Predictive approach of COVID-19 propagation via multiple-terms sigmoidal transition model

Abdelbasset Bessadok-Jemai, Abdulrahman A. Al-Rabiah[*]

Chemical Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh, 11421, Saudi Arabia

A B S T R A C T

The COVID-19 pandemic with its new variants has severely affected the whole world socially and economically. This study presents a novel data analysis approach to predict the spread of COVID-19. SIR and logistic models are commonly used to determine the duration at the end of the pandemic. Results show that these well-known models may provide unrealistic predictions for countries that have pandemics spread with multiple peaks and waves. A new prediction approach based on the sigmoidal transition (ST) model provided better estimates than the traditional models. In this study, a multiple-term sigmoidal transition (MTST) model was developed and validated for several countries with multiple peaks and waves. This approach proved to fit the actual data better and allowed the spread of the pandemic to be accurately tracked. The UK, Italy, Saudi Arabia, and Tunisia, which experienced several peaks of COVID-19, were used as case studies. The MTST model was validated for these countries for the data of more than 500 days. The results show that the correlating model provided good fits with regression coefficients (R2) > 0.999. The estimated model parameters were obtained with narrow 95% confidence interval bounds. It has been found that the optimum number of terms to be used in the MTST model corresponds to the highest R^2, the least RMSE, and the narrowest 95% confidence interval having positive bounds.

1. Introduction

The COVID-19 pandemic has devastated health care systems, forced the closure of educational institutions and daily life activities, and plunged the world into crisis. Despite introducing the vaccine COVID-19, the situation remained challenging due to the emergence of several variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The emergence of SARS-CoV-2 strains is a matter of concern. Several new outbreak strains could develop in the future, leading to a severe epidemic resurgence similar to that seen in South Africa (Analytica, 2020; Volz et al., 2021, pp. 2012–2020). Thus, an end to the pandemic would only be possible if successful vaccines against the circulating variants are made available in sufficient quantities worldwide (Fontanet et al., 2021). However, to cope with pandemic situations, health authorities always need a

---

List of symbols

Parameter   Definition Value/Unit

SIR model Eqs. (1)–(4)
I	Number of infected individuals count
N	Overall number of individuals (S + I + R) count
R	Number of recovered individuals count
S	Number of susceptible individuals count
β	Effective contact rate of the disease day^{-1}
γ	Inverse mean infectious period day^{-1}

Logistic model Eq. (5)
Y	Cumulative number of reported cases (C) or deaths (D) count
Y_0	Initial value of reported cases (C) or deaths (D) count
Y_∞	Final projected number of cases (C_∞) or deaths (D_∞) count
d	Number of days
I	Time constant day^{-1}

Sigmoidal Transition model Eqs. (6)-(8)
Y, Y_0, Y_∞	Same as in Logistic model
$d_{1/2}$	Half time period of a given peak days
m	Hill slope exponent dimensionless

statistical estimate of future cases to prepare at a strategic and logistical level, for example, by providing ventilators and intensive care units.

The forecast models for the spread of COVID-19 could be helpful for local authorities to take the necessary measures on time to contain the spread of the pandemic. Therefore, developing reliable and accurate forecasting models to deal with such pandemics is essential. The number of verified COVID-19 infected people worldwide has risen to nearly 464 million, with over 6 million deaths due to the disease, according to Worldometers (2022).

The increasing number of COVID-19 reported cases and the multiple wave reoccurrence worldwide due to the virus variants prompted researchers to emphasize modernizing COVID-19 prediction research by incorporating the latest data updates into more efficient and complicated COVID-19 models. Reliable prediction of COVID-19 disease transmission dynamics, like numerous other scientific concerns related to the disease, is an essential element of the research. The study of epidemic dynamics is a widely used research topic in mathematical modeling and simulation. Since the pandemic began, numerous studies have effectively predicted the final extent and duration of the global pandemic COVID-19 (Barrio, Kaski, Haraldsson, Aspelund, & Govezensky, 2021; Qasim, Ahmad, Yoshida, Gould, & Yasir, 2020; X. Zhou et al., 2020). The indicators are the cumulative daily number of infected persons, deaths, and recovered persons. These parameters can be obtained systematically from official sources. Researchers and health officials use the daily confirmed data to forecast the final numbers of infected populations. In addition, deaths are also recorded and closely monitored as this is the most critical indicator of the pandemic impact parameter. Any country that fails to contain the spread of the pandemic may infect the most vulnerable population and increase the number of deaths. Therefore, the impact of the countermeasures taken by health and government authorities can be indirectly assessed by the number of cases and deaths recorded each day. Thus, methods and models for data analysis are essential to bring the final estimated figures as close as possible to the actual statistics and evaluate the authorities' performance.

Researchers have used several mathematical models to predict the number of infected cases and the duration of epidemic disease (Allen, Brauer, den Driessche, & Wu, 2008; Brauer, Castillo-Chavez, & Feng, 2019; Ndii & Supriatna, 2017; Neves & Guerrero, 2020; Odagaki, 2021; Ramos, Ferrández, Vela-Pérez, Kubik, & Ivorra, 2021; Roberts & Heesterbeek, 2003). For example, susceptible-infected-susceptible (SIS) models were used to formulate nonlinear occurrence rates in paired epidemic assumptions (Kabir, Kuga, & Tanimoto, 2019; Meng, Zhao, Feng, & Zhang, 2016), susceptible-infected-recovered (SIR) models employ epidemic recovery rates (Alshammari & Khan, 2021), susceptible-exposed-infected-recovered (SEIR) type models predict epidemic proliferation to constitute the exposed class (Aba Oud et al., 2021; Li, Meng, & Wang, 2018; Tian & Yuan, 2017). However, the SIR model is the most widely used mathematical model among epidemiologists, ecologists, and sociologists to study the potential for disease pandemics (Arazi & Feigel, 2021; Cooper, Mondal, & Antonopoulos, 2020; Getz, Salter, Muellerklein, Yoon, & Tallam, 2018; Rodrigues, 2016).

Roda et al. showed that the SIR model performs better than the susceptible-exposed-infected-resistant model in representing the information contained in the confirmed-case data (Roda, Varughese, Han, & Li, 2020). This indicates that the

prediction of more complex models can be unreliable compared to a simpler model. (Fanelli & Piazza, 2020) employed the susceptible-infected-recovered-deaths model (SIRD) to predict the peak of COVID-19 in Italy, which was placed around March 21st, 2020. The authors obtained a peak number of infected people of about 26,000 (excluding recovered and dead) and deaths number of about 18,000 at the end of the epidemics. The authors used data corresponding to the period between January 22nd, 2020, and March 15th, 2020 (Fanelli & Piazza, 2020). However, by July 26, 2020, the number of deaths in Italy was 35,112 (Tang et al., 2014). Thus, the SIRD model failed to predict the pandemic output when used at the early stage.

In the past, epidemiology models (e.g., SIR or logistic type models) were often used to forecast the final size and period of the epidemic by fitting model parameters to actual recorded data (Malavika et al., 2021; Malhotra & Kashyap, 2020; Martcheva, 2015). In many cases, the forecasting is inaccurate in size and period (Batista, 2020). The correlating models may provide a good fit indicated by an R2 > 0.99. However, the various model coefficients are estimated with large bounded confidence intervals. Small changes in one of the model parameters can lead to a significant deviation in other parameters. After more than two years, it has become apparent that the global spread of the COVID-19 pandemic is not like any other pandemic. Irregularities and multiple spikes make it difficult to accurately predict the final numbers based on standard epidemiological models.

