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.
A study on the efficiency of the estimation models of COVID-19

Mohammed N. Alenezi\textsuperscript{a}, Fawaz S. Al-Anzi\textsuperscript{b,∗}, Haneen Alabulrazzaq\textsuperscript{a}, Ammar Alhusaini\textsuperscript{b}, Abdullah F. Al-Anzi\textsuperscript{c}

\textsuperscript{a} Computer Science & Information Systems Department, Public Authority for Applied Education & Training, Kuwait
\textsuperscript{b} Computer Engineering Department, Kuwait University, Kuwait
\textsuperscript{c} Electrical & Computer Engineering Department, American University, Kuwait

\textbf{ABSTRACT}

Today, the world is fighting against a dangerous epidemic caused by the novel coronavirus, also known as COVID-19. All have been impacted and countries are trying to recover from the social, economic, and health devastations of COVID-19. Recent epidemiology research has concentrated on using different prediction models to estimate the numbers of infected, recovered, and deceased cases around the world. This study is primarily focused on evaluating two common prediction models: Susceptible – Infected – Recovered (SIR) and Susceptible – Exposed – Infected – Recovered (SEIR). The SIR and SEIR models were compared in estimating the outbreak and identifying the better fitting model for forecasting future spread in Kuwait. Based on the results of the comparison, the SEIR model was selected for predicting COVID-19 infected, recovered, and cumulative cases. The data needed for estimation was collected from official sites of the Kuwait Government between 24 February and 1 December 2020. This study presents estimated values for peak dates and expected eradication of COVID-19 in Kuwait. The proposed estimation model is simulated using the Python Programming language on the collected data. The simulation was performed with various basic reproduction numbers (between 5.2 and 3), the initial exposed population, and the incubation rate. The results show that the SEIR model was better suited than the SIR model for predicting both infection and recovery cases with \( R_0 \) values ranging from 3 to 4, \( E_0 = 80 \) and \( \alpha = 0.2 \).

Introduction

The world is now undergoing a challenging period due to an unparalleled spread of an infectious virus, named coronavirus or COVID-19. The effects from the family of coronaviruses can range from a simple cold to more dangerous forms similar to Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). COVID-19, also in this family, was reported on 31 December 2019 in Wuhan, China. Due to the expeditious transmission of the virus, the Chinese Government implemented several containment measures to control this outbreak from spreading, including a complete lockdown and suspension of all forms of transportation to Wuhan during late January 2020 [1]. Thailand was the second country to report cases on 13 January 2020 [2]. COVID-19 has now spread to 218 countries and territories. As of 6 December 2020, nearly 66.6 million infected cases, 42.8 million recovered cases, and 1.53 million deceased cases were reported due to COVID-19 [3].

On 30 January 2020, the total world count of confirmed cases reached 8,096. COVID-19 was declared an epidemic and a global public health emergency by the World Health Organization (WHO) [4]. With its rapid spreading becoming a serious challenge across the world, the WHO announced COVID-19 as a global pandemic on 11 March 2020 [5]. Most impacted countries imposed a full or partial lockdown to control the spread. One of the best measures to be implemented to reduce the possibility of spreading is social distancing. This involves people maintaining a distance of 1 meter, wearing masks covering the mouth and nose, and gloves to limit transmission. The pandemic quickly impacted various areas such as economics, education, politics, etc., along with public health all over the world. Furthermore, we saw an increase in poverty and unemployment rates.

When considering how to best control the spread of COVID-19, the virus’s vigorous infectious behavior, ambiguity in transmission methods, expanded incubation period, complications in detecting the virus, etc., must be taken into account. Most countries began joint efforts...
to prevent the transmission and eradication of the virus [6]. Pharmaceutical companies, research departments, and health departments are developing multiple vaccines and treatments to prevent and bolster COVID-19 recoveries. There are now vaccines in the later stages of development and testing.

