing on surfaces by various routes, such as coughing or sneezing from infected individuals, remain infectious for a given period (Rothe et al. 2020). Studies have found that asymptomatic people with the virus and who are in the incubation period are also contagious, although the incubation period is not the period in which the virus spreads most efficiently (Zhou et al. 2020). Since the virus infects the respiratory tract, the most efficient transmission occurs through contact of the hands, eyes, mouth, and nose with surfaces contaminated with droplets and respiratory secretions expelled when infected individuals cough or sneeze (Singhal 2020). The World Health Organization (WHO) declared the outbreak a pandemic on March 11, 2020 due to its global scale (Acıbadem Sağlık Grubu 2020).

In the period between January 2020 and the time of writing (April 26, 2020, 02:02), 2,914,093 cases and 202,985 deaths had been reported worldwide (Worldometers 2020). Within the specified period, the countries with total cases numbering above 100,000 (number of cases presented in parentheses) include the United States (956,679), Spain (223,759), Italy (195,351), France (161,488), Germany (156,418), the United Kingdom (148,377) and Turkey (107,773), respectively (Worldometers 2020). The inclusion of Turkey in this ranking indicates the severity of COVID-19 infections in our country.

COVID-19 has a very high transmission and spread rate when compared to previous viral diseases. The outbreak progressed very rapidly, and quickly became a global issue, spurring intensive studies in the subject and highlighting a need to develop decision-making mechanisms based on the course of the infection (Ankaralı et al. 2020). In this context, deterministic or stochastic modeling of infections can be very helpful in understanding their mechanisms and processes, and in the development of preventive or therapeutic strategies (Ankaralı et al. 2020, Godio et al. 2020). For instance, Tuan et al. (2020) employed a Caputo fractional-order derivative method for predicting the behavior of the COVID-19 pandemic. Using the reports from around the world from January 22, to April 11, 2020, they concluded that the pandemic still will be active globally. Torrealba-Rodriguez et al. (2020) used Gompertz and logistic mathematical methods to forecast the number of the positive COVID-19 cases in Mexico, from February 27, 2020 to May 08, 2020. They also employed artificial neural networks to see the performance of the soft computing method. The predicted number of cases by Gompertz, logistic, and inverse artificial neural network (ANN) models were 47,576, 42,131, and 44,245 cases, respectively. The authors also predicted the total number of cases until the end of pandemic. They predicted 469,917 cases for Gompertz model, 59,470 cases for the logistic model, and 70,714 cases for the inverse ANN model, respectively. In another study, Zhang (2020) employed singular and nonsingular kernels for a mathematical modelling of the possibility of the COVID-19 spread within a given general population. Avila-Ponce de León et al. (2020) used the Susceptible-Exposed-Infected-Asymptomatic-Recovered-Dead (SEIARD) epidemic model mathematical analysis and state-level forecast of the COVID-19 pandemic in Mexico until the end of November 2020. In a similar modeling study, Ivorra et al. (2020) employed a Susceptible-Exposed-Infectious-Hospitalized Recovered-Dead (θ-SEIHRD) model (θ represents the fraction of infected people that has been detected) for predicting the spread of the coronavirus disease. Using the data from the undetected infections in China, they were able to adapt the model to the official reports from WHO and predict the spread of the disease in China. Furthermore, Kucharski et al. (2020) reported that the transmission and control of the COVID-19 can be represented mathematically. Using the officially reported data from January 2020 to February 2020 as the input for a stochastic model, they reported that there is a more than 50% chance of the infection in locations with similar transmission potential to Wuhan in early January once there are at least four independently introduced cases.

In this study, a prediction modeling was conducted for our country undergoing a difficult COVID-19 pandemic period on the basis of data reported for 47-day period between March 10, 2020 and April 25, 2020 (records for restricted period before June 01, 2020 normalization process). Thus, it is aimed to obtain the prediction of "when our country may have ended" this epidemic if the COVID-19 measures are strictly followed in Turkey and to interpret the numerical findings for Turkey. In this context, it is aimed to show the importance of the loss of process caused by the imprudence related to the epidemic after the normalization calendar (June 01 2020) in our country. Moreover in process modeling performed in the study for progress of restricted period research is based on development of new mathematical formulations in connection with the cumulative number of positive cases by days, the number of infected patients per day from the outset, the cumulative number of recovered patients, and the number of deaths per day from the outset, use of different search algorithms in analysis of these issues, and validation by testing by means of many statistical performance indicators.

2. Materials and Methods

2.1. Data Collection and Criteria

The present study is based on the data of a sampling period of 47 days (restricted process before the normalization circular dated 01 June 2020) between March 10, 2020 and April 25, 2020 (days 69 and 116 of the COVID-19 pandemic based on January 01, 2020 start date) related to the outbreak in Turkey. The assessment criteria included (*i*) the cumulative number of positive cases by days (total cases), (*ii*) the number of infected patients per day from the outset (daily new cases), (*iii*) the cumulative number of recovered patients by days (total recoveries), and (*iv*) the number of deaths per day from the outset (daily deaths).

2.2. Computational Procedure

Modeling studies were conducted by using MATLAB® R2018a (V9.3.0.713579, 64-bit (win64), Academic License Number: 40578168, MathWorks Inc., Natick, MA) and DataFit® (version 8.1.69, Oakdale Engineering, PA, USA, RC549) softwares running on a Casper Excalibur (Intel® Core™ i7-7700HQ CPU, 2.81 GHz, 16 GB of RAM, 64-bit) PC.

