Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Time-variant reliability-based prediction of COVID-19 spread using extended SEIVR model and Monte Carlo sampling

Mahdi Shadabfar a, Mojtaba Mahsuli a, Arash Sioofy Khoojine b, c, Vahid Reza Hosseini d

a Center for Infrastructure Sustainability and Resilience Research, Department of Civil Engineering, Sharif University of Technology, Tehran 145888-9694, Iran
b Faculty of Economics and Business Administration, Yibin University, Yibin 644000, China
c School of Mathematics and Statistics, Shanghai Jiao Tong University, Shanghai, China
d Institute for Advanced Study, Nanchang University, Nanchang 330031, China

ARTICLE INFO
Keywords:
COVID-19 spread
Reliability model
Monte Carlo sampling
Exceedance probability
Transmission rate
Recovery rate
Mortality rate

ABSTRACT
A probabilistic method is proposed in this study to predict the spreading profile of the coronavirus disease 2019 (COVID-19) in the United State (US) via time-variant reliability analysis. To this end, an extended susceptible-exposed-infecte

vaccinated-recovered (SEIVR) epidemic model is first established deterministically, considering the quarantine and vaccination effects, and then applied to the available COVID-19 data from US. Afterwards, the prediction results are described as a time-series of the number of people infected, recovered, and dead. Upon introducing the extended SEIVR model into a limit-state function and defining the model parameters including transmission, recovery, and mortality rates as random variables, the problem is transformed into a reliability model and analyzed by the Monte Carlo sampling. The findings are subsequently given in the form of exceedance probabilities (EPs) of the three main outputs, namely, the maximum number of infected cases, the total number of recovered cases, and the total number of fatal cases. Afterwards, by incorporating time into the formulation of the reliability problem, the EPs are calculated over time and presented as 3D probability graphs, illustrating the relationship between the number of cases affected (i.e., infected, recovered, or dead), exceedance probability, and time. The results for the US demonstrate that, by the end of 2021, the number of the infected (active) cases decreases to 0.8 million and number of cases recovered and fatalities increases to 41.3 million and 0.6 million, respectively.

Introduction
The coronavirus disease 2019 (COVID-19) was detected in the City of Wuhan, Hubei Province, in December 2019, and quickly spread to other parts of China. The virus was then discovered in different countries but with some delay, and rapidly turned into an epidemic. The problem was finally declared as a pandemic by the World Health Organization (WHO) on March 11, 2020. The WHO also issued a series of preliminary regulatory determinations for healthcare services against the emerging disease and called on all nations to cooperate in its control. The disease then spread to the rest of the world, causing health crises in one region after another. Despite public vaccination in a limited number of countries, programs such as social isolation, physical distancing, and wearing masks are employed as the main control strategies to reduce the exponential growth of COVID-19.

Mathematical models can contribute to estimating transmission, recovery, and mortality rates in different countries and demonstrate the spread of disease [35–38]. Various nations have also taken varying levels of measures, and therefore the pattern of disease transmission differs from one country to another. Since the onset of COVID-19, several modeling methods have been proposed, in which the real data of the affected areas have been used to calibrate the models and different characteristics of the disease transmission [1–3]. As well, intervention strategies adopted by various countries have been considered to simulate the infection behavior [39–42]. These models have then been presented to decision-makers as tools of predicting the extent and the severity of the disease outbreak and providing essential information to determine the type and the effectiveness of intervention strategies [4]. Such models, including the susceptible-exposed-infected-recovered (SEIR) and its modified variants have been thus far practiced in
several studies to simulate the COVID-19 spreading profile [5–7].

Even though these methods are effective tools and typically provide valuable insight into understanding the rate and the speed of disease spread within the society, they suffer from one major drawback, that is, they are deterministic [8]. This leads to two further issues: First, the data reported by governments and related agencies fail to comprehensively reflect the state of the disease in the society and only represent detected and diagnosed cases. Many others are not entered into the diagnostic systems and are not reported in the data. Therefore, simply relying on these data and fitting models to them does not necessarily mean that the systems and are not reported in the data. Therefore, simply relying on these data and fitting models to them does not necessarily mean that the results are reliable [9]. Second, input parameters, such as infection, recovery, and mortality rates, are not well-known parameters and are normally associated with significant uncertainty [10]. Assuming these input parameters as fixed values will incorporate a margin of error into calculations and may lead to unreliable outcomes. In fact, a deterministic assumption about the spread of an epidemic means an illogical certainty assumption for each case affected by the disease and the treatment state of each patient. A more rational solution is to take the problem of disease transmission as a probabilistic model to determine the probability of transmission, recovery, and mortality rates by assuming input parameters as random variables.