The available data provide an opportunity to assess the reliability of the predictive models before applying them to data from countries where the epidemic is still in the middle of its course. After such evaluation, one can confidently use such models to predict progression. In addition to the SIR and logistic models, the sigmoid Transition (ST) type model will be explored. The sigmoidal transition (ST) model (also called a Hill equation) has been widely used in biology, pharmacology, and physiology to model dose-response curves (Cairns, Robinson, & Loiselle, 2008; Di Veroli et al., 2015; Gadagkar & Call, 2015; Gesztelyi et al., 2012; Meddings, Scott, & Fick, 1989; Pessoa et al., 2014). In particular, Di Veroli et al. have used the Hill model to propose a modified multi-phasic model to better fit irregularly evolving data in a very similar manner to the present work. Their approach provided a sound approach to fitting dose-response curves with various levels of complexity.

This study proposes a novel approach to accurately predict the spread of the COVID-19 pandemic with multiple peaks and waves. At first, the SIR, logistic, and sigmoidal transition (ST) models will be evaluated for the pandemic data of Italy. A multiple-term sigmoidal transition (MTST) model will be presented and used to predict the final size and period of COVID-19 spread for a selected number of countries that experienced multiple peaks and waves.

2. Methodology

The world reported COVID-19 data revealed multiple peaks, or multiple waves spread worldwide. The irregularities in these various spreads (exhibiting multiple peaks) make any standard epidemiological model unreliable in predicting the final sizes or mortality numbers. This is especially true during ongoing pandemic spreads for authorities to precisely predict the evolution of the pandemic. Modeling the current pandemic requires new insights into how to treat and analyze daily data.

The following subsections describe standard models used to study pandemic spread, e.g., the SIR and the logistic. In addition, the sigmoidal transition type model is introduced and used to forecast COVID-19 cases adequately.

We considered three forecasting models for the analysis and forecasting purposes: SIR, logistic, and ST. The model equations are given in terms of the number of cases (C), or in terms of the number of deaths (D). The model equations are described in the following subsections.

2.1. SIR model

As described by (Batista, 2020), this model can be represented by the following system of differential equations:

$$\frac{dS}{dt} = -\frac{\beta}{N} SI, \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta}{N} SI - \gamma I, \tag{2}$$

$$\frac{dR}{dt} = \gamma I, \tag{3}$$

where $S(t)$ is the number of susceptible individuals but not infected with the disease at time t, $I(t)$ is the number of infected individuals at time t, $R(t)$ is the number of recovered individuals at time t, β is the effective contact rate of the disease, $1/\gamma$ is the mean infectious period.

The parameters and the initial values must be estimated from the actual available data. The overall number of individuals (N) is calculated from S, I and R:

$$N = S + I + R \tag{4}$$

where $S(0) = S_0$, $I(0) = I_0$, and $R(0) = R_0$

A collection of codes based on the SIR model, using MATLAB, is available for researchers and local authorities to forecast the number of cases and the duration of the epidemic in different countries (H. Zhou & Lee, 2011).

2.2. Logistic model

The logistic type model can be written as follows (Garnier & Quetelet, 1838):

$$Y(d) = \frac{Y_\infty}{1 + \left(\frac{Y_\infty}{Y_0} - 1\right)e^{-r.d}} \tag{5}$$

where Y is the cumulative number of reported cases (C) or deaths (D), d is the number of days, while Y_∞, Y_0, and r are fitting parameters corresponding to the final projected number of cases or deaths after the end of the epidemic, the initial value, and the time constant, respectively.

2.3. Sigmoidal transition model

The transition sigmoid type model (also known as Hill equation) is written as follows (Yo = 0) (Gesztelyi et al., 2012):

$$Y(d) = \frac{Y_\infty}{1 + \left(\frac{d}{d_{1/2}}\right)^{-m}} \tag{6}$$

where, Y is the cumulative number of reported cases (C) or deaths (D), d is the number of days, while Y_∞, $d_{1/2}$, m are all positive fitting parameters corresponding to the final projected number of cases or deaths after the end of the epidemic, the half time period, and the hill slope exponent, respectively.

3. Results and analysis

The statistical data of the first six months of the pandemic for Italy (as a case study) has been used and analyzed based on the three models. A detailed analysis of the data for Italy during the pandemic progression is given in the following subsections.

The multiplicity of waves and peaks is linked to the appearance of virus variants throughout the pandemic spread, so the specific corresponding dates serve to explain the rise of the recorded data. According to the World Health Organization (WHO), several virus variants have been identified worldwide. These variants are summarized in Table 1. The UK variant (Alpha, B.1.1.7) is thought to have first appeared in autumn 2020.

3.1. SIR model analysis

The first six months of the COVID-19 pandemic data within Italy were obtained and analyzed using the SIR model fitting. The obtained results are depicted in Fig. 1 (a, b), which compares the actual cases, and SIR predicted evolution (plot output of the MATLAB codes developed by (Batista, 2020)).

It could be observed from Fig. 1 that the deviation of the predicted curve from the actual 1st six months' data (one single peak) illustrates the limitations of such a model. This deviation would be even more drastic when using the comprehensive data (more than one year and a half of the pandemic). This shows that the SIR model provides unrealistic estimates, even for one wave period.

Fig. 1 shows plots obtained from a MATLAB program that was developed by (Batista, 2020). The program uses a specific color scheme to describe the different parameters of the SIR fit results such as the model curve (blue line), the upper and lower estimate bounds (red color dashed lines), the final number of cases (C_{end} represented by green dashed line), the peak date (vertical red line), and the periods of ascending, descending, and final period of the wave (indicated by red, yellow and green bands, respectively). The parameters (R_n) and (R_o) displayed on top of Fig. 1 are model parameters defined by Batista in his Matlab program. Based on epidemiological terminology (Dharmaratne et al., 2020), R_o refers to the number of cases

Table 1
Dominant mutants worldwide.

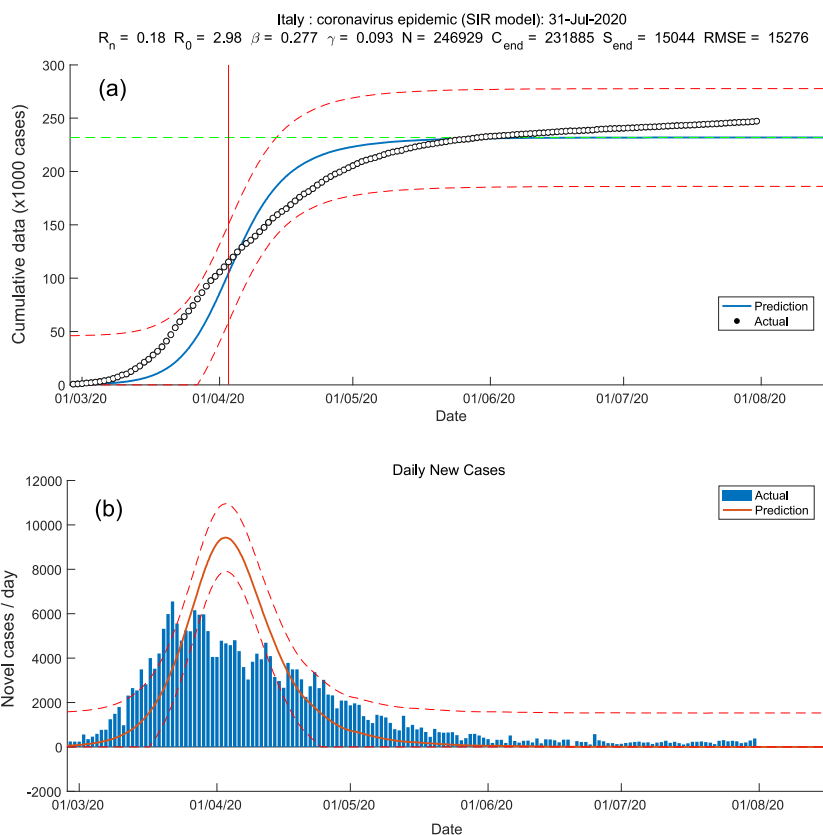| WHO Label	Pango Lineage	Earliest documented Samples	Transmissibility	Immune Evasiveness	Vaccine Effectiveness
Alpha	B.1.1.7	United Kingdom (Fall 2020)	+ + +	— —	✓
Beta	B.1.351	South Africa	+	+ + + +	✓
Gamma	P.1	Brazil	+ +	+ +	✓
Delta	B.1.617.2	India (April 2021)	+ + + +	+ +	✓
Lambda	C.37	Peru	+ + + +	+ +	✓