The parameters of the logistic functions proposed for the modeling of the cumulative numbers of positive cases and recoveries by days until April 25, 2020 for Turkey were determined using the Nelder–Mead simplex method (direct search) algorithm (N-M SDSA, proposed by John Ashworth Nelder and Roger Mead in 1965) and the BFGS (Broyden–Fletcher–Goldfarb–Shanno) Quasi-Newton cubic line-search algorithm (BFGS Q-N LSA, introduced by William Cooper Davidon in 1959) within the MATLAB® numeric computing environment (Çetinkaya and Yetilmezsoy 2019).

The collected data associated with the restricted course of the COVID-19 pandemic in Turkey (between 10 March 2020 and 25 April 2020) were transported directly from Microsoft® Excel, which was used as an open database connectivity data source, and then the nonlinear regression-based analysis was implemented. As the regression models were solved, they were automatically sorted according to the goodness-of-fit criteria within the framework of the DataFit® interface. The nonlinear convergence parameters of the exponential functions proposed for the modeling of the daily new cases and the daily deaths were determined using the DataFit® program, with calculation options selected as follows: regression tolerance = 1×10^{-10}, maximum number of iterations = 250 and nonlinear iteration limit = 10. The regression analysis was performed using the Richardson's extrapolation method and the double-precision Levenberg-Marquardt method (Abdul-Wahab et al. 2020, Yetilmezsoy et al. 2020, Yetilmezsoy 2016).

2.3 Statistical Analysis

The predictive performance of the models and the estimate errors were calculated using various statistical performance indicators (determination coefficient (R^2), adjusted coefficient of multiple determination (R^2_{adj}), mean absolute error (MAE), root mean squared error (RMSE), root mean squared error - systematic ($RMSE_S$), root mean squared error - unsystematic ($RMSE_U$), proportion of systematic error (PSE), index of agreement (IA), fractional variance (FV), coefficient of variation of RMSE (CV(RMSE)), standard error of the estimate (SEE), sum of residuals (SR), residual average (RA), and Fisher's *F*-statistic) reported in the literature (Abdul-Wahab et al. 2020, Yetilmezsoy et al. 2020, Foroughi et al. 2018, Yetilmezsoy et al. 2018, Yetilmezsoy 2016, Yetilmezsoy et al., 2009). In the statistical analysis, an alpha (α) level of 0.05 (or 95% confidence) was considered to test the significance of the derived functions.

3. Results

3.1. Prediction Models and Quantitative Evaluations on Infection Process

According to the 47-day data from March 10, 2020 to April 25, 2020 for the restricted course of the COVID-19 pandemic in Turkey, the derived prediction models for the cumulative number of positive cases by days, the number of infected patients per day from the outset, the cumulative number of recovered patients, and the number of deaths per day from the outset are expressed, respectively, in Equations (1) and (3) (logistic functions), and Equations (2) and (4) (exponential functions).

$$f(t,a,b,c) = \frac{c}{1+\exp\left[\frac{-(t-b)}{a}\right]} = \frac{120120.7108}{1+\exp\left[\frac{-(t-103.8380)}{6.0630}\right]} \quad (1)$$

$$f(t) = \exp\left[433.6586 - \frac{7863.0701}{t} - 75.2815.\ln(t)\right] \quad (2)$$

$$f(t,a,b,c) = \frac{c}{1+\exp\left[\frac{-(t-b)}{a}\right]} = \frac{115499.9998}{1+\exp\left[\frac{-(t-123.7692)}{6.1093}\right]} \quad (3)$$

$$f(t) = \exp\left[319.8079 - \frac{6060.3337}{t} - 55.2989\cdot\ln(t)\right] \quad (4)$$

The logistic regression has been broadly utilized to portray the growth of a population. An infection can be defined as the growth of the population of pathogenic agents, so that a logistic model seems plausible for the analysis of case-control studies in epidemiological researches. In the logistic formula, t denotes the time (day) and three parameters a, b, and c refer the infection speed, the day with the maximum infections occurred, and the expected number of infected patients at the end of the outbreak, respectively. According to this model, at high time values, the number of infected individuals approaches c, where this point can be considered as the end of the infection. Mathematically, the first derivative of this function begins to decrease at the inflection point b, and therefore, after this peak point, the severity of the epidemic becomes less disruptive and tends to decrease (Malato 2020).

According to Equation (1), the total number of cases in Turkey at the end of the restricted course of the COVID-19 pandemic was predicted to be above 120,000 (R^2 = 0.9990, Standard Error of estimate (SEE) = 628.3215). According to Equation (1), the day of the maximum number of infections was predicted to be April 11–13, 2020 (days 102–104 of the COVID-19 outbreak from the outset), considering the standard deviation values. The estimated values were obtained using the N-M SDSA within the MATLAB® platform, where the optimization problem was solved by employing "fminsearch" built-in code for the multidimensional unconstrained non-linear minimization (number of non-linear iterations (NNI) = 165, function counts (FC) = 306, and elapsed time = 3.364532 seconds). Figure 1 shows the outputs of the logistic model given in Equation (1) based on the data from between March 10, 2020 and April 25, 2020, during the restricted course of the COVID-19 pandemic in Turkey (the cumulative number of positive cases by days).