The primary contribution of this paper is to propose an extended SEIVR model, which considers the effects of quarantine and vaccination while accounting for prevailing uncertainties in the observed data. In fact, the existing data and future forecasts alike are uncertain. That is, the data provided do not represent the total number of infected, recovered, and dead cases. There is always a hidden population that has not been detected and reported by the healthcare system. Therefore, the uncertainty is considered in the model parameters to reflect the uncertainty in both existing and future data. To this end, assuming eight operators of (1) susceptible, (2) insusceptible, (3) vaccinated, (4) exposed, (5) infective, (6) quarantined, (7) recovered, and (8) dead, an extended SEIVR model is established and applied to the data reported for the United State (US) as a country severely infected with COVID-19. The problem output is then reported based on three main parameters, namely, the maximum number of the infected, the total number of the recovered, and the final number of fatalities. Then, by defining the input parameters, including transmission, recovery, and mortality rates as random variables, the problem is transformed into a reliability model to compute exceedance probabilities (EP) of the output. Next, by incorporating time into the problem formulation and converting the model into a time-variant reliability analysis, the results are presented as temporally-varying probabilities.

The content of the paper is organized as follows. In Section “Extended SEIVR model”, the extended SEIVR model is formulated, and the deterministic form of the problem is presented and implemented on the official data released for the US. In Section “Probabilistic formulation of the extended SEIVR model”, the main idea behind the probabilistic definition of the extended SEIVR model is presented. In Section “Reliability model”, a reliability model is developed to account for prevailing uncertainties, which transforms the problem into a probabilistic one. In Section “Time-variant reliability analysis”, a time-variant reliability analysis is conducted and the resulting exceedance probabilities over time are presented as three-dimensional (3D) graphs and contour plots. Finally, the research is summarized and concluded in Section “Conclusion”.

Extended SEIVR model

Model formulation

The mathematical modeling of an epidemic for investigating intensity and prohibition mechanisms is a widely accepted approach that can provide decision support on control measures [11–14]. A brief overview of some aspects of the development of mathematical modeling, extraction and processing of epidemiological data can be found here [11,12]. Additionally, new mathematical models in solving disease transmission ODEs, such as the fractional method, can be found in [13,14]. Apart from COVID-19, the mathematical tools were also utilized to investigate other diseases. In this regard, Naresh et al. [15] proposed and analyzed the nonlinear mathematical model of HIV/AIDS transmission, Tan et al. [16] used the mathematical model to analyze the H1N1 influenza epidemic data, and Diaz et al. [17] analyzed the West African Ebola virus. The basic mathematical model for an epidemic spread is generally recognized as a susceptible-exposed-infected-recovered (SEIR) model. This model is employed in epidemiology to calculate the number of SEIR individuals in a population at any time [18,19]. It can be further applied to interpret the development in the number of people demanding medical treatment over the course of an epidemic [20]. The entire population is accordingly divided into four classes, i.e., $S$ as the number of susceptible, $E$ as the number of the exposed, $I$ as the number of infected, and $R$ as the number of recovered cases. Later on, by incorporating more details of the disease spreading process in the problem formulation, more accurate and comprehensive forms of the basic SEIR model were proposed by different researchers [21–23]. De la Sen et al. [21] generalized the SEIR model and created a time-varying compartmental model. Lu et al. [22] added the hospitalization compartment to the basic SEIR model, and Liu et al. [23] modified the basic SEIR model to capture the evolution trajectory of the COVID-19. Zamir et al. [42] studied non-pharmaceutical interventions (NPIs) with mathematical models. Singh et al. [38] presented a fractional form of differential equations to provide an efficient computational model for predicting the number of people infected with the disease. Danane et al. [40] introduced a white noise and Levy jump perturbations in the model compartments to construct a stochastic model and compared the results with numerical simulations. Atananga [43] adopted a fractal-fractional integral solution to predict the spread of the epidemic. Gao et al. [44] investigated the spread of the disease as a dynamic system by presenting a model called Bata-Hosts-Reservoir-People-Coronavirus (BHRPC). Gao et al. [45] used the $q$-homotopy analysis transform method ($q$-HATM) to examine unreported cases in the data and subsequently, studied the spread of the disease with updated data. Each of these methods has its advantages, but they generally solve the problem from a deterministic point of view. To fill the gap, a general framework is proposed in this paper to combine any desired deterministic method, such as the extended SEIVR model used in this study, with Monte Carlo sampling to take the model uncertainty into account and express the results probabilistically. This leads to predictions of the spread of the epidemic given a target probability level.