Fig. 1. SIR model fitting results for Italy's data until the July 31, 2020. (a) Cumulative infection data, and (b) daily data.

directly caused by an infected individual during his infectious period. R_o is thus a measure of the ability of a disease to spread amongst a given population. The reproduction number (R_n) represents the transmissibility of a disease.

3.2. Sigmoidal and logistic analysis

The first six months' predictions data obtained from sigmoidal transition (ST) and logistic models for Italy are summarized in Table 2 and Table 3, respectively. These estimates were obtained using the MATLAB curve fitting tool. By comparing Tables 2 and 3, it could be observed that the ST model outperforms the logistic model starting from 80 days of the pandemic. The predicted final cases error percentage fell under 2% for the sigmoidal model, while the logistic model varied from about 14% down to 5%.

The fitted data after 162 days are shown in Fig. 2 (a, b). Clearly, the sigmoidal transition model (Fig. 2(b)) shows a very close correspondence compared to the logistic model (Fig. 2(b)). The prediction bounds are very narrow for the sigmoidal fit model compared to that of the logistic fit.

Table 2
Final cases prediction by the sigmoidal transition model during the pandemic progression of Italy.

Days	C_∞	N	$d_{1/2}$	R^2	Error (%)
30	12432455	3.52	143.1	0.9995	5066.5
40	194606	4.18	38.08	0.9995	−22.54
50	192479	4.17	37.97	0.9998	−23.39
60	220131	3.78	40.89	0.9995	−12.38
80	245887	3.44	43.9	0.9994	−2.13
100	246905	3.42	44.03	0.9996	−1.72
120	246078	3.44	43.92	0.9997	−2.05
140	246179	3.44	43.93	0.9997	−2.01
160	247117	3.4	44.07	0.9997	−1.64
162	247269	3.4	44.09	0.9997	−1.58

Table 3

Final cases prediction by the logistic model during the pandemic progression of Italy.

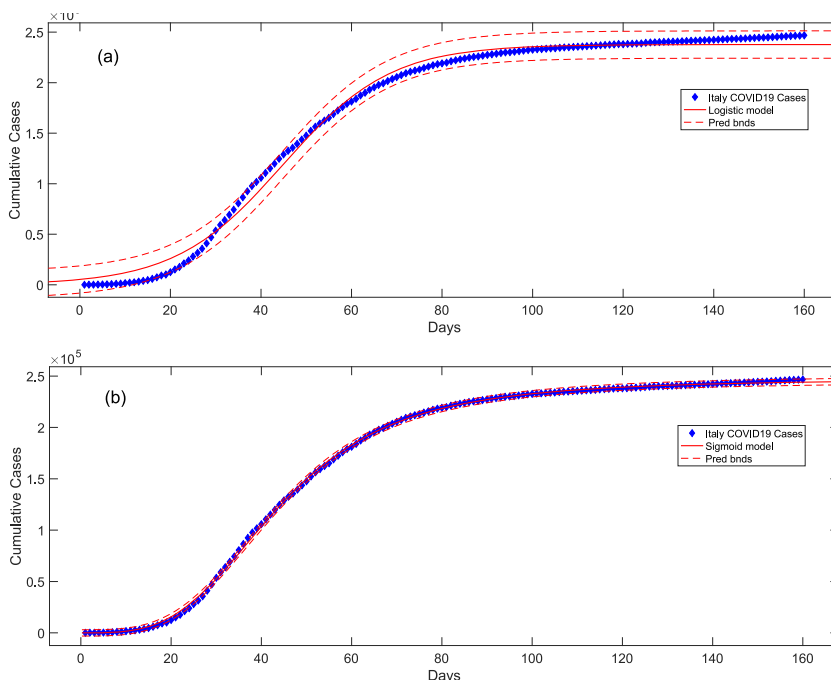| Days	C_o	C_∞	R	R^2	Error (%)
30	279	116782	0.195	0.9991	−53.52
40	351	130510	0.184	0.9997	−48.05
50	650	155190	0.158	0.9988	−38.23
60	1305	182353	0.131	0.9967	−27.42
80	2809	214825	0.104	0.9950	−14.49
100	3755	226269	0.095	0.9950	−9.94
120	4363	231543	0.090	0.9949	−7.84
140	4837	234950	0.087	0.9947	−6.4823
160	5271	237722	0.084	0.9941	−5.38
162	5317	237997	0.084	0.9940	−5.27

Fig. 2. Comparison between (a) logistic and (b) sigmoidal regression fits for Italy's data (1st six months).

Table 4

MATLAB curve fitting results (Italy's first six months).

Sigmoidal transition $C = C_\infty/(1+(d/d_{1/2})^{-m})$		Logistic $C = C_\infty/(1+(C_\infty/C_o -1)*e^{(-r*d)})$	
Parameter	Value (95% conf. interval)	Parameter	Value (95% conf. interval)
C_∞	2.473e+05 (2.467e+05, 2.478e+05)	C_∞	2.38e+05 (2.364e+05, 2.396e+05)
M	3.399 (3.365, 3.434)	R	0.08411 (0.08082, 0.08741)
$d_{1/2}$	44.09 (43.95, 44.23)	C_o	5317 (4541, 6092)
R^2	RMSE	R^2	RMSE
0.9997	1583.9	0.9940	6841.1

The curve fitting results for Italy's first six months are given in Table 4. The goodness of fit parameters for the sigmoidal model are better than that of the logistic model: (R2 = 0.9997, RMSE = 1583.9) compared to (R2 = 0.9940, RMSE = 6841.1), respectively. Where R2 is the regression coefficient and the RMSE is the root mean square difference estimate.