Kuwait faces a serious threat due to the novel coronavirus, and its impact is seen everywhere. COVID-19 was first discovered in Kuwait on 24 February 2020; the first five cases were related to citizens who returned from abroad. Trying to limit localized infection rates, the Kuwait Government imposed precautionary measures such as quarantines, banning flights to and from selected countries, closing down retail shops, announcing a public holiday from 12 March 2020 onward. A partial curfew was implemented on 22 March 2020 between 5:00 PM and 4:00 AM daily, and the curfew timings were changed twice (from 5:00 PM to 6:00 AM and again from 4:00 PM to 8:00 AM). Finally, the government issued a complete lockdown from 10 May 2020 to 30 May 2020 [7].

In Kuwait, 142,993 individuals had been infected with COVID-19 as of 1 December 2020, out of which 138,507 had recovered, 881 deceased, and 3,605 currently receiving treatment. As of 1 June 2020, the population of Kuwait was approximately 4,776,407, according to data collected from the Public Authority for Civil Information (PACI) [8]. The Ministry of Health (MOH) and a small number of private hospitals and clinics are Kuwait’s primary healthcare providers. The entire bed capacity is around 8,200, almost 7,118 from MOH and 1,082 from private sector hospitals [9].

It is essential to generate an accurate estimation model for forecasting the impacts of COVID-19 in all fields like health, social, and economics for supporting the decision-makers to make appropriate mandates to handle these uncertainties. The estimation model helps to forecast what should happen in the near future. These predictions will support authorities to take precautionary and preventive measures like arranging sufficient care and treatment plans. Due to the nature of pandemics, it is not possible to make precise estimates. However, researchers and scientists tried to model this pandemic based on some scientifically proven estimation methods. According to the estimation
The SIR and SEIR estimation models are two simple and effective compartmental models used for modeling a pandemic. The Susceptible-Infected-Recovered model (SIR) is a popular model used for forecasting a pandemic [10]. The SIR model considers the total population as three compartments: Susceptible population, Infected population, and Recovered population. The infected population is referred to as the total number of individuals infected with COVID-19 and are capable of spreading the virus to others. The people who are all susceptible to infection but not yet infected are in the Susceptible compartment. Those who are deceased or recovered are in the Recovered compartment. These phases are considered as the three progressive phases of an epidemic. The SIR model is effective for estimating the percentage of the population who might need medical support. Based on the SIR estimation model, an individual who has recovered from the virus has acquired immunity and will not be reinfected.

The SEIR model is an extension of the SIR model which introduced the compartment labeled Exposed to the already existing SIR. Those exposed to infected individuals and who did not become infected are categorized as an Exposed population [11]. SEIR considers the entire population under the susceptible compartment. It also considers the people who recovered have acquired a lifelong immunity against the virus.

Here, we are going to model the COVID-19 outbreak using the SIR and SEIR models, analyze and compare both in terms of accuracy to find the best model for predicting the outbreak using Kuwait as a case study. After selecting the most relevant model, it will be used to predict the COVID-19 spread in Kuwait with different basic reproduction numbers ($R_0$). For estimating the epidemic, we assumed that the country has a constant population during the estimation period (ignoring deaths, births, and migration during the estimation period). This research is mainly done to identify the best estimation method for modeling the COVID-19 pandemic considering the infected, recovered, and cumulative cases based on various $R_0$ values, which will be explained in the upcoming sections. All required information needed for this study is collected from government authorized sites during the period of 24 February to 1 December 2020 [7]. Our study is also trying to identify the
best values for peak dates and the expected decline of this pandemic in Kuwait.

The rest of the paper is organized as follows. The “Background” section discusses various estimation models available for modeling the pandemics. In section “SEIR-based prediction”, we describe the SEIR model, a deterministic compartmental model, which we used to model the COVID-19 pandemic in this research. A comparison of the two popularly used deterministic compartmental models, SIR and SEIR, is performed and illustrated in section “Comparison of SIR and SEIR”. Finally, the analysis and conclusions of the results are provided in sections “Results and discussion”, and “Conclusion”, respectively.

Background

Regression models

Antonio Guterres, the UN Secretary-General, pronounced COVID-19 as the most hazardous calamity since the Second World War. The coronavirus outbreak created a frightening global predicament. It severely impacted the daily lives of people around the world. It has also affected the economic, health, social, and political aspects in every impacted country. Most countries recommended imposing travel bans and other restrictions; large-scale quarantines were set up all over the world as an attempt to impede the COVID-19 spread.