In this study, an extended SEIVR model, initially developed by Peng et al. [24], is utilized and extended to establish a deterministic model for the spread of COVID-19 in the US. In the definition of this model, the population is divided into eight states. First, a large portion of the population is considered susceptible ($S$), which could later become insusceptible ($P$), vaccinated ($V$), or exposed [11] ($E$). Based on social exposure, exposed cases could be transferred to the infective state [2] ($I$). If a patient is diagnosed with the disease, the patients will be quarantined ($Q$) according to the disease severity and infection. Upon treatment via hospital medical care or self-medication, most of the quarantined cases recover ($R$) and some die ($D$) due to the disease severity. The relationship between these eight states is schematically shown in Fig. 1. For simplicity, it is assumed that the recovered and vaccinated people disappear from the disease cycle and will not be infected again. In order to consider reinfection, it can be assumed that vaccinated and recovered individuals are exposed to the disease at a certain rate. Therefore, two new terms must be introduced into the differential equation, one to transform the vaccinated into the exposed, $\text{1}$ Cases who might be infected but are not yet infectious in a latent period.

$\text{2}$ Cases with infectious capacity but not yet quarantined.
Exposed is modeled as a matrix form, as follows:

\[
\begin{bmatrix}
\frac{dS}{dt} \\
\frac{dE}{dt} \\
\frac{dI}{dt} \\
\frac{dD}{dt} \\
\frac{dR}{dt} \\
\frac{dQ}{dt} \\
\frac{dP}{dt} \\
\frac{dV}{dt}
\end{bmatrix} = \begin{bmatrix}
-a - \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\gamma & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \gamma & -\delta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \delta & -\kappa(t) - \lambda(t) & 0 & 0 & 0 & 0 \\
0 & 0 & \kappa(t) & 0 & 0 & 0 & 0 & 0 \\
\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
S(t) \\
E(t) \\
I(t) \\
D(t) \\
R(t) \\
Q(t) \\
P(t) \\
V(t)
\end{bmatrix},
\]

This equation is then solved numerically using the Runge Kutta method [25,26]. Assuming that \( f(Y_n, t_n) = A(t_n) \cdot Y_n + F(t_n) \), where \( h = 110 \) and \( t_n \) is the days between Feb. 15, 2020 and March 13, 2021, Runge Kutta coefficients, i.e., \( k_1, \ldots, k_4 \), are calculated as follows:

\[
k_1 = hf(Y_n, t_n), \quad k_2 = hf(Y_n + k_1, t_n), \quad k_3 = hf(Y_n + k_2/2, t_n + h/2), \quad k_4 = hf(Y_n + k_3, t_n + h),
\]

Next, the \( Y_{n+1} \) components are obtained as time histories.

\[
Y_{n+1} = Y_n + 1/6(k_1 + 2k_2 + 2k_3 + k_4),
\]

As can be seen, the recovery rate (\( \lambda \)) and the mortality rate (\( \kappa \)) are functions of time. This enables the algorithm to use the updated values of \( \lambda \) and \( \kappa \) at any point in time to better incorporate the system dynamics. For this purpose, the following exponential functions are used to model \( \kappa \) and \( \lambda \) [24,27],

\[
\kappa(t) = \kappa_0 + \exp(-\kappa_1 \cdot (t + t_1)), \quad \lambda(t) = \frac{\lambda_0}{1 + \exp(-\lambda_1 \cdot (t - t_2))}
\]

where \( \kappa_0, \kappa_1, \tau_1, \lambda_0, \lambda_1, \) and \( t_1 \) are model parameters to be determined via regression. Here, parameters \( \kappa_0, \kappa_1, \lambda_0, \) and \( \lambda_1 \) have the dimension of the inverse time, and \( \tau_1 \) and \( t_1 \) have the dimension of time. It is also of note that parameters \( \kappa \) and \( \lambda \) converge to \( \kappa_0 \) and \( \lambda_0 \) over time.

**Model implementation**

After establishing the problem formulation, data collection and model analysis are performed. For this purpose, based on the online dataset provided by Johns Hopkins University to monitor the worldwide spread of COVID-19 [28], the data of infected (active), recovered, and fatal cases in the US were collected up until March 13, 2021.
In addition to time-dependent variables, i.e., $x(t)$ and $j(t)$, there are five parameters, i.e., $\alpha, \beta, \rho, \gamma$, and $\delta$, in the extended SEIVR model that must be determined according to the actual data collected. For this purpose, a regression problem was established in which the optimal values of the parameters were calculated by minimizing the Root Mean Square Error (RMSE) and fitting the extended SEIVR model to the data. That is, a cost function is based on the RMSE between the observed data and model predictions. Thereafter, the cost function is minimized so as to find the optimal value of the parameters that lead to the minimum prediction error. The results are reported in Table 1. These values are the best fit for the US data. In the case of using the proposed SEIVR model for other countries and regions, different optimal parameters would be obtained.