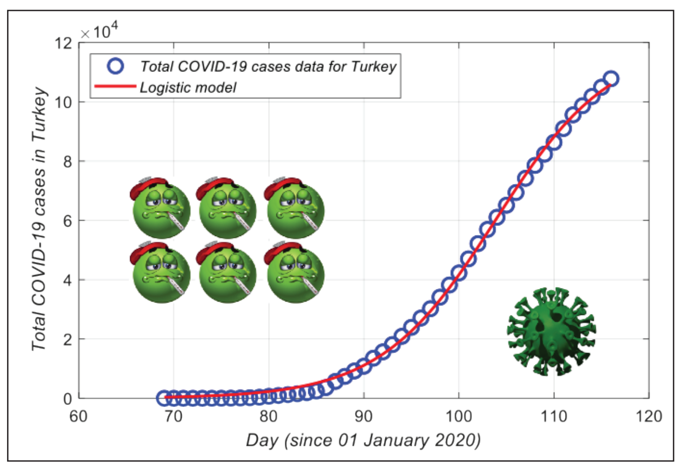

Figure 1. Prediction of the cumulative number of positive cases by days in Turkey during the COVID-19 pandemic (between March 10 2020 and April 25 2020) with a logistic model.

According to Equation (2), the number of daily COVID-19 cases could be predicted to drop below 1,000 by the first week of May 2020 (May 06–10, 2020) and below 100 by the last week of May 2020 (May 24–28, 2020) (R^2 = 0.9652, SEE = 343.3034). The number of recovered patients per day was calculated using the cumulative prediction results garnered from Equation (3), and the highest number of recoveries was predicted to occur at the beginning of May

2020 (R^2 = 0.9973, SEE = 121.2619). Figure 2 depicts the outputs of the exponential model (the number of infected patients per day from the outset) given in Equation (2) and the calculated daily recoveries (from Equation (3)) based on the data from between March 10, 2020 and April 25, 2020, during the restricted course of the COVID-19 pandemic in Turkey. Additionally, for the aim of fulfilling both computational and educational purposes of this study, a sample MATLAB® code, which is written to visualize the time-courses of daily new COVID-19 cases and daily recoveries in Turkey, is depicted in Figure 3.

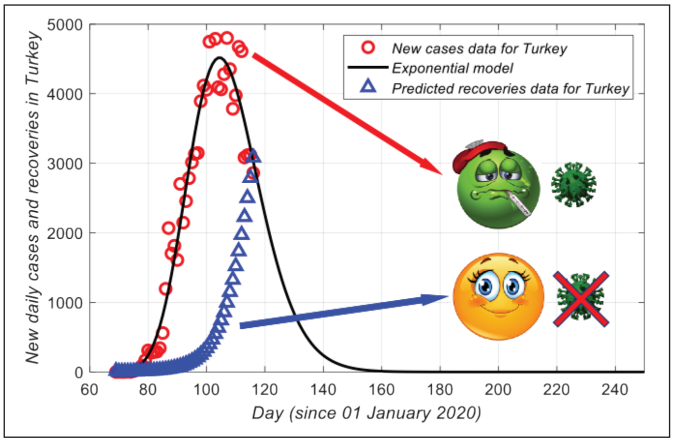

Figure 2. Prediction of the number of infected patients per day and the number of recovered patients per day from the outset during the COVID-19 pandemic in Turkey (between March 10, 2020 and April 25, 2020).

```
clear, clc
load 'dailycase';
load 'dailycoverycalc';
y1 = dailycase(:,2);
y2 = dailycoverycalc(:,2);
y22 = y2([1:48]);

x = [69:116]';
x1 = [69:250]';

a1 = 433.65862239;
b1 = -7863.07008407;
c1 = -75.28151528;

fcase = exp(a1+b1./x1+c1.*log(x1));

plot(x,y1,'or','MarkerSize',8,'LineWidth',2)
hold on
plot(x1,fcase,'-k','LineWidth',2)
plot(x,y22,'^b','MarkerSize',8,'LineWidth',2)
axis([60 250 0 5000]);
xlabel('\itDay (since 01 January 2020)')
ylabel('\itNew daily cases and recoveries in Turkey')
legend('\itNew cases data for Turkey','\itExponential model',...
    '\itPredicted recoveries data for Turkey','Location','NorthEast')
set(gca,'FontSize',12)
grid on
```

Introduction of the collected COVID-19 data (daily new cases and calculated daily recoveries) for Turkey

Definition of variable limits, model parameters, and model structure for the exponential function (Equation (2))

Setting of the plotting properties and the axis limits

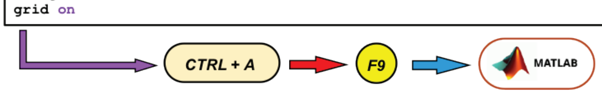

Figure 3. A sample MATLAB® script created for the visualization of daily new COVID-19 cases (obtained from Equation (2)) and daily recoveries (calculated from Equation (3)).