Many researchers modeled and forecasted the COVID-19 pandemic. Most of the studies were focused on tracing the spread of COVID-19 to analyze and predict its infection rate, recovery rate, and expected eradication. For evaluating and forecasting the COVID-19 pandemic, researchers used various models like deterministic compartmental models (DCM), agent-based models (ABM), and logistic growth models [12]. Although there are a variety of models available for forecasting various pandemics, statistical models produce better results. Regression-based models are of a different type of model, and all of them are used by various researchers in their studies. Linear regression of various orders (2nd, 3rd, and so on), Locally Weighted Linear Regression (LOESS), Generalized Linear Model (GLM), Poisson, and logistic regression, are all examples of various regression models. Regression-based methods are mainly founded on the number and type of independent and dependent

Fig. 5. Comparison of actual and estimated cumulative infection between 24th February 2020 and 28th May 2020 using SEIR model.

Fig. 6. Comparison of actual and estimated cumulative recovery between 24th February 2020 and 28th May 2020 using SIR model.
parameters and the shape of the regression curve.

Regression analysis is the most common method used to predict and analyze the relationship between two or more dependent and independent variables. Independent variables can be used for estimating the target or dependent variable using previous values. Regression analysis is used to analyze the relationship between those variables and predict future values based on this analysis. The regression model can be linear or nonlinear. If the model contains only linear parameters, then it is known as linear regression. We can also represent a polynomial regression as linear regression. The relationship between the dependent and independent variables in a polynomial regression of second-order (also known as a second-order model or quadratic model) with one explanatory variable is given by Eq. (1) and with two variables is explained in Eq. (2), respectively.

\[
y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \\
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_11 x_1^2 + \beta_22 x_2^2 + \epsilon
\]

Where \(x, x_1, \) and \(x_2\) are explanatory variables or features, and \(y\) is the dependent variable. The value of the dependent variable, \(y\) may be the number of death cases, recovered cases, confirmed cases, etc. \(x\) is used to represent features like gender, region, age, number of tests conducted,
Fig. 10. Evaluation results of SEIR and SIR based on RSS values.

Fig. 11. The daily cases of infection from 29th May 2020 to 1st December 2020.

Fig. 12. The daily cases of recovery from 29th May 2020 to 1st December 2020.
etc. $\beta_0$, $\beta_1$, and $\beta_2$ are constants which represent bias or intercept ($\beta_0$), slopes or weights ($\beta_1$, and $\beta_2$). $\epsilon$ represents the possible error obtained in this model since in any real-world situations, the regression model may not be able to estimate the correct target value, and it signifies the noise in these relationships. In the quadratic model, the linear effect parameter and quadratic effect parameter are $\beta_1$ and $\beta_2$ respectively. When $x = 0$, the value of $y$ gives the intercept value($\beta_0$).

Eq. (3) represents the relationship between independent and dependent variables using a third-order polynomial regression model. The relationship between the target and independent variables would be non-linear or curvilinear in the polynomial regression model.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$  \hspace{1cm} (3)

Where $\beta_1$, $\beta_2$, $\beta_3$, ..., $\beta_p$ are coefficients of variables. The polynomial regression can be represented as a linear regression with multiple explanatory variables when $x^i$ is represented by $x_i$ for all $i (1, 2, 3, \ldots)$.