4. Multiple-terms sigmoidal transition model

The irregularity of the recorded COVID-19 pandemic data leads to unrealistic predictions using the usual models. Based on the complete data sets (more than 1½ years), the three previous models failed to predict the correct patterns due to the

numerous spikes and waves associated with the pandemic. The sigmoidal model was found to have better agreement compared to the other two models. Since the sigmoidal model (Eq. (6)) works well for a single data distribution with one peak, it can be extended for a data distribution with multiple peaks (k-peaks). As a first approach, the number of terms is equivalent to the number of peaks. Nonetheless, the optimum number of terms to be used for a given set of data should be systematically obtained based on the statistical parameters R^2, RMSE, and 95% confidence interval. A developed generalized form of the sigmoidal transition model for multiple peaks is thus given in Eq. (7):

$$Y(d) = \frac{Y_{\infty,1}}{1 + \left(\frac{d}{d_{1/2,1}}\right)^{-m_1}} + \frac{Y_{\infty,2}}{1 + \left(\frac{d}{d_{1/2,2}}\right)^{-m_2}} + \cdots + \frac{Y_{\infty,k}}{1 + \left(\frac{d}{d_{1/2,k}}\right)^{-m}} = \sum_{i=1}^{k} \frac{Y_{\infty,i}}{1 + \left(\frac{d}{d_{1/2,i}}\right)^{-m_i}} \tag{7}$$

The forecasted final number of cases (or mortality) is obtained from Eq. (8) at an infinite time (when $\left(\frac{d}{d_{d_{1/2,i}}}\right)^{-m_i} \to 0$), by summing all the $Y_{\infty,i}$:

$$Y_\infty = \sum_{i=1}^{k} Y_{\infty,i} \tag{8}$$

The parameters of Eq. (7) and Eq. (8) are defined in section 2.3.

This novel modeling scheme will be applied to official data of several countries (UK, Italy, Saudi Arabia, and Tunisia) to validate its accuracy for 500 days. The countries have been selected as case studies for illustration purposes.

4.1. United Kingdom (UK) case study

The first reported case in the United Kingdom was on January 31, 2020, and 158,488 cases had been reported in England as of May 21, 2020 (Qasim et al., 2020). From the beginning of the pandemic, it was clear that the epidemic would have a severe impact on certain socioeconomic groups, usually those already most affected by health inequalities.

Fig. 3 lays out the overall fatality data for the UK (more than 1½ year data); clearly, just as all other countries, the UK has endured multiple peaks (waves) pandemic spread. Predicting the ultimate data using standard models (SIR) would be unrealistic. The transitional type evolution of the pandemic would be best described using the multiple terms sigmoidal transition (MTST) model (using as many terms as the peaks) as given by Eqs. (7) and (8). As shown in Fig. 3, the UK has recorded roughly 4 or 5 peaks, requiring five terms sigmoidal fit (R2 = 1.00; RMSE = 262.95).

As can be read from Table 5, the fit parameters are obtained with relatively narrow confidence intervals; as can be deduced from these data, the ultimate mortality toll (D_∞) is the sum of the $D_{\infty,i}$ (calculated using Eq. (8)).

Fig. 3 shows that the UK variant coincides with the period between 200 and 300 days after the first death caused by this variant. As shown in Fig. 3, a peak was recorded during this period, followed by an even more pronounced peak, reflecting the population impact of such a variant both locally and internationally.

The UK health authorities then launched a widespread vaccination campaign, which proved successful, as shown in Fig. 3 (data after 400 days). Indeed, despite the appearance of the Indian variant (Delta, B.1.617.2), the mortality numbers stayed very low.
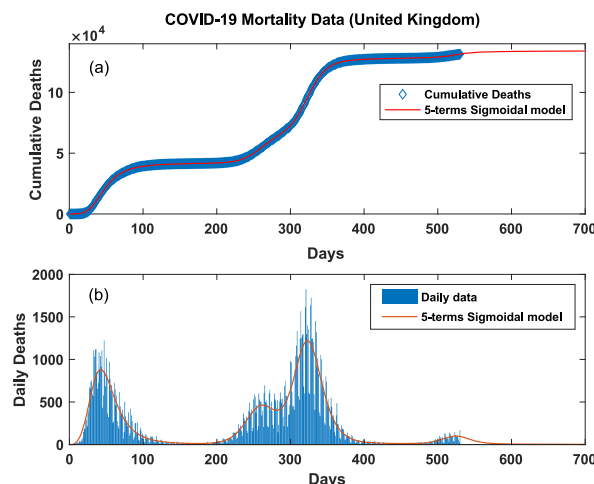

Fig. 3. Mortality data for the UK: (a) cumulative data, (b) daily data.

Table 5
MATLAB curve fitting results (sensitivity analysis for Saudi Arabian data).

Data size	No. terms	Ultimate cumulative value		Peak		R^2	RMSE
(days)		D_∞	Upper bound	(day)	(count)		
512	4 terms	9513	11,128	450	16	0.9999	21.265
	5 terms	9165	10,023	450	16	1.0000	11.278
400	5 terms	9567	16,270	460	18	1.0000	10.808
350	5 terms	9164	21525	420	28	1.0000	11.438

4.2. Italy case study

Like the UK, the Italian pandemic data show multiple peaks and waves, reflecting the mutation of the virus and its impact (Fig. 4). Again, Italy experienced the second peak (at about 300 days) after the first 250 days, which coincided with the spread of the UK variant. The fourth peak (at about 400 days) appears to coincide with the emergence of the Indian variant. However, thanks to the successful vaccination campaign, mortality figures in Italy have fallen dramatically and have not changed after that.

4.3. Saudi Arabia case study

The pandemic data from Saudi Arabia show several peaks; nevertheless, the sanitary measures taken by Saudi health authorities seemed to prevent the wide spread of the various COVID19 virus variants. Fig. 5 shows that mortality rates declined between 200 and 300 days (UK variant). Due to the vaccination campaign, KSA managed to prevent the impact of both the British and Indian variants (i.e., a very moderate increase in mortality rates in the period between 400 and 500 days).

From Fig. 5(b), Saudi Arabian data seemingly reflect 2 or 3 waves, but looking closely at the data, multiple changes (decreases and increases) have been recorded during the first wave (a 300 day long) and for various durations. In fact, prevention policies and protocols, human behavior, etc., during a given wave may lead to an occasional increase (or a decrease during better management of the pandemic) in the recorded data leading to the appearance of a "sub-peak" causing divergence of the fitting scheme. Clearly, a first surge has been recorded after about 50 days of the start of the pandemic. After a peak at 100 days, the decreasing trend has been halted at around 140 days to exhibit a slight increase before decreasing again. So for Saudi Arabia and following a sensitivity analysis (§4.3.1b), 5 terms are optimal for fitting the recorded data.

4.3.1. Sensitivity analysis of the MTST model for Saudi Arabian data

(a) *Effects of Data Size*: The sensitivity of the MTST model to the data size has been further explored for Saudi Arabian pandemic mortality data. The fitting results have been obtained for 350, 400, and more than 512 days of data.
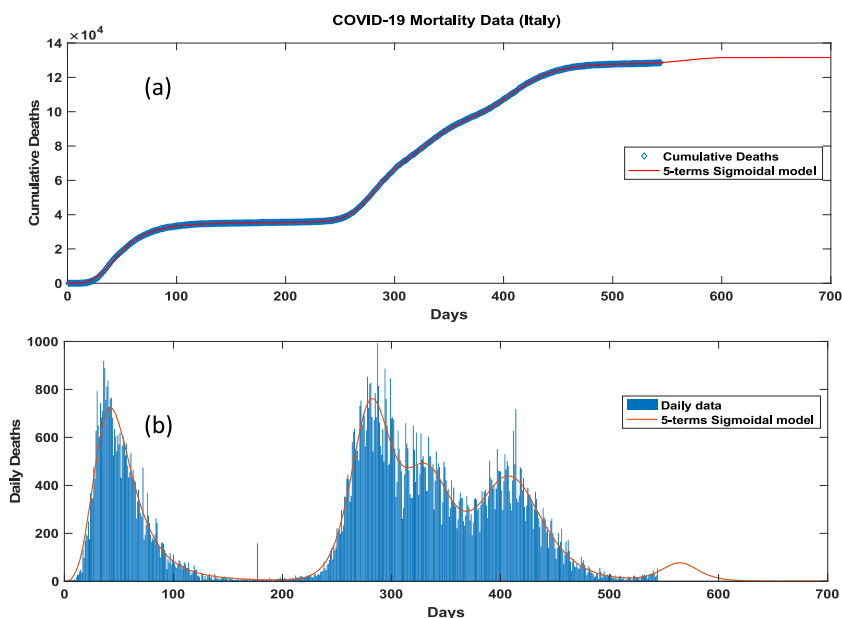