The number of recovered patients per day was calculated using the cumulative prediction results garnered from Equation (3), and the highest number of recovered patients was predicted to occur at the beginning of May 2020 using the BFGS Q-N LSA in the M-file Editor within the framework of MATLAB® software. The optimization problem was solved by employing "fminunc" built-in code to find the local minimum of the error function (number of non-linear iterations (NNI) = 18, function counts (FC) = 76, and elapsed time = 1.839137 seconds). According to Equation (3), the total number of recovered patients in Turkey at the end of the COVID-19 pandemic was predicted to be above 115,000 (R^2 = 0.9973, SEE = 121.2619). Figure 4 illustrates the outputs of the logistic model given in Equation (3) based on the data from between March 10, 2020 and April 25, 2020, during the COVID-19 epidemic in Turkey (the cumulative number of recovered patients by days).

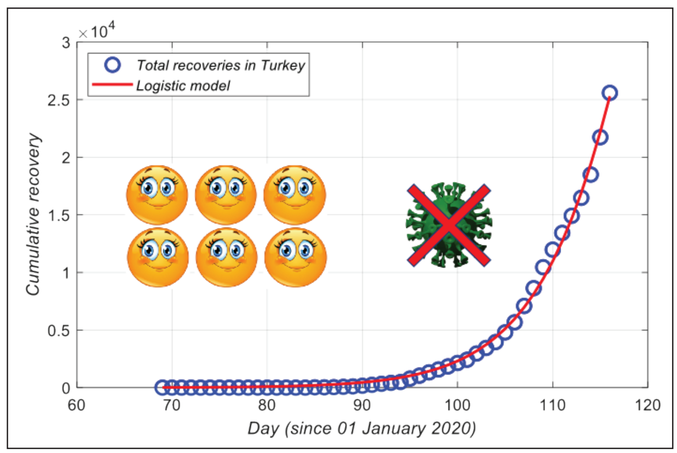

Figure 4. Prediction of the cumulative number of recovered patients by days in Turkey during the COVID-19 pandemic (between March 10, 2020 and April 25, 2020) with a logistic model.

According to Equation (4), the daily deaths attributable to the COVID-19 coronavirus in Turkey could be predicted to decrease significantly at the beginning of June 2020 (R^2 = 0.9811, SEE = 6.8250), and the COVID-19 pandemic in numerical terms could be estimated to be over by the end of July 2020 (July 29–31, 2020) at around day 213 of the outbreak from the outset. Figure 5 delineates the outputs of the exponential model given in Equation (4) based on the data from between March 10, 2020 and April 25, 2020, during the restricted course of the COVID-19 pandemic in Turkey (the number of deaths per day from the outset).

3.2 Statistical Appraisal of Models' Performance in Forecasting the COVID-19 pandemic

Descriptive statistical indicators are utilized as helpful mathematical tools to evaluate the prediction accuracy of proposed models as well as the quantity of the estimation error. For the derived exponential models (Equation (2) for daily new cases and Equation (4) for daily deaths), regression variable results including model coefficients (*a*, *b*, *c*), standard errors (SE), *t*-statistics, and the corresponding *p*-values are summarized in Table 1.

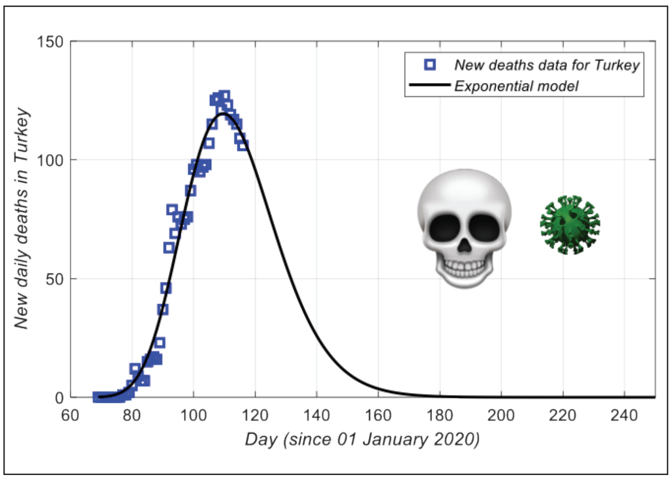

Figure 5. Prediction of the number of deaths per day from the outset during the COVID-19 pandemic in Turkey (between March 10, 2020 and April 25 2020) based on an exponential model.

While larger absolute *t*-ratio value shows the more significant parameter in the regression model, the variable with the lowest *p*-value will denote the most significant one (Yetilmezsoy et al. 2020, Yetilmezsoy 2016). Considering the magnitudes of the absolute *t*-ratios along with the respective *p*-values ($p < 0.050$) of the regression coefficients and variance analysis-based values within a confidence level of 95% (Table 1), the exponential models showed significant fits ($F = 623.4014 > F_{\alpha,df,(n-df+1)} = F_{0.05,2,45} = F_{tabulated} = 3.2043$, $p = 0.0000 < \alpha = 0.05$ for Equation (2), and $F = 1169.1143 > F_{tabulated}$, $p = 0.0000 < \alpha = 0.05$ for Equation (4)) to the COVID-19 related data (collected between 10 March 2020 and 25 April 2020) in the estimation of the number of infected patients per day from the outset (daily new cases) and the number of deaths per day from the outset (daily deaths) in Turkey. The statistical evaluators considered for the appraisal of the estimation performance of the proposed exponential models are presented in Table 2.