The actual recovery and mortality rates and the functions fitted to them are shown in Fig. 2. Furthermore, the extended SEIVR model fitted to the active, as well as recovered and fatal cases are illustrated in Fig. 3. The grey area in Fig. 3 shows the actual data as of March 13, 2020 and the solid line represents the forecast data. It is estimated that the number of active cases will reach about 0.8 million by Jan. 01, 2022. Also, the maximum number of people recovered reaches about 41.3 million and the total number of fatalities would be as high as 0.6 million cases.

**Probabilistic formulation of the extended SEIVR model**

As briefly described in the Introduction, the deterministic SEIVR model suffers from two major drawbacks. First, it usually is not a precise representation of the future observations. Second, the model fails to demonstrate how reliable its predictions are, even if the model can provide a perfect fit to past observations. The data themselves are uncertain as they merely account for the cases that are detected by healthcare centers. In fact, there is always a hidden infected part that is not detected and not reflected in the data.

This study proposes a method to boost the reliability of the SEIVR model in predicting the future state of the epidemic. Therefore, the hypothesis is not to make use of a particular deterministic SEIVR model to predict the epidemic state but to quantify the uncertainty and establish a feasible range within which the answer can follow a normal distribution (Fig. 4). Thus, the highest probability density falls within the region wherein the actual data is located. Of note, the upper and lower bounds need not be precisely determined upon assuming the distribution function. In fact, the assumed probability distribution function is unbounded but the probability at tails is extremely small, which makes the realization of events in these regions improbable.

By combining the proposed model with the Monte Carlo sampling, the SEIVR model is subjected to analysis repeatedly, given the realization of random variables. By generating enough samples, the sample space is well-covered. This process can be summarized in three main steps:

1. A normal distribution function is assumed for the SEIVR model parameters, i.e., $\alpha, \beta, \gamma, \delta, k_0$, and $\lambda_0$, based on the values obtained from the deterministic model. Then, $n$ random samples are generated using the assumed distribution function.
2. The SEIVR model is run for each random sample and its results are stored in a response matrix.
3. Using the data obtained in Step 2, the EPs for the infected ($I$), recovered ($R$), and dead ($D$) cases are computed at the desired point in time and a probability contour plot is then drawn to represent the results probabilistically.

**Reliability model**

The main output of the extended SEIVR model, i.e., the number of $I$, $R$, and $D$ cases, is notably contingent upon model parameters, i.e., $\alpha, \beta, \gamma, \delta, k_0$, and $\lambda_0$. In Section “Model implementation”, the extended SEIVR model showed that the values of $\alpha = 6.4 \times 10^{-2}, \beta = 4.049, \rho = 4 \times 10^{-3}, \gamma = 8 \times 10^{-3}, \delta = 6.6 \times 10^{-2}, k_0 = 7.997 \times 10^{-4}$, and $\lambda_0 = 1.6 \times 10^{-2}$ leads to the best fit of the model with the actual data from the US. In this step, however, instead of assuming the model parameters as deterministic values, they are considered as random variables except $\rho$ which is considered constant. The probabilistic characteristics of these variables are given in Table 2.

As illustrated, the mean values of the variables are here assumed equal to the best-fit values. This means that the highest probability of occurrence is assigned to the output of the SEIVR analysis. Also, a coefficient of variation of 10% is considered for each variable to account for the other scenarios that arise due to the uncertainty of these parameters. In the following, three limit-state functions are defined as shown below:

\[
g_1(x, t) = I_t - I(x, t),
\]

\[
g_2(x, t) = R_t - R(x, t),
\]

\[
g_3(x, t) = D_t - D(x, t),
\]

where $I(x, t)$, $R(x, t)$, and $D(x, t)$ represent the number of infected, recovered, and dead cases obtained from the SEIVR analysis as a function of the vector of random variables $x = \{\alpha, \beta, \gamma, \delta, k_0, \text{and} \lambda_0\}$ and time, and $I_0$, $R_0$, and $D_0$ are their corresponding thresholds, shown in Fig. 3 [3]. The occurrence of $g(x, t) \leq 0$ also suggests that the prediction of the SEIVR model exceeds the threshold. Therefore, the EP can be defined as:

\[
p_{I_t} = P(\{g_1(x, t) \leq 0\}),
\]

\[
p_{R_t} = P(\{g_2(x, t) \leq 0\}),
\]

\[
p_{D_t} = P(\{g_3(x, t) \leq 0\}),
\]

To solve the established reliability problem, various solutions have been proposed in the literature [29,30], including the Monte Carlo sampling method used in this study [31-33]. To perform the analysis, 20,000 random samples are generated for each random variable in accordance with the probability distributions in Table 2. The histograms of the generated samples are shown in Fig. 5. Next, the limit-state functions in Eqs. (20)-(22) are calculated by conducting an SEIVR analysis for each realization of the underlying random variable. The histogram of samples generated for each limit-state function is shown in Fig. 6.