Consider the third-order polynomial regression equation of one independent variable explained in 3 can be viewed as a linear regression model with three independent variables as explained in 4. $x^i, x^2$, and $x^3$ are rewritten as $x_1, x_2$, and $x_3$ respectively.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$  \hspace{1cm} (4)

A regression model that used to analyze and predict the influence of numerous continuous independent parameters on various target parameters is referred to as the Generalized Linear Model (GLM) [13]. It can be represented as a suitable generalization of already available regression models such as linear or polynomial. GLM is not able to predict the exact value of the target value using explanatory variables. The difference is calculated as a possible error in this case also. The main components of GLM are a random component, a link function $g(\cdot)$, and a linear predictor. The conditional distribution of target parameters $Y_i$ is referred to as a random component. Eq. (5) represents a linear model.
relationship between independent and dependent parameters, which is viewed as a linear predictor. The value of $i$ will be 1, 2, ..., $n$, and a link function which represents how the mean $E(Y_i) = \mu_i$ relates to the linear predictor:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \epsilon_i$$  \hspace{1cm} (5)

A non-parametric regression model used for smoothening the regression curve or line in volatile time-series is known as Locally Weighted Linear Regression (LOESS) [14]. The scatter plot is used for getting the best fitting data. The regression curve is smoothened with the help of local subsets. The first step in the LOESS method is to identify a smoothing parameter. After identifying a parameter, the model selects $k$ nearest neighbors of an independent variable to be smoothened ($x_0$). LOESS algorithm is applied to each point of $x_0$, which reassigns the weights to its nearest neighbors.

A regression model used for predicting the discrete dependent or response parameter is known as Poisson regression [15]. This model’s main assumption is that the response variable is considered positive counts, and it follows the Poisson distribution. This model is mainly applicable for analyzing the rates having positive counts as values. It is similar to Logistic regression, which is mainly used for calculating ratios having values between 0 and 1.

The logistic regression model is also a regression model used to predict or analyze the target or response parameter using explanatory or independent parameters under consideration [16]. The logistic regression model is best suited for analyzing and estimating the growth of pandemic or epidemic diseases. The model assumes that the epidemics are growing exponentially in the initial stage, then it reaches a steady increase phase and diminishes its rate of growth. The logistic regression-based epidemic model calculates the count of infected cases using Eq. (6) having $C(0) = C_0$ as the initial condition [16].

$$C = \frac{K}{1 + \left(\frac{K - C_0}{C_0}\right)e^r}$$  \hspace{1cm} (6)

where $C$, $r$, and $K$ defines the count of infected people, infection rate, and final epidemic size respectively. The rate of change of infection at time $t$ is calculated using Eq. (7) [16].

$$\frac{dC}{dt} = Cr\left(1 - \frac{C}{K}\right)$$  \hspace{1cm} (7)

The estimation of time at which the epidemic reaches its maximum growth rate of $t_e$ is explained in (8).
Results in Physics 26 (2021) 104370

\[ t_p = \frac{\ln \left( \frac{K}{C_0} \right)}{r} \tag{8} \]

Eqs. (9) and (10) estimate the values of peak count of infected cases and the maximum rate of growth at peak period.

\[ C_p = \frac{K}{2} \tag{9} \]

\[ \frac{dC}{dt_p} = rK \tag{10} \]

Eq. (11) is used for fitting the regression model with the actual confirmed cases of infection. The value \( b_i = b_i(t) \) represents the actual estimate of infection where \( 1 \leq i \leq n \).

\[ C = \frac{b_1}{1 + b_2 e^{-b_3 t}} \tag{11} \]

Deterministic compartmental models

There exist numerous models for analyzing and estimating the epidemic spread. Deterministic compartmental models (DCM) are non-linear models used for modeling the epidemic spread. They mainly use differential equations for modeling the outbreak. The most commonly used DCM models are the Susceptible Infected Recovered (SIR) model, the Susceptible Exposed Infected Recovered (SEIR) model, and the Autoregressive Integrated Moving Average (ARIMA) model.

Susceptible-infected-recovered (SIR) model

The SIR model assumes total population is a sum of three different parameters namely, Susceptible \( (S) \), Infected \( (I) \), and Recovered \( (R) \) as explained in Eq. (12) \([17]\). Many researchers used the SIR model to analyze and estimate various diseases such as HIV and Ebola \([18,19]\). In the SIR model, the total population is represented by \( N \) \([10]\). A susceptible population is a subset of the total population who all are healthy, but they are at risk of becoming infected. Persons infected by the disease are known as the Infected population. Those who are