According to the statistical results in Table 2, the proposed exponential models demonstrated satisfactory correlations ($R^2 = 0.9652$ for Equation (2) and $R^2 = 0.9811$ for Equation (4)) in the prediction of daily new cases and daily deaths associated with the COVID-19 pandemic in Turkey. These results revealed that only 3.48% and 1.89% of the total

Table 1. Regression variable results for the derived exponential models (Equations (2) and (4))

Regression coefficients	SE	t-ratio	p-value
EM-1	$f(t) = \exp\left[433.6586 - \frac{7863.0701}{t} - 75.2815 \cdot \ln(t)\right]$		
a = 433.6586	33.9017	12.7916	0.0000
b = -7863.0701	613.2810	-12.8213	0.0000
c = -75.2815	6.0280	-12.4887	0.0000
EM-2	$f(t) = \exp\left[319.8079 - \frac{6060.3337}{t} - 55.2989 \cdot \ln(t)\right]$		
a = 319.8079	26.9378	11.8721	0.0000
b = 6060.3337	492.2000	-12.3127	0.0000
c = 55.2989	4.7793	-11.5705	0.0000
Fisher's *F*-statistic	$F = \frac{MSS_{reg}}{MSS_{res}} = \frac{(SS_{reg}/df_1)}{(SS_{res}/df_2)} = \frac{(\sum_{i=1}^{n}(P_i - O_m)^2))/(k)}{(\sum_{i=1}^{n}(O_i - P_i)^2)/(n-(k+1))}$		

***SE:** Standard error, **t:** Time (day), **EM-1:** Exponential model-1 (Equation (2)), **EM-2:** Exponential model-2 (Equation (4)), and p-values < 0.050 are considered to be significant.*

*O, P, m, reg, res, and i are the subscripts indicating the observed, predicted, mean, regression, residual, and index of data points, respectively, **SS:** Sum of squares, **MSS:** Mean sum of squares, **df_1:** k and **df_2:** (n–(k+1)) are the degrees of freedom (herein $df_1 = 2$ and $df_2 = 45$), n is the number of data points (herein n = 48), and k is the number of coefficients in the regression model (k = 2).*

Table 2. Statistics of the nonlinear regression-based analysis conducted for the construction of the exponential equations (Equations (2) and (4))

Statistical evaluator	Calculation	Models	
		EM-1	EM-2
Standard error of the estimate	$SEE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n-2}}$	343.3034	6.8250
Sum of residuals	$SR = \sum_{i=1}^{n}(O_i - P_i)$	-219.5354	-20.8101
Residual average	$RA = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)$	-4.5737	-0.4335
Determination coefficient	$R^2 = \frac{\sum_{i=1}^{n}(P_i - O_m)^2}{\sum_{i=1}^{n}(O_i - P_i)^2 + \sum_{i=1}^{n}(P_i - O_m)^2}$	0.9652	0.9811
Adjusted coefficient of multiple determination	$R^2_{adj} = 1 - \left[(1 - R^2)\left(\frac{n-1}{n-p-1}\right)\right]$	0.9636	0.9803

O, P, m, and i are the subscripts indicating the observed (or collected), predicted, mean, and index of data points, respectively, n is the number of observations, p is the total number of explanatory variables in the regression model (without including the constant term), EM-1 = Exponential model-1 (Equation (2)), and EM-2 = Exponential model-2 (Equation (4)).

variations were not described by the proposed exponential models (for Equations (2) and (4), respectively). Moreover, the high magnitude of the adjusted determination coefficients (R^2_{adj} = 0.9636 for Equation (2) and R^2_{adj} = 0.9803 for Equation (4)) corroborated the importance of the derived exponential models. Additionally, it is worth noting that if the model includes so many parameters, and the sample size is not that much big the R^2_{adj} may be significantly lower than the R^2 value (Abdul-Wahab et al. 2020, Yetilmezsoy et al. 2020, Yetilmezsoy et al. 2009). For the proposed exponential formulations, R_{adj}^2 and R^2 values were very close to each other, confirming that the size of the sample (n = 48) used in the present study was appropriate for the modeling purpose.

Finally, various statistical indicators (e.g. R^2, SEE, MAE, RMSE, $RMSE_S$, $RMSE_U$, PSE, IA, FV, CV(RMSE)) were computed to assess the prediction performance of the proposed logistic functions (Equations (1) and (3)). Summary of the statistical analysis implemented for the assessment of the predictive performance of the proposed logistic models is tabulated in Table 3.

Based on the statistical values presented in Table 3, the proposed logistic functions showed quite strong correlations (R^2 = 0.9990 for Equation (1) and R^2 = 0.9973 for Equation (3)) in the predictions of the cumulative number of positive cases by days (total cases) and the cumulative number of recovered patients by days (total recoveries) related to the restricted course of the COVID-19 pandemic in Turkey. The relevant statistics corroborated that only 0.10% and 0.27% of the total variations were not accounted by the proposed logistic functions (for Equations (1) and (3), respectively), revealing that the derived formulations satisfactorily estimated the expected targets with considerably small deviations. Additionally, IA values (0.9997 and 0.9992 for Equations (1) and (3), respectively) were found to be very close to 1, indicating that very good agreements were attained between the collected COVID-19 data (the cumulative number of positive cases by days and the cumulative number of recovered patients by days) and outputs of the logistic models (for Equations (1) and (3), respectively) for Turkey from March 10, 2020 to April 25, 2020. Moreover, the low values of the coefficient of variation (3.8772% and 9.1605% for Equations (1) and (3), respectively) suggested a high degree of accuracy and a good deal of the reliability of the proposed logistic equations, as reported in the previous studies (Abdul-Wahab et al. 2020, Yetilmezsoy et al. 2020, Yetilmezsoy et al. 2009).