The probability $p_f$ is obtained by solving the following multifold integral:

\[
p_f = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \cdots dx_n,
\]

Table 1

| Parameter | Value | Parameter | Value |
|-----------|-------|-----------|-------|
| $\alpha$  | 0.064 | $\gamma$  | 0.008 |
| $\beta$   | 4.049 | $\delta$  | 0.066 |
| $\rho$    | 0.004 | $k_0$     | 0.016 |
| $\kappa$  | 7.997 x 10^{-4} | $\lambda_0$ | 0.032 |
| $\epsilon$ | 0.0165 | $\tau_1$ | 160.788 |
| $\tau_1$  | 68.485 | –         | –     |
where \( \psi(x) \) = step function that equals unity when \( g(x) \leq 0 \) and zero otherwise. It follows that \( p_f \) is the expectation of \( \psi(x) \) with respect to distribution \( f_X(x) \); hence
\[
p_f = \frac{1}{N} \sum_{k=1}^{N} \psi(x_k), \tag{27}
\]
where \( N \) denotes the total number of samples, here 20,000, and \( x_k \) represents the vector of realizations of \( x \) in the \( k \)th sample. The three probabilities are obtained as \( p_f = 49.35\% \), \( p_s = 48.8\% \), and \( p_d = 48.85\% \).

To assess the accuracy of results, it is of utmost importance to ensure that the Monte Carlo sampling is converged. In other words, there is a need to confirm that the number of random samples utilized in the Monte Carlo sampling is sufficient and that the resulting probability changes only negligibly by adding further random samples. Therefore, the probability is calculated for different sample sizes and then plotted in a graph to demonstrate the convergence trend. The results are illustrated in Fig. 7 for all the three limit-state functions. As observed, upon using 5,000 random samples, the resulting probabilities for all three limit-state functions are well converged and the subsequent fluctuations are negligible.
The EPs of $I$, $R$, and $D$ were calculated for the assumed thresholds. However, a more comprehensive expression of the problem is to calculate the EP for a range of thresholds for $I_0$, $R_0$, and $D_0$. Therefore, the parameters $I_0$, $R_0$, and $D_0$ are defined as decision variables, which can take various values. An external loop is added to the Monte Carlo sampling to set the parameters $I_0$, $R_0$, and $D_0$ and to introduce them into the algorithm. Accordingly, by each run of the loop, the EP corresponding to a particular value of $I_0$, $R_0$, and $D_0$ is calculated and stored in vectors. The results for a sufficiently wide range of $I_0$, $R_0$, and $D_0$ thresholds are shown in Table 3. By conducting the analysis for a large number of $I_0$, $R_0$, and $D_0$ thresholds, the EP is plotted against $I_0$, $R_0$, and $D_0$ in Fig. 8. As seen, EP sharply decreases as $I_0$, $R_0$, and $D_0$ are increased and the probability falls below 5% for $I_0 > 1.4 \times 10^6$, $R_0 > 5.79 \times 10^7$, and $D_0 > 8.18 \times 10^5$, which indicates that such events are unlikely.

From a practical point of view, the exceedance probability curves provided in Figure 8 can be used in estimating the spread of the epidemic, and facilitate informed decisions to control it. To this end, a target probability that represents the accepted risk must be considered. Then, the assumed probability level can be used to determine the number of the infected, the recovered, and the dead. For example, by accepting a probability of 10% for exceeding the forecast, the number of the infected, the recovered, and the dead is estimated at $1.25 \times 10^6$, $5.5 \times 10^7$, and $7.7 \times 10^5$, respectively, using the graphs provided in Figure 8. If the
target probability is considered as a smaller value, a more conservative estimate is obtained, indicating more people will grapple with the disease.