---

Fig. 16. The forecasted cumulative recovery highlighting the rise, peak, and predicted decline of COVID-19 using the SEIR model with various \( R_0 \) values having \( E_0 = 80 \) and \( \alpha = 0.2 \).
recovered from the pandemic have acquired immunity and are referred to as the recovered population. The total deceased population is also counted as recovered in the SIR model [20].

\[ N = S + I + R \]  

(12)

The SIR model works based on the assumption that the total population is constant during the period of epidemic analysis and prediction, which means no deaths and births are considered in that duration. The model estimates the changes in Susceptible (S), Infected (I), and Recovered (R) populations as differential equations explained in Eqs. (13)–(15) respectively [21,22].

\[ \frac{dS}{dt} = -\frac{\beta IS}{N} \]  

(13)

\[ \frac{dI}{dt} = \frac{\beta IS}{N} - rI \]  

(14)

\[ \frac{dR}{dt} = rI \]  

(15)

Fig. 17. Forecasted infection and recovery for various \( R_0 \) values with \( E_0 = 47 \) and \( \alpha = 0.25 \).

Fig. 18. Forecasted infection and recovery for various \( R_0 \) values with \( E_0 = 477 \) and \( \alpha = 0.0769 \).
The SIR model is used to model the pandemic considering some initial conditions. The values of the initial susceptible population, initial infected population, and the initial recovered population are denoted by $S(0)$, $I(0)$, and $R(0)$, respectively.

Infection rate represents the rate at which the susceptible population is becoming infected per day and is indicated by $\beta$. In contrast, the recovery rate, $\gamma$ indicates the recovery rate from the infection with acquired immunity [23]. The fraction of infection rate to the recovery rate as in Eq. (16) is referred to as a basic reproduction number and is denoted by $R_0$.

$$R_0 = \frac{\beta}{\gamma} \tag{16}$$

**Susceptible-exposed-infected-recovered (SEIR) model**

An advancement of the SIR estimation model is known as the SEIR model which introduces a new compartment as Exposed ($E$) to the already established compartments of SIR. The SEIR model assumes the total population as a sum of these four compartments, as shown in Eq. (17). It considers the entire population to be in the susceptible compartments.
Those exposed to infected persons, but do not become infected, are labeled as an Exposed population. The SEIR estimation model also assumes the total population is constant in the entire duration of estimation.

\[ N = S + E + I + R \]  

(17)

The Eqs. (18)-(21) explained the rate of change of \( S, E, I, \) and \( R \) with time using the SEIR model [24].

\[
\frac{dS}{dt} = \beta IS - \frac{\beta IS}{N} \tag{18}
\]

\[
\frac{dE}{dt} = \frac{\beta IS}{N} - \alpha E \tag{19}
\]

\[
\frac{dI}{dt} = \alpha E - \gamma I \tag{20}
\]

\[
\frac{dR}{dt} = \gamma I \tag{21}
\]

The rate at which exposed persons become infected is referred to as incubation rate, \( \alpha \) [2]. \( T_I \) and \( T_R \) are referred to as serial or infectious period and incubation or latent period, respectively. The value of \( \beta \) is calculated as a fraction of the reproduction number and period, \( \alpha \) and \( \gamma \) as reciprocals of incubation period and serial period respectively, as shown in Eqs. (22)–(24) [25,24]. The reciprocal infection rate is known as the contact period (1/\( \beta \)).

\[
\beta = \frac{R_0}{T_I} \tag{22}
\]

\[
\alpha = \frac{1}{T_I} \tag{23}
\]

\[
\gamma = \frac{1}{T_R} \tag{24}
\]

There are several measures available for evaluating the generated estimation model. Some of them are Residual Sum of Squares (RSS), Coefficient of Determination (\( R^2 \)), and Root Mean Squared Error (RMSE). RSS estimates the error between actual and estimated values. It is a statistical method used for identifying the variance in the actual data.
Table 5: Evaluation measures of the SEIR-based model using various $R_0$ values with $R_0 = 94$ and $\alpha = 0.167$.