As seen from Table 3, descriptive statistical performance indicators, such as PSE (0.3010 and 0.1311 for Equations (1) and (3), respectively) and FV (0.0117 and 0.0100 for Equations (1) and (3), respectively), also indicated that the derived logistic regressions yielded fairly small residuals and showed a remarkable estimation performance on the prediction of the cumulative number of positive cases by

Table 3. Statistical performance of the logistic models (Equations (1) and (3)) used for the predictions of the cumulative number of positive cases by days (total cases) and the cumulative number of recovered patients by days (total recoveries).

Statistical performance indicator	Calculation	Models	
		LM-1	LM-2
Determination coefficient	$R^2 = \dfrac{\sum_i^n (P_i - O_m)^2}{\sum_i^n (O_i - P_i)^2 + \sum_i^n (P_i - O_m)^2}$	0.9990	0.9973
Standard error of the estimate	$SEE = \sqrt{\dfrac{\sum_{i=1}^n (O_i - P_i)^2}{n-2}}$	628.3215	121.2619
Mean absolute error	$MAE = \dfrac{1}{n}\sum_{i=1}^n \lvert P_i - O_i \rvert$	1123.2016	255.5167
Root mean squared error	$RMSE = \left(\dfrac{1}{n}\sum_{i=1}^n [P_i - O_i]^2\right)^{0.5}$	1278.7545	348.6285
Root mean squared error - systematic	$RMSE_s = \left(\dfrac{1}{n}\sum_{i=1}^n [P_{reg} - O_i]^2\right)^{0.5}$	615.0922	118.7088
Root mean squared error - unsystematic	$RMSE_U = \left(\dfrac{1}{n}\sum_{i=1}^n [P_{reg} - P_i]^2\right)^{0.5}$	1121.1042	327.7958
Proportion of systematic error	$PSE = (RMSE_s)^2 / (RMSE_u)^2$	0.3010	0.1311
Index of agreement	$IA = 1 - \left[\dfrac{\frac{1}{n}\sum_{i=1}^n (P_i - O_i)^2}{\frac{1}{n}\sum_{i=1}^n (\lvert P_i - O_m \rvert + \lvert O_i - O_m \rvert)^2}\right]$	0.9997	0.9992
Fractional variance	$FV = \dfrac{2(\sigma_o - \sigma_p)}{(\sigma_o + \sigma_p)}$ $\sigma_o = \sqrt{\dfrac{\sum_{i=1}^n (O_i - O_m)^2}{n-1}}$ $\sigma_p = \sqrt{\dfrac{\sum_{i=1}^n (P_i - P_m)^2}{n-1}}$	0.0117	0.0100
Coefficient of variation of RMSE	$CV(RMSE) = \left(\dfrac{RMSE}{O_m}\right) \times 100$	3.8772	9.1605

O, P, m, reg, and i are the subscripts indicating the observed, predicted, mean, regression, and index of data points, respectively, σ is the standard deviation, n is the number of data points, LM-1 = Logistic model-1 (Equation (1)), and LM-2 = Logistic model-2 (Equation (3)).

days (total cases) and the cumulative number of recovered patients by days (total recoveries). Furthermore, considering the nonlinear nature of the COVID-19 outbreak-based data, values of SEE, MAE, RMSE, $RMSE_S$, and $RMSE_U$ obtained from the proposed logistic functions were reasonably low, and their maximum value did not exceed 0.0154 per thousand of Turkey's total population (83,154,997 million people according to the records on December 31 2019).

3.3. Validation of the Model Output

In this study, under the light of data from the 47-day period between March 10, 2020 and April 25, 2020, a predictive modeling was performed for the cumulative number of positive cases by days, the number of infected patients per day from the outset, the cumulative number of recovered patients, and the number of deaths per day from the outset. Predictions in the scope of the study are forecasts based on the continuation of restrictions prior to the normalization calendar (01 June 2020). Therefore, to evaluate the consistency of the model outputs, a validation study was carried out including the records between March 10 2020 and April 25 2020. In this context, a set of log data obtained for 85 determination coefficient (R^2) values for logistic model-1 (Equation (1)) for "the cumulative number of positive cases by days" was determined as 0.9719, for exponential model-1 (Equation (2)) "the number of infected patients per day from the outset" was calculated as 0.8925, logistic model-2 (Equation (3)) for "the cumulative number of recovered patients" was computed as 0.9991 and exponential model-2 (Equation (4)) for "the number of deaths per day from the outset" was determined as 0.9533. Additionally, each criterion (including the cumulative number of positive cases by days, the number of infected patients per day from the outset, the cumulative number of recovered patients by days, and the number of deaths per day from the outset) was statistically evaluated in terms of the non-parametric Mann–Whitney (M–W) U (or the Wilcoxon rank-sum) test and the Kruskal–Wallis (K–W) test with the Conover–Inman method. Results were appraised with two-tailed p values to reflect the statistical significance between paired groups (α = 0.05 or 95% confidence). Both the M–W U test and the K–W test revealed that there was no statistically significant difference between the predictions of the logistic model-1 (Equation (1)) and the corresponding observed data ($p_{M\text{–}W}$ = 0.1725 and $p_{K\text{–}W}$ = 0.1733). Therefore, the null hypothesis (H_0) was not rejected in favor of the alternative hypothesis (H_a), because the p value was higher than the chosen significance level of 0.05 (or 95% confidence). Similarly, for the 85-day sampling period between March 10, 2020 and May 31, 2020, the non-parametric tests indicated that there was insufficient evidence for a noticeable difference between the predictions of exponential model-1 (Equation (2)) and the officially reported data ($p_{M\text{–}W}$ = 0.1138 and $p_{K\text{–}W}$ = 0.1141). Moreover, the non-parametric tests demonstrated that no sufficient evidence was found for a significant difference between the estimations of the logistic model-2 (Equation (3)) and the respective recorded values ($p_{M\text{–}W}$ = 0.7582 and $p_{K\text{–}W}$ = 0.7593). Likewise, the non-parametric tests concluded that no sufficient evidence was generated for a significant difference between the predictions of the exponential model-2 (Equation (4)) and the corresponding official records ($p_{M\text{–}W}$ = 0.8515 and $p_{K\text{–}W}$ = 0.8522). The above noted statistical results obtained from the non-parametric analysis confirmed with 95% certainty that the proposed models satisfactorily described the restricted course (for data records before the normalization circular dated June 01, 2020) of the COVID-19 pandemic in Turkey.