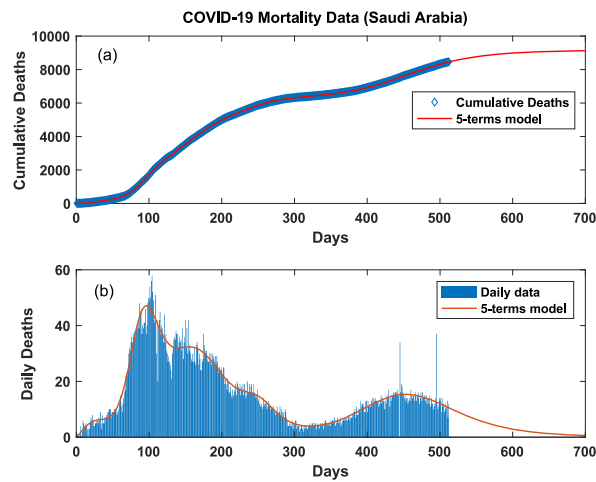**Fig. 4.** Mortality data for Italy: (a) cumulative data, (b) daily data.

Fig. 5. Mortality data for Saudi Arabia: (a) cumulative, (b) daily data. (5-terms MTST model).

Firstly, the complete data set (512 days) was fitted using 4 and 5 terms of the MTST model and revealed that the 5-terms fit results had $R^2 = 1$ and a significantly lower RMSE than that of the 4-terms ($R^2 = 1$ & RMSE = 10.28 compared to $R^2 = 0.9999$ & RMSE = 21.26). This indicates that an optimization of the number of terms is required to obtain the most appropriate number of terms to use in the MTST model.

To study the sensitivity to the data size during an ongoing wave, the MTST model has also been used to fit partial data (350 and 400 days). The forecasted curves based on partial data (350 and 400 days) are given in Fig. 6. Both partial data have been fitted with a 5-term MTST model. Based on 350 days only, the final peak occurs earlier with a higher peak value compared to the results based on 400 days which had a peak occurring slightly later with a relatively lower peak level.

The sensitivity analysis for the Saudi Arabian data is summarized in Table 5. From this sensitivity analysis, it appears that using data size just before and just after the actual final peak does not have a significant impact on the ultimate forecasted value nor the regression quality (equivalent R2 and RMSE). On the other hand, partial data size has a significant impact on the upper estimation bound of the ultimate value and a slight impact on the day where the peak occurs when using 350 days
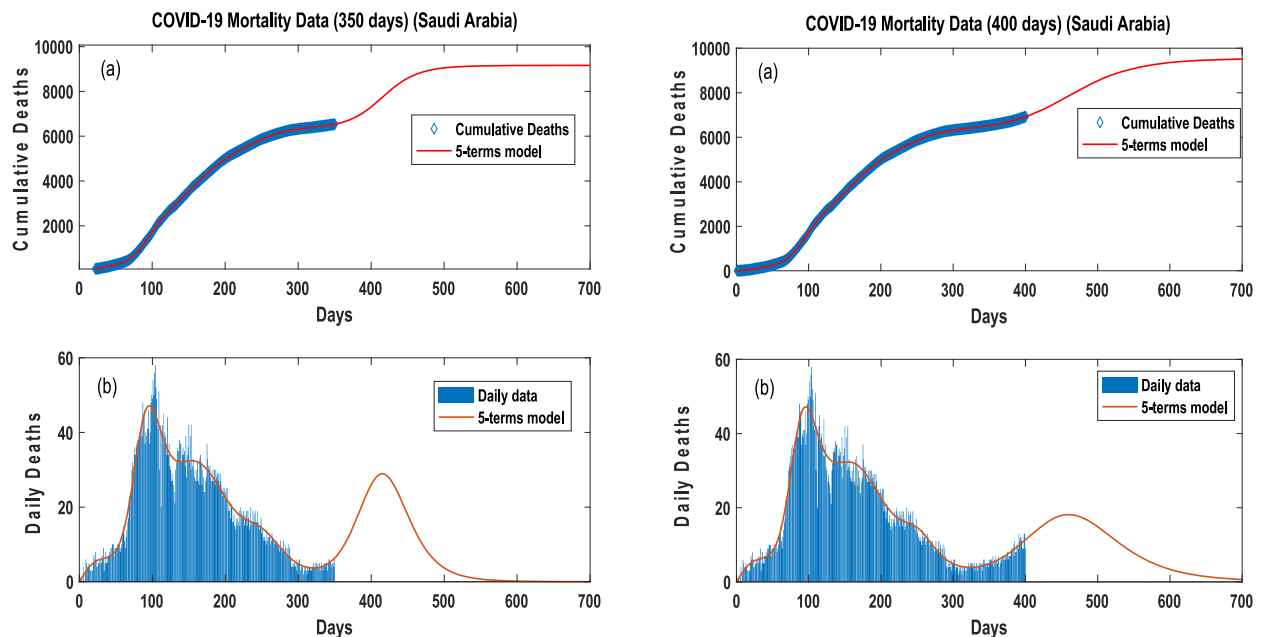

Fig. 6. Mortality data for Saudi Arabia (350 & 400 days data): (a) cumulative, (b) daily data.
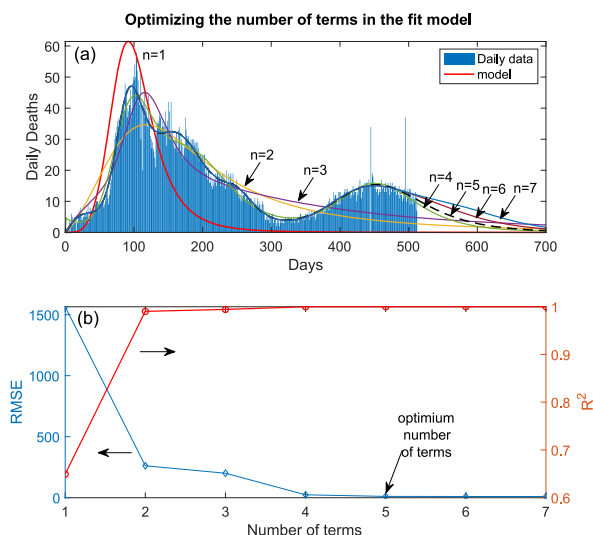
Fig. 7. MTST model sensitivity analysis for Saudi Arabia: (a) daily data model fits for nterms = 1 to 7 (nterms = 5 as optimum scenario), (b) evolutions of RMSE and R^2 vs. nterms.

instead of the complete set of data. Obviously, the more data we use, the more accurate and more reliable the fit estimates are (i.e. narrower estimate confidence bounds).

(b) *Effects Of The Number Of Terms*: The impact of the number of terms and their significance has also been explored for Saudi Arabian pandemic mortality data as a case study. Fitting of the complete data set (of 512 days) has been performed using up to 7 terms in an attempt to systematically choose the optimum number of terms.