**Vaccination effect**

In many countries, including the US, public vaccination is being pursued as the primary vehicle for controlling the COVID-19 outbreak. However, the effect of vaccination largely depends on the vaccination rate, which in turn depends on the national healthcare infrastructure, facilities, and policies. The previous section assessed the state of the pandemic in early 2022 considering the current vaccination rate in the US. This is implemented by assuming “susceptible” cases as the target population of vaccination. This part of the population was then vaccinated at a specified rate, $\rho$. Vaccinated people are deemed to have enough immunity to get out of the disease cycle. The parameter $\rho$ is not regarded as a random variable, because it assumes that the data provided are accurate statistics of the number of vaccinations with negligible uncertainty. The optimal value of the parameter $\rho$ shown in Table 1 is obtained by fitting the proposed SEIVR model to the US data. In this section, the state of the pandemic is investigated with vaccination rates falling or rising compared to the current situation. For this purpose, reliability sensitivity analysis is employed [34]. In fact, by defining vaccination as a decision variable, the change in the exceedance probability of the maximum number of active cases in early 2022 is investigated. Seven scenarios are considered in which the vaccination rate is assumed to be 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 4.0 times that of the current situation. Then, an individual reliability analysis is performed for each scenario and the exceedance probability curve is extracted. The results are shown in Fig. 9. This figure shows that increasing the vaccination rate sharply decreases the number of active cases. For example, the number of active cases at an exceedance probability of 5% for vaccination rates of 0.5, 1.0, 2.0, and 4.0 are 2.0, 1.4, 0.5, and 0.2 million, respectively. This means that, assuming a 5% exceedance probability, the US needs to quadruple the vaccination rate to reduce the number of active cases to less than 0.2 million by early 2022.

**Time-variant reliability analysis**

In the previous section, the probability diagrams were extracted at a single point in time. However, since the results of the SEIVR model

| $n$ | $I_0 \times 10^6$ | Probability (%) | $R_0 \times 10^7$ | Probability (%) | $D_0 \times 10^5$ | Probability (%) |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|
| 1   | 0.2            | 100            | 2              | 100            | 3.5            | 100            |
| 2   | 0.4            | 99.54          | 2.8            | 96.70          | 4.3            | 98.03          |
| 3   | 0.6            | 86.58          | 3.6            | 75.14          | 5.1            | 83.59          |
| 4   | 0.8            | 52.99          | 4.4            | 37.22          | 5.9            | 54.42          |
| 5   | 1              | 26.50          | 5.2            | 13.39          | 6.7            | 27.80          |
| 6   | 1.2            | 11.82          | 6              | 3.38           | 7.5            | 11.90          |
| 7   | 1.4            | 4.87           | 6.8            | 0.99           | 8.3            | 4.12           |
| 8   | 1.6            | 2.10           | 7.6            | 0.27           | 9.1            | 1.46           |
| 9   | 1.8            | 0.80           | 8.4            | 0.05           | 9.9            | 0.48           |
| 10  | 2              | 0.30           | 9.2            | 0.05           | 10.7           | 0.20           |
| 11  | 2.2            | 0.10           | 10             | 0.05           | 11.5           | 0.05           |
| 12  | 2.4            | 0.05           | 10.8           | 0              | 12.3           | 0.02           |

Table 3 Results of reliability analysis for different values of $I$, $R$, and $D$.
analysis are a function of time, this section calculates the EP for an entire timespan. To implement the time-variant model, the continuous time variable is broken down into a limited number of discrete points using a fixed time step. This discretization simplifies the problem because it enables the algorithm to treat time as a constant at any point in time. Employing this technique, time-variant reliability analysis turns into a series of time-invariant models. Then, each of these reliability problems is separately analyzed using the algorithm described in the previous sections and the result of each one is stored in the form of a probability diagram.

To this end, a second external loop is added to the algorithm, which sets a new value for time in each run of the loop and feeds it into the inner loop. Then, according to the time introduced into the system, the inner loop is used to calculate the probability curve and store it in a vector. Afterwards, this process continues until the probability graph covers the entire timespan under consideration. As such, time is added to the graph as the third dimension. A 3D graph is hence produced that illustrates time, probability, and either of the infected, recovered, or fatal cases in a single plot. This process is schematically shown in Fig. 10. The resulting 3D graphs and contour plots are plotted in Fig. 11. In this figure, a color scale code is utilized to demonstrate the probability level.

The advantage of this diagram over the deterministic graphs presented previously is that the results cover a wide range of possible scenarios. This increases the reliability of the answer because the uncertainty in the data is reflected in the output of the model. In addition, the graph is presented in the form of a probability distribution over time. That is to say, at any given time, there is a probability of exceedance for any state of disease transmission. For example, suppose the number of infected cases after 400 days is of interest. Fig. 11b shows that at \( t = 400 \), there are 90%, 44%, and 14.6% probabilities that the number of infected cases exceeds 1, 1.5, and 2 million, respectively. That is, the results are no longer limited to a deterministic prediction in time, and each prediction corresponds to a certain probability level. Therefore, based on the target reliability, the probabilistic output can be used to investigate the spread of the disease in the country.