4. Discussion

The present estimations are in line with the reported values in literature. For instance, Djilali and Ghanbari (2020) reported that the COVID-19 epidemic would disappear in Turkey by the end of July (after 97 days starting from the date April 15 2020), which indicates a good accordance with our study (probably around July 29–31 2020). Moreover, Kırbaş et al. (2020) predicted the number of total cases in six different European countries including Turkey. Based on the 55-day data (between March 12 2020 and May 03 2020) taken from official website of European Centre for Disease and Control (ECDC), they reported that the predicted number of the total COVID-19 cases for the Turkey was approximately 143,503, which is in a good agreement with the predicted values in our study (over 120,000 cases). Furthermore, Aslan et al. (2020) proposed a susceptible-exposed-infectious-quarantine-recovered (SEIQR) type deterministic model (coupled with the system of ordinary differential equations (ODEs)) to evaluate the dynamics of the COVID-19 outbreak in Hubei (China) and Turkey and to emphasize the crucial roles of testing, quarantine, and social distancing in the control of this pandemic. They reported that Turkey would have about 148,100 total cases if the number of individuals in quarantine and the number of COVID-19 tests were increased by the authorities. The authors also stated that the outbreak would almost finish at the end of July 2020 depending on the change in quarantine rate and the rate of reported cases in Turkey. The results obtained within the scope of the investigation indicate that there is a good agreement between the current study and the estimates of SEIQR-based study.

The predictions of the present study do not take into account unexpected situations (such as failure to comply with curfews and the need to wear of masks/gloves, violations of social distancing rules, and the occurrence of a new wave of cases) other than the current restricted course (restricted process before the normalization circular dated June 01 2020), and are predictions based on the continuation of the data (the results for the 47-day sampling period between March 10 2020 and April 25 2020) under the hygiene measures applied to counter COVID-19 in the country. The current restricted course would be negatively affected, and predictions cannot be expected to be consistent in extreme circumstances, such as travel and religious services, social distancing violations, uncontrolled crowds at religious sites and shrines, especially during Eid al-Fitr (Ramadan Festival) on May 24–26 2020, as well as spontaneous gatherings following the weekend curfews.

Finally, it should be noted that the modeling with the limited epidemiological data is a fairly laborious task for forecasting and tracking the subsequent course (particularly normalization process) of COVID-19 pandemic. In any case, it is as well early to draw an authoritative conclusion due to the dynamics in accessible information, current habits of individuals in the community, and their reactions to nationwide preventive measures or restrictions. It is seen that as the number of days in which epidemic data is collected increases, the forecast performance of the models will also increase. However, updating and calibrating the model parameters according to the course of the epidemic will contribute to the process to be formulated more successfully in deterministic terms.

5. Acknowledgments

The computational analysis implemented in this study has been financially supported by Turkish Academy of Sciences (TÜBA) as a part of Prof. Dr. Kaan Yetilmezsoy's "The Outstanding Young Scientist Award (TÜBA-GEBİP)" of the year 2018.

6. References

Abdul-Wahab, SA., Omer, ASM., Yetilmezsoy, K., Bahramian, M. 2020. Modelling the clogging of gas turbine filter houses in heavy-duty power generation systems. *Math. Comput. Model. Dyn. Syst.* 26 (2): 119-143. Doi: 10.1080/13873954.2020.1713821.

Acıbadem Sağlık Grubu. 2020. Corona Virüsü (Koronavirüs) Nedir, Belirtileri Nelerdir? https://www.acibadem.com.tr/koronavirus/koronavirus-corona-virusu-nedir-belirtileri/

Ankaralı H., Ankaralı, S., Erarslan N. 2020. COVID-19, SARS-CoV2, Enfeksiyonu: Güncel epidemiyolojik analiz ve hastalık seyrinin modellemesi. *Anadolu Kliniği Tıp Bilimleri Dergisi* 25 (1): 1-22. Doi: 10.21673/anadoluklin.707038.

Aslan, H. I., Demir, M., Wise, MM., Lenhart, S. (2020). Modeling COVID-19: Forecasting and analyzing the dynamics of the outbreak in Hubei and Turkey. *medRxiv*. Doi: 10.1101/2020.04.11.20061952.