Fig. 7(a) summarizes the different model curves corresponding to the various numbers of terms used. Fig. 7(b) shows that using more terms leads to an increase in R^2 and a decrease in the RMSE; thus, a better fit is obtained for nterms $\geq$ 4. It should be noted, however, that the optimum number of terms to choose is the one corresponding to narrow 95% confidence intervals having positive lower and upper limits of the model parameters estimates (D_∞, $d_{1/2}$, and m). Table 6a shows fit results for nterms = 5, 6, and 7 scenarios, clearly indicating that R^2 and RMSE alone are not sufficient to decide whether or not a fit result

Table 6a

Sensitivity analysis of the MTST model fit parameters to the number of terms for KSA.

$D_{\infty,i}$	value (95% conf. interval)	n_i	value (95% conf. interval)	$d_{1/2,i}$	value (95% conf. interval)
(a) 5-terms MTST fit parameters ($\mathbf{R^2}$ = 1.0000; **RMSE** = 11.28)					
$D_{\infty,1}$	452.9 (293, 612.8)	m_1	2.063 (1.731, 2.396)	$d_{1/2,1}$	49.97 (34.14, 65.81)
$D_{\infty,2}$	2359 (2133, 2584)	m_2	6.64 (6.311, 6.97)	$d_{1/2,2}$	97.95 (96.96, 98.95)
$D_{\infty,3}$	3160 (2882, 3438)	m_3	6.148 (5.668, 6.627)	$d_{1/2,3}$	174.6 (173.3, 175.9)
$D_{\infty,4}$	462 (335, 589.1)	m_4	15.79 (13.38, 18.19)	$d_{1/2,4}$	252.1 (250.5, 253.6)
$D_{\infty,5}$	2731 (2663, 2799)	m_5	10.18 (9.943, 10.42)	$d_{1/2,5}$	461.7 (460.4, 463)
$\boldsymbol{D_\infty}$ **= 9165 (max. 10023)** *acceptable overall fit parameters and intervals bounds*					
(b) 6-terms MTST fit parameters ($\mathbf{R^2}$ = 1.0000; **RMSE** = 9.94)					
$D_{\infty,1}$	501.4 (294.9, 707.8)	m_1	1.986 (1.665, 2.306)	$d_{1/2,1}$	54.64 (34.4, 74.89)
$D_{\infty,2}$	2277 (2022, 2531)	m_2	6.751 (6.379, 7.123)	$d_{1/2,2}$	97.63 (96.57, 98.69)
$D_{\infty,3}$	3217 (2909, 3525)	m_3	6.021 (5.502, 6.539)	$d_{1/2,3}$	174.1 (172.7, 175.6)
$D_{\infty,4}$	471.7 (312.5, 631)	m_4	14.87 (12.14, 17.59)	$d_{1/2,4}$	253 (250.8, 255.1)
$D_{\infty,5}$	1521 (−6415, 9457)	m_5	11.79 (1.966, 21.61)	$d_{1/2,5}$	437.8 (357.8, 517.9)
$D_{\infty,6}$	1534 (−8082, 1.115e+04)*	m_6	10.18 (−19.9, 40.25)*	$d_{1/2,5}$	521.8 (285.4, 758.2)
$\boldsymbol{D_\infty}$	**9522 (max. 28002)** *relatively high upper bound and some negative lower intervals bounds*				
(c) 7-terms MTST fit parameters ($\mathbf{R^2}$ = 1.0000; **RMSE** = 9.96)					
$D_{\infty,1}$	521.5 (295.4, 747.7)	m_1	1.962 (1.644, 2.28)	$d_{1/2,1}$	56.54 (34.47, 78.61)
$D_{\infty,2}$	2236 (1977, 2495)	m_2	6.803 (6.419, 7.187)	$d_{1/2,2}$	97.46 (96.4, 98.52)
$D_{\infty,3}$	3273 (2965, 3580)	m_3	5.926 (5.416, 6.435)	$d_{1/2,3}$	174.3 (172.8, 175.7)
$D_{\infty,4}$	443.6 (267.9, 619.3)	m_4	15.35 (12.11, 18.59)	$d_{1/2,4}$	253.2 (250.8, 255.5)
$D_{\infty,5}$	1769 (−1.66, 1.69) e+05	m_5	11.57 (−65.35, 88.49)	$d_{1/2,5}$	441.1 (−735.7, 1618)
$D_{\infty,6}$	942.8 (−2.27, 2.27) e+06*	m_6	11.13 (−2319, 2341)	$d_{1/2,6}$	520 (−4.24, 4.34) e+04
$D_{\infty,7}$	615.6 (−7.02, 7.02) e+06*	m_7	14.01 (−7.46, 7.46) e+04*	$d_{1/2,7}$	593 (−3.624, 3.64) e+05
$\boldsymbol{D_\infty}$ **= 9802 (max. 9466842)** *unrealistic upper bounds and numerous negative lower intervals bounds*					

Table 6b
Summary of the selection criteria of the number of terms to be used in the MTST model for KSA.

n	R^2	RMSE	Reject/Accept
1	0.6493	1565.3	Reject for inaccurate D_∞ ; low R^2; high RMSE
2	0.9902	261.75	Reject for negative lower bound of D_0 ; low R^2; high RMSE
3	0.9942	200.36	Reject for negative lower bounds; low R^2; high RMSE
4	0.9999	23.46	Reject for negative lower bound of D_1, and $d_{1/2,2}$
5	1.0000	11.28	Accept for optimum fit criterion
6	1.0000	9.94	Reject for negative lower bounds and relatively high upper bound of D_∞ (Table 6a)
7	1.0000	9.96	Reject for negative lower bounds and unrealistic high upper bound of D_∞ (Table 6a)

is acceptable. In fact, both (6) and (7) terms fit models give exceptional R^2 and RMSE performances but both significantly fail to provide realistic fit parameters values and reasonable 95% confidence intervals (having negative lower intervals bounds).

Table 6b summarizes fit results for nterms varying from 1 to 7 while given the rejection criteria for the various n-terms scenarios. The most critical rejection criterion is in fact too wide 95% confidence interval with a negative lower bound. Because the fit parameters are basically positive parameters (see §2.3), no negative bound is accepted.

4.4. Tunisia case study

Fig. 8 shows that Tunisia managed to control the spread of the pandemic in the initial phase (up to 200 days). However, following the emergence of the different variants and the lack of an effective vaccination campaign, combined with the inability of the authorities to enforce hygiene regulations and laws, Tunisia has experienced several consecutive peaks of pandemic spread. This continuous trend of repetitive peaks reflects the failure to maintain a minimum level of control over the spread of the pandemic. The 4th peak appears to coincide with the emergence of the Indian variant. Despite a marked increase in deaths, Tunisia is well on its way to controlling the spread.