Conclusion

This study aimed to provide a probabilistic solution to predict the spreading profile of COVID-19 in the society while accounting for prevailing uncertainties in the observed data. To this end, an extended SEIVR model was first developed using the available data from the US as a case study to predict the number of the infected, recovered, and fatal cases over time. Then, by defining the model parameters including transmission, recovery, and mortality rates as random variables, the model was reformulated as a reliability problem, and the deterministic SEIVR model was transformed into a probabilistic problem. Next, Monte Carlo sampling was utilized to calculate the EP for three parameters of the final number of fatalities, the final number of recovered cases, and the maximum number of the infected cases. The problem was next redefined as a time-variant reliability analysis to obtain the probability distributions of the said parameters over time.

The main conclusions drawn from this study are as follows:

- The deterministic results of the extended SEIVR model reveal that, in view of the current circumstances, the descending trend in the number of active infected cases in the US will continue until Jan 01, 2022, when it will reach about 0.8 million. In addition, after accounting for the uncertainties in the model and conducting a probabilistic analysis, it was observed that the number of active infected cases can exceed this value. For example, there is a 20% probability for reaching to 1.1 million active infected cases.

- The results of the deterministic analysis show that the number of the recovered cases gradually increases to nearly 41.34 million. The probabilistic analysis of the model demonstrates the probability that the number of the recovered cases exceed 50 million is about 20%.

- The results of the deterministic SEIVR model indicate that the number of fatalities will reach nearly 0.6 million. Probabilistic analysis reveals an exceedance probability of 20% for 0.7 million fatalities.

- The results of reliability sensitivity analysis show that the number of active cases decreases sharply by increasing the vaccination rate. For instance, quadrupling the vaccination rate reduces the number of active cases to less than 0.2 million by early 2022.

It is emphasized that the results of this study are based on the current state of COVID-19 pandemic in the US. In the case of events that alter the current circumstances, such as enforcing stricter social distancing or wider public vaccinations, the results will also be altered.

Author statement

Mahdi Shadabfar: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review editing; Mojtaba Mahsuli: Investigation, Methodology, Supervision, Validation, Writing - original draft, Writing - review editing; Arash Sioofy Khoojine: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review editing; Vahid Reza Hosseini: Data curation, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing - original draft.
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The first author gratefully acknowledges Iran’s National Elites Foundation for a postdoctoral fellowship at Sharif University of Technology. The second author also appreciatively thanks the Deputy of Research and Technology of Sharif University of Technology for supporting this research.

References

[1] Zeng Y, Guo X, Deng Q, Luo S, Zhang H. Forecasting of covid-19: spread with dynamic transmission rate. J Saf Sci Resilience 2020;1:91–6. https://doi.org/10.1016/j.jnlssr.2020.07.005.

[2] Lin C, Lau AK, Fung JC, Guo C, Chan JW, Yeung DW, Zhang Y, Bo Y, Hossain MS, Zeng Y, Luo XQ. A mechanism-based parameterisation scheme to investigate the association between transmission rate of covid-19 and meteorological factors on plains in china. Sci Total Environ 2020;737:140348. https://doi.org/10.1016/j.scitotenv.2020.140348.

[3] Lin J, Huang W, Wen M, Li D, Ma S, Hua J, Hu H, Yin S, Qian Y, Chen P, Zhang Q, Yuan N, Sun S. Containing the spread of coronavirus disease 2019 (covid-19): meteorological factors and control strategies. Sci Total Environ 2020;744:140935. https://doi.org/10.1016/j.scitotenv.2020.140935.

[4] Eikenberry SE, Mancuso M, Iboi E, Pankavich S, Eikenberry K, Kuang Y, Kostelich E, Gumel AB. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. Infect Disease Model 2020;5:293–308. https://doi.org/10.1016/j.idm.2020.04.001.

[5] Tang L, Zhou Y, Wang L, Parkayastha S, Zhang L, He J, Wang F, Song PX.K. A review of multi-compartment infectious disease models. Int Stat Rev 2020;88:462-513. https://doi.org/10.10111/insr.12402.

[6] Mohamadou Y, Halidou A, Kpen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. Appl Intell 2020;50:3913–25. https://doi.org/10.1007/s10489-020-01770-9.

[7] Kudryashov NA, Chmykhov MA, Viporovitch M. Analytical features of the SIR model and their applications to COVID-19. Appl Math Model 2021;90:466–73. https://doi.org/10.1016/j.apm.2020.08.057.

[8] Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. Clin Microbiol Infect 2013;19:999–1005. https://doi.org/10.1111/1469-0691.12308.

[9] Alberti T, Faranda D. On the uncertainty of real-time predictions of epidemic growths: a covid-19 case study for china and Italy. Commun Nonlinear Sci Numer Simul 2020;90:105372. https://doi.org/10.1016/j.cnsns.2020.105372.