Avila-Ponce de León, U., Pérez, ÁGC., Avila-Vales, E., 2020. An SEIARD epidemic model for COVID-19 in Mexico: Mathematical analysis and state-level forecast. *Chaos Solitons Fract.* 140: 110165. Doi: 10.1016/j.chaos.2020.110165.

Çetinkaya, AY., Yetilmezsoy, K. 2019. Evaluation of anaerobic biodegradability potential and comparative kinetics of different agro-industrial substrates using a new hybrid computational coding scheme. *J. Clean. Prod.* 238: 117921. Doi: 10.1016/j.jclepro.2019.117921.

Djilali, S., Ghanbari, B., 2020. Coronavirus pandemic: A predictive analysis of the peak outbreak epidemic in South Africa, Turkey, and Brazil. *Chaos Solitons Fract.* 138: 109971. Doi: 10.1016/j.chaos.2020.109971.

Foroughi, M., Chavoshi, S., Bagheri, M., Yetilmezsoy, K., Samadi, MT. 2018. Alum-based sludge (AbS) recycling for turbidity removal in drinking water treatment: an insight into statistical, technical, and health-related standpoints. *J. Mater. Cycles Waste Manag.* 20(4): 1999-2017. Doi: 10.1007/s10163-018-0746-1.

Godio, A., Pace, F., Vergnano, A. 2020. SEIR modeling of the Italian epidemic of SARS-CoV-2. *Preprints* 2020(5): 1-10. Doi: 10.20944/preprints202004.0073.v1.

Ivorra, B., Ferrández, MR., Vela-Pérez, M., Ramos, AM. 2020. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun. Nonlinear Sci. Numer. Simul.* 88: 105303. Doi: 10.1016/j.cnsns.2020.105303.

Kırbaş, İ., Sözen, A., Tuncer, AD., Kazancıoğlu, FŞ. 2020. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fract.* 138. Doi: 10.1016/j.chaos.2020.110015.

Kucharski, AJ., Russell, TW., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, RM., Sun, F., Jit, M., Munday, JD., Davies, N., Gimma, A., van Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, CI., Clifford, S., Quilty, BJ., Bosse, NI., Abbott, S., Klepac, P., Flasche, S. 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* 20: 553-558. Doi: 10.1016/S1473-3099(20)30144-4.

Malato, G., 2020. Covid-19 infection in Italy. Mathematical models and predictions. A comparison of logistic and exponential models applied to Covid-19 virus infection in Italy. https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., Hoelscher, M., 2020. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N. Engl. J. Med. 382: 970-971. Doi: 10.1056/NEJMc2001468.

Singhal, T., 2020. A review of coronavirus disease-2019 (COVID-19). Indian J. Pediatr. 87: 281-286. Doi: 10.1007/s12098-020-03263-6.

Torrealba-Rodriguez, O., Conde-Gutiérrez, RA., Hernández-Javier, AL., 2020. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. *Chaos Solitons Fract.* 138: 109946. Doi: 10.1016/j.chaos.2020.109946.

Tuan, NH., Mohammadi, H., Rezapour, S., 2020. A mathematical model for COVID-19 transmission by using the Caputo fractional derivative. *Chaos Solitons Fract.* 140: 110107. Doi: 10.1016/j.chaos.2020.110107.

Worldometers, 2020. COVID-19 Coronavirus Pandemic, Coronavirus Cases: Turkey. https://www.worldometers.info/coronavirus/#countries

Yetilmezsoy, K. 2016. A new simple model for the prediction of waste sludge flow rate in the steady-state completely mixed activated sludge process. *Environ. Eng. Manag. J.* 15 (12): 2613-2630. Doi: 10.30638/eemj.2016.288.

Yetilmezsoy, K., Demirel, S., Vanderbei, R.J. (2009). Response surface modeling of Pb (II) removal from aqueous solution by *Pistacia vera* L.: Box–Behnken experimental design. *J. Hazard. Mater.* 171(1-3): 551-562. Doi: 10.1016/j.jhazmat.2009.06.035.

Yetilmezsoy, K., Erhuy, CG., Ates, F., Bilgin, MB. 2018. Implementation of fuzzy logic approach to estimate the degree of expulsion and spattering index and weld strength in projection welding. *J. Braz. Soc. Mech. Sci. Eng.* 40(6): 283. Doi: 10.1007/s40430-018-1210-9.

Yetilmezsoy, K., Özçimen, D., Koçer, AT., Bahramian, M., Kıyan, E., Akbin, HM., Goncaloğlu, Bİ. 2020. Removal of Anthraquinone Dye via Struvite: Equilibria, Kinetics, Thermodynamics, Fuzzy Logic Modeling. *Int. J. Environ. Res.* 14 (5): 541-566. Doi: 10.1007/s4174 2-020-00275 -0.

Zhang, Z., 2020. A novel covid-19 mathematical model with fractional derivatives: Singular and nonsingular kernels. *Chaos Solitons Fract.* 139: 110060. Doi: 10.1016/j.chaos.2020.110060.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., Guo, Q., Song, T., He., J., Yen, HL., Peiris, M., Wu, J., 2020. SARS-CoV-2 viral load in upper respiratory specimens of ınfected patients. *N. Engl. J. Med.* 382:1177-1179. Doi: 10.1056/NEJMc2001737.