4.5. MTST model fitting summary for the studied countries

Table 7 summarizes the projected ultimate number of fatalities for the four studied countries. Only the upper bound is indicated along with the average estimated D_∞. It is well established from Figures (3-6) that the multiple terms sigmoidal transition model (MTST) fits well the mortality data curves for all countries included in this study. The fitting model
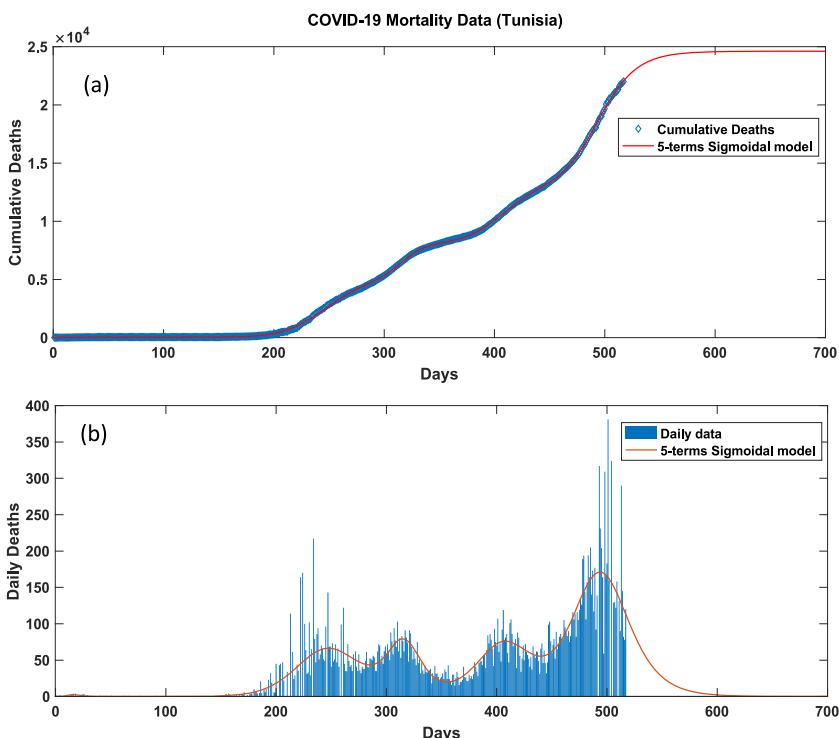

Fig. 8. Mortality data for Tunisia: (a) cumulative data, (b) daily data.

Table 7
Multi-Terms Sigmoidal Transition model fit parameters for UK, Italy, KSA, and Tunisia.

$D_{\infty,i}$	value (95% conf. interval)	m_i	value (95% conf. interval)	$d_{1/2,i}$	value (95% conf. interval)
(d) UK: 5-terms MTST fit parameters ($\mathbf{R^2} = 1.0000$; $\mathbf{RMSE} = 262.95$)					
$D_{\infty,1}$	41690 (41500, 41870)	m_1	3.81 (3.76, 3.85)	$d_{1/2,1}$	48.31 (48.13, 48.49)
$D_{\infty,2}$	24390 (23330, 25440)	m_2	18.44 (17.68, 19.19)	$d_{1/2,2}$	260.9 (260.1, 261.7)
$D_{\infty,3}$	59060 (57720, 60400)	m_3	25.85 (25.35, 26.36)	$d_{1/2,3}$	323.7 (323.5, 324)
$D_{\infty,4}$	5045 (1363, 8727)	m_4	3.87 (0.67, 7.07)	$d_{1/2,4}$	413.8 (127, 700.6)
$D_{\infty,5}$	4479 (1467, 8812)	m_5	43.41 (16.22, 70.6)	$d_{1/2,5}$	524.7 (505.6, 543.8)
$\boldsymbol{D_\infty}$	***134650 (max. 150401)***				
(e) Italy: 4-terms MTST fit parameters ($\mathbf{R^2} = 1.0000$; $\mathbf{RMSE} = 180.83$)					
$D_{\infty,1}$	35680 (35640, 35720)	m_1	3.66 (3.63, 3.68)	$d_{1/2,1}$	48.62 (48.52, 48.72)
$D_{\infty,2}$	36170 (35120, 37220)	m_2	22.61 (22.2, 23.01)	$d_{1/2,2}$	281.9 (281.3, 282.5)
$D_{\infty,3}$	23050 (21510, 24600)	m_3	23.52 (22.06, 24.98)	$d_{1/2,3}$	333.6 (332.7, 334.4)
$D_{\infty,4}$	33330 (32670, 34000)	m_4	20.86 (20.51, 21.2)	$d_{1/2,4}$	409.6 (408.9, 410.3)
$\boldsymbol{D_\infty}$	***128240 (max. 131540)***				
(f) KSA: 5-terms MTST fit parameters ($\mathbf{R^2} = 1.0000$; $\mathbf{RMSE} = 10.278$)					
$D_{\infty,1}$	452.9 (293, 612.8)	m_1	2.063 (1.731, 2.396)	$d_{1/2,1}$	49.97 (34.14, 65.81)
$D_{\infty,2}$	2359 (2133, 2584)	m_2	6.64 (6.311, 6.97)	$d_{1/2,2}$	97.95 (96.96, 98.95)
$D_{\infty,3}$	3160 (2882, 3438)	m_3	6.148 (5.668, 6.627)	$d_{1/2,3}$	174.6 (173.3, 175.9)
$D_{\infty,4}$	462 (335, 589.1)	m_4	15.79 (13.38, 18.19)	$d_{1/2,4}$	252.1 (250.5, 253.6)
$D_{\infty,5}$	2731 (2663, 2799)	m_5	10.18 (9.943, 10.42)	$d_{1/2,5}$	461.7 (460.4, 463)
$\boldsymbol{D_\infty}$	***9165 (max. 10023)***				
(g) Tunisia: 4-terms MTST fit parameters ($\mathbf{R^2} = 0.9999$; $\mathbf{RMSE} = 48.88$)					
$D_{\infty,1}$	4466 (4356, 4577)	m_1	14.69 (14.29, 15.08)	$d_{1/2,1}$	241.1 (240.1, 242)
$D_{\infty,2}$	3865 (3713, 4018)	m_2	22.79 (21.66, 23.92)	$d_{1/2,2}$	313.1 (312.5, 313.8)
$D_{\infty,3}$	4491 (4305, 4678)	m_3	26.83 (25.55, 28.12)	$d_{1/2,3}$	407.1 (406.3, 408)
$D_{\infty,4}$	12040 (11620, 12460)	m_4	27.65 (26.59, 28.7)	$d_{1/2,4}$	495 (494.2, 495.7)
$\boldsymbol{D_\infty}$	***24865 (max. 25733)***				

regression coefficient R2 is unity, and RMSE is very low for all countries, indicating that MTST is more effective for analyzing and predicting multi-waves pandemics.

5. Conclusions

The nature and spread of the COVID-19 pandemic make the predictions of current models unrealistic. The multi-peaks outbreaks of COVID-19 and its variants require a more elaborated modeling approach. The multiple terms sigmoidal transition (MTST) model fits well the mortality curve compared to the logistic model. Moreover, the MTST model reproduces the different outbreaks well and is very close to the actual outbreak data. A sensitivity analysis of the MTST model using different data sizes (for Saudi Arabia) revealed that using data size just before and just after the actual final peak does not have a significant impact on the ultimate forecasted value nor the regression quality (equivalent R^2 and RMSE). Divergence of the fitting scheme mainly occurs due to (i) inadequate initial estimate bounds for the different model parameters and (ii) unrealistic number of terms. Furthermore, the sensitivity of the MTST model to the number of terms showed that it is possible to find the optimum number of terms to be used corresponding to high R2, low RMSE, and narrow 95% confidence intervals of the different fit parameters. While for a given choice of the number of terms, any unrealistically wide 95% confidence interval with a negative lower bound leads to rejection of the scenario. Therefore, it can be concluded that the MTST model is more effective compared to the logistic model for pandemics with multiple outbreaks. The MTST model will be beneficial for the relevant authorities to predict the consequences of the pandemic at an earlier stage and take countermeasures to minimize the impact. This approach allows to reasonably narrow down the estimate bounds when the SIR or logistic models produce unreliable high confidence intervals in the estimate of fitting parameters such as the final cases size or the ultimate number of deaths.