| $\beta$ | $\gamma$ | $R_0$ | $R^2$ | RMSE | RSS |
|---------|---------|-------|-------|------|-----|
| (a) Cumulative Infection |
| 0.117  | 0.0225 | 5.2   | 0.961286169 | 899.3024757 | 76830769.56 |
| 0.128  | 0.0333 | 3.84  | 0.87968236 | 1584.208227 | 238422992.2 |
| 0.129  | 0.0337 | 3.83  | 0.88897347 | 1522.951596 | 220341248.5 |
| 0.13   | 0.0355 | 3.66  | 0.853164406 | 751.1411079 | 291406783 |
| 0.132  | 0.035 | 3.77  | 0.911537892 | 1359.405894 | 175588516.5 |
| 0.119  | 0.024 | 4.96  | 0.950789719 | 924.5852103 | 81211492.06 |
| 0.138  | 0.04  | 3.38  | 0.817395946 | 1953.115489 | 362392710.9 |
| 0.138  | 0.04  | 3.45  | 0.88644334 | 1538.843069 | 224963609.1 |
| (b) Cumulative Recovery |
| 0.117  | 0.0225 | 5.2   | 0.860077366 | 791.6658713 | 59539810.91 |
| 0.128  | 0.0333 | 3.84  | 0.969489801 | 369.6753204 | 12986805.04 |
| 0.129  | 0.0337 | 3.83  | 0.97630004 | 225.812541 | 10094705.56 |
| 0.13   | 0.0355 | 3.66  | 0.975863192 | 326.8045794 | 10270682.88 |
| 0.132  | 0.035 | 3.77  | 0.988598982 | 225.9796969 | 4851355.638 |
| 0.119  | 0.024 | 4.96  | 0.892962976 | 692.4125802 | 45546342.21 |
| 0.135  | 0.04  | 3.38  | 0.986304303 | 247.6794656 | 5827786.181 |
| 0.138  | 0.04  | 3.45  | 0.983266194 | 273.7795562 | 7120561.047 |

Values, which was not determined by the generated estimation model. RSS measures help to identify the optimal values of infection and recovery rates ($\beta$ and $\gamma$), which estimates the possible error rate with the selected $\beta$ and $\gamma$. Eq. (25) explains how the RSS measures are calculated.

$$\text{RSS} (\beta, \gamma) = \sum_{i=1}^{\text{len}} (y_{\text{actual}}(t) - y_{\text{predicted}}(t))^2$$

Another statistical measure used for model evaluation is the coefficient of determination ($R^2$), which is used as a goodness-of-fit measure. It is measured as a percentage of variance in the target variable estimated using the independent parameters. It calculates the strength of the relationship of the generated prediction model with the actual target variable. The value of $R^2$ is between 0 and 1 (ranging from 0 to 100%). The method for calculating the $R^2$ value is given in Eq. (26).

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Where TSS refers to the total sum of squares which is calculated as a sum of squared variation of the predicted parameter, $y_i$, $i \in \text{len}$ from its total mean, $\bar{y}$ and is explained in Eq. (27).

$$\text{TSS} = \sum_{i=1}^{\text{len}} (y_i - \bar{y})^2$$

RMSE represents the standard deviation of the actual values from estimated data points or the regression line. RMSE is calculated using Eq. (28).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{\text{len}} (E_i - A_i)^2}{N}}$$

Where $E_i$ refers to the estimated value at point $i$ and $A_i$ is the corresponding actual value.

**Autoregressive integrated moving average (ARIMA) model**

Another statistical model for analyzing and forecasting the future growth of time-dependent information is known as ARIMA. It is the most commonly used statistical-based method for estimating and analyzing changes in the time-dependent data [26]. The AutoRegression (refers to AR in ARIMA) model is used to identify the relationship between the observed data and other lagged observations. Differentiation is used to make the time series stationary; it is considered a pre-processing step (Integrated). By considering the residual error dependency and observed data, Moving Average (MA) is performed for lagged observations. Lag polynomials are used in the ARIMA model as shown in Eqs. (29) and (30) [27].