[10] Chen-Charpentier BM, Stanescu D. Epidemic models with random coefficients. Math Comput Model 2010;52:1004–10. https://doi.org/10.1016/j.mcm.2010.01.014.

[11] Brauer F. Mathematical epidemiology: past, present, and future. Infect Disease Model 2017;2:113–27. https://doi.org/10.1016/j.idm.2017.02.001.

[12] Walters CE, Meslé MM, Hall IM. Modelling the global spread of diseases: a review of current practice and capability. Epidemics 2018;25:1–8. https://doi.org/10.1016/j.epidem.2018.05.007.

[13] Baleanu D, Mohammadi H, Rezapour S. A fractional differential equation model for the COVID-19 transmission by using the Caputo-Fabrizio derivative. Adv Differ Equ 2020;299:1–27. https://doi.org/10.1186/s13662-020-02762-4.

[14] Sweilam N, Al-Mekhlafi S, Baleanu D. A hybrid fractional optimal control for a novel Coronavirus (2019-nCoV) mathematical model. J Adv Res 2020. https://doi.org/10.1016/j.jare.2020.08.006.

[15] Naresh R, Tripathi A, Sharma D. Modelling and analysis of the spread of aids epidemic with immigration of HIV infectives. Math Comput Model 2009;49:880–92. https://doi.org/10.1016/j.mcm.2008.09.013.

[16] Tan X, Yuan L, Zhou J, Zheng Y, Yang F. Modeling the initial transmission dynamics of influenza a H1N1 in Guangdong Province, China. Int J Infect Dis 2013;17:e479–84. https://doi.org/10.1016/j.ijid.2012.11.018.

[17] Diaz P, Constantine P, Kalmbach K, Jones E, Pankavich S. A modified seir model for the spread of ebola in western africa and metrics for resource allocation. Appl Math Comput 2018;294:141–55. https://doi.org/10.1016/j.amc.2017.11.039.
[18] He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlinear Dyn 2020;101:1667–80. https://doi.org/10.1007/s11071-020-05743-y.

[19] Piovella N. Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic. Chaos Solitons Fractals 2020;140:110243. https://doi.org/10.1016/j.chaos.2020.110243.

[20] Giordano G, Blanchini F, Bruno R, Colaneri P, Filippo AD, Matteo AD, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020;26:855–60. https://doi.org/10.1038/s41591-020-08833-7.

[21] la Sen MD, Agarwal RP, Ibeas A, Alonso-Quesada S. On a generalized time-varying SEIR epidemic model with mixed point and distributed time-varying delays and combined regular and impulsive vaccination controls. Adv Differ Equ 2010;1:1–42. https://doi.org/10.1155/2010/281612.

[22] Chen Y, Chen Y, Ren G, Xu C, Wang S, Yin Z. A fractional-order SEHDR model for COVID-19 with inter-city networked coupling effects. Nonlinear Dyn 2020;101: 1717–30. https://doi.org/10.1007/s11071-020-05848-4.

[23] Liu M, Ning J, Du Y, Cao J, Zhang D, Wang J, Chen M. Modelling the evolution trajectory of COVID-19 in Wuhan, China: experience and suggestions. Public Health 2020;183:76–80. https://doi.org/10.1016/j.puhe.2020.05.001.

[24] Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. medRxiv 2020. https://doi.org/10.1101/2020.02.16.20023465.

[25] Mwalili S, Kimathi M, Ojiambo V, Gathungu D, Mbogo R. SEIR model for COVID-19 dynamics incorporating the environment and social distancing. BMC Res Notes 2020;13:1–5. https://doi.org/10.1186/s13104-020-05192-1.

[26] Kovalnogov VN, Simos TE, Tsitouras C. Runge–Kutta pairs suited for sir-type epidemic models. Math Methods Appl Sci 2021;44:5210–6. https://doi.org/10.1002/mma.7104.

[27] Cheynet E. Generalized SEIR Epidemic Model (fitting and computation) 2020. https://doi.org/10.5281/zenodo.3911854.

[28] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4. https://doi.org/10.1016/S1473-3099(20)30120-1.

[29] Didiveev O, Maden H. Structural reliability methods. Chichester, UK: John Wiley and Sons; 2007.

[30] Choi S-K, Grandhi RV, Canfield RA. Reliability-based structural design. 1st ed. London: Springer-Verlag; 2007.

[31] Zio E. The Monte Carlo simulation method for system reliability and risk analysis. London: Springer-Verlag; 2013.

[32] Far MS, Wang Y. Approximation of the monte carlo sampling method for reliability analysis of structures. Math Problems Eng 2016;1:1–9. https://doi.org/10.1155/2016/5726565.