Declaration of competing interest

None Declared

Acknowledgment

References

Aba Oud, M. A., Ali, A., Alrabaiah, H., Ullah, S., Khan, M. A., & Islam, S. (2021). A fractional order mathematical model for COVID-19 dynamics with quarantine, isolation, and environmental viral load. *Advances in Difference Equations, 2021*(1). https://doi.org/10.1186/s13662-021-03265-4

Allen, L. J. S., Brauer, F., den Driessche, P., & Wu, J. (2008). *Mathematical epidemiology* (Vol. 1945). Springer.

Alshammari, F. S., & Khan, M. A. (2021). Dynamic behaviors of a modified SIR model with nonlinear incidence and recovery rates. *Alexandria Engineering Journal, 60*(3), 2997–3005.

Analytica, O. (2020). *New COVID-19 strain brings South Africa more headaches*. Emerald Expert Briefings (oxan-es).

Arazi, R., & Feigel, A. (2021). Discontinuous transitions of social distancing in the SIR model. *Physica A: Statistical Mechanics and its Applications, 566*, Article 125632.

Barrio, R. A., Kaski, K. K., Haraldsson, G. G., Aspelund, T., & Govezensky, T. (2021). A model for social spreading of Covid-19: Cases of Mexico, Finland and Iceland. *Physica A: Statistical Mechanics and its Applications, 582*, Article 126274.

Batista, M. (2020). Estimation of the final size of the COVID-19 epidemic. *medRxiv, 16*, 2020.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology* (Vol. 32). Springer.

Cairns, S. P., Robinson, D. M., & Loiselle, D. S. (2008). Double-sigmoid model for fitting fatigue profiles in mouse fast-and slow-twitch muscle. *Experimental Physiology, 93*(7), 851–862.

Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons ＼& Fractals, 139*, Article 110057.

Dharmaratne, S., Sudaraka, S., Abeyagunawardena, I., Manchanayake, K., Kothalawala, M., & Gunathunga, W. (2020). Estimation of the basic reproduction number (R0) for the novel coronavirus disease in Sri Lanka. *Virology Journal, 17*(1), 1–7.

Di Veroli, G. Y., Fornari, C., Goldlust, I., Mills, G., Koh, S. B., Bramhall, J. L., et al. (2015). An automated fitting procedure and software for dose-response curves with multiphasic features. *Scientific Reports, 5*(1), Article 14701. https://doi.org/10.1038/srep14701

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals, 134*, Article 109761. https://doi.org/10.1016/j.chaos.2020.109761

Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., & Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet, 397*(10278), 952–954.

Gadagkar, S. R., & Call, G. B. (2015). Computational tools for fitting the Hill equation to dose-response curves. *Journal of Pharmacological and Toxicological Methods, 71*, 68–76.

Garnier, J.-G., & Quetelet, A. (1838). *Correspondance mathématique et physique* (Vol. 10). Impr. d'H. Vandekerckhove.

Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B., & Tosaki, A. (2012). The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences, 66*(4), 427–438.

Getz, W. M., Salter, R., Muellerklein, O., Yoon, H. S., & Tallam, K. (2018). Modeling epidemics: A primer and numerus model builder implementation. *Epidemics, 25*, 9–19. https://doi.org/10.1016/j.epidem.2018.06.001

Kabir, K. M. A., Kuga, K., & Tanimoto, J. (2019). Analysis of SIR epidemic model with information spreading of awareness. *Chaos, Solitons ＼& Fractals, 119*, 118–125.

Li, F., Meng, X., & Wang, X. (2018). Analysis and numerical simulations of a stochastic SEIQR epidemic system with quarantine-adjusted incidence and imperfect vaccination. *Computational and Mathematical Methods in Medicine, 2018*, 1–14. https://doi.org/10.1155/2018/7873902

Malavika, B., Marimuthu, S., Joy, M., Nadaraj, A., Asirvatham, E. S., & Jeyaseelan, L. (2021). Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models. *Clinical Epidemiology and Global Health, 9*, 26–33.

Malhotra, B., & Kashyap, V. (2020). *Progression of COVID-19 in Indian states-forecasting endpoints using SIR and Logistic Growth models*.

Martcheva, M. (2015). *An introduction to mathematical epidemiology* (Vol. 61). Springer.

Meddings, J. B., Scott, R. B., & Fick, G. H. (1989). Analysis and comparison of sigmoidal curves: Application to dose-response data. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 257*(6), G982–G989.

Meng, X., Zhao, S., Feng, T., & Zhang, T. (2016). Dynamics of a novel nonlinear stochastic SIS epidemic model with double epidemic hypothesis. *Journal of Mathematical Analysis and Applications, 433*(1), 227–242.

Ndii, M. Z., & Supriatna, A. K. (2017). Stochastic mathematical models in epidemiology. *Information, 20*, 6185–6196.

Neves, A. G. M., & Guerrero, G. (2020). Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and Sao Paulo state, Brazil. *Physica D: Nonlinear Phenomena, 413*, Article 132693.

Odagaki, T. (2021). Exact properties of SIQR model for COVID-19. *Physica A: Statistical Mechanics and Its Applications, 564*, Article 125564.

Pessoa, D., Souto-Maior, C., Gjini, E., Lopes, J. S., Ceña, B., Codeço, C. T., et al. (2014). Unveiling time in dose-response models to infer host susceptibility to pathogens. *PLoS Computational Biology, 10*(8), Article e1003773.

Qasim, M., Ahmad, W., Yoshida, M., Gould, M., & Yasir, M. (2020). Analysis of the worldwide corona virus (COVID-19) pandemic trend;A modelling study to predict its spread. *MedRxiv, 3*, Article 20048215 (January), 2020.03.30.

Ramos, A. M., Ferrández, M. R., Vela-Pérez, M., Kubik, A. B., & Ivorra, B. (2021). A simple but complex enough θ-SIR type model to be used with COVID-19 real data. Application to the case of Italy. *Physica D: Nonlinear Phenomena, 421*, Article 132839.

Roberts, M. G., & Heesterbeek, J. A. P. (2003). *Mathematical models in epidemiology* (Vol. 215). EOLSS.

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling, 5*, 271–281. https://doi.org/10.1016/j.idm.2020.03.001

Rodrigues, H. S. (2016). *Application of SIR epidemiological model: New trends. ArXiv Preprint ArXiv:1611.02565*.

Tang, X., Thankappan, S. K., Lee, P., Fard, S. E., Harmon, M. D., Tran, K., et al. (2014). Polymeric biomaterials in tissue engineering and regenerative medicine. *Natural and Synthetic Biomedical Polymers*, 351–371.

Tian, B., & Yuan, R. (2017). Traveling waves for a diffusive SEIR epidemic model with non-local reaction. *Applied Mathematical Modelling, 50*, 432–449.

others Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., et al. (2021). *Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data*. MedRxiv.

Zhou, H., & Lee, J. (2011). Nanoscale hydroxyapatite particles for bone tissue engineering. *Acta Biomaterialia, 7*(7), 2769–2781.

others Zhou, X., Ma, X., Hong, N., Su, L., Ma, Y., He, J., et al. (2020). *Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model*. MedRxiv.