$$\varphi(L)(L-1)^d y_t = \theta(L) \varepsilon_t$$

Where $p$, $d$, and $q$ in these equations should be greater than or equal to 0. The ARIMA model with $d = 0$ (ARIMA($p,0,q$)) is ARMA($p,q$) model. When $d$ and $q$ are equal to 0 (ARIMA(p,0,0)) is AR($p$) model, and finally, when $p$ and $d$ are equal to 0 (ARIMA(0,0,0)) is MA($q$) model. In almost all cases, the value of $d$ is 1 (difference in time-series data is 1). ARIMA model with $p = 0$, $q = 0$, and $d = 1$ is a special case and is known as the Random Walk model and corresponding $\gamma_i$ is estimated using Eq. (31) [27].

$$y_t = y_{t-1} + \varepsilon_t$$

Many researchers performed analysis and estimation of COVID-19 epidemic outbreak all over the world using different estimation models. They all tried to forecast the peak values and expected ending time of this epidemic. COVID-19 spread in Italian regions are analyzed and estimated by Distante et al. [25]. The peak values infection and period are estimated using the SEIR estimation model. They studied the epidemic spread and concluded that the outbreak reaches its maximum value in Italy’s northern regions by March-end and Southern regions of Italy by the first week of April 2020. They calculated the basic reproduction number $R_0$, and predicted the future dates for Kuwait. According to their obtained results, Kuwait reaches its peak between July 23rd and August 22nd of 2020. They also found that the lockdown and other preventive measures taken by Kuwait’s government have proven to be effective in reducing the number of cases.

Syed and Sibgatullah [29] analyzed and estimated the COVID-19 outbreak in Pakistan using the SIR estimation model. They did the analysis based on the data collected from the National Database of their country. They forecasted the peak value and time at which COVID-19 reached its peak in Pakistani areas, estimating the peak on 26 May 2020. Their conclusion was that unless the authorities imposed strict policies to control the epidemic growth, 90% of the total population would be affected by the epidemic before the last week of July.

An SIR-based estimation model is generated for modeling the growth of the COVID-19 epidemic in Bangladesh by Rahman et al. [20]. They analyzed and forecasted the spread of coronavirus. They studied and analyzed the impact of various preventive measures imposed by their government for controlling the outbreak, like social distancing. They forecasted the final size of infection in their country at 3,782,558, and the epidemic would reach its peak value on the 92nd day. Their study...
concluded that social distancing has an effective impact on controlling the epidemic’s spread, and strict social distancing is one of the best measures to control the epidemic’s growth.

Batista [30] analyzed and estimated the COVID-19 epidemic spread in China, South Korea, and the rest of the world using SIR based estimation model. He did this study to estimate the final size of this outbreak in these regions. He forecasted these estimates using both the SIR and logistic model and evaluated his model using $R^2$ score. He et al. [31] proposed an SEIR-based model for analyzing and forecasting COVID-19 based on some control measures, including quarantine, hospital, etc. They modeled the epidemic considering collected information from Hubei Province. They used a particle swarm optimization algorithm for identifying the various parameters of the proposed model. Their study identified that the parameters may be changed based on the scenarios. They suggested quarantine and treatment are the best methods for controlling the epidemic. Lounis and Azevedo [32] modeled COVID-19 in Algeria using the classical and generalized SEIR model. They tried to forecast the future 100 days out based on the official confirmed cases in Algeria between April 2020 and early August 2020. They forecasted the counts of cumulative infection and deaths up to November 2020.

A model’s suitability for prediction depends on the problem at hand. Compartmental models like SIR, and SEIR are widely used to predict a pandemic’s spread because they are deterministic models and can work easily with a large population size. They can also be used to analyze the effect of various control strategies imposed by the authorities. ARIMA model, on the other hand, efficiently manages the outliers and can be used for both seasonal and non-seasonal data. The accuracy of ARIMA depends on how the observed data (training set), and/or parameters, are being modeled. For the purpose of this study, the SIR and SEIR models were chosen.

**SEIR-based prediction**

The SEIR model is an extension of the SIR model used to analyze and forecast the epidemic outbreak. The main parameters of the SEIR-based estimation model are incubation rate ($\alpha$), infection rate ($\beta(t)$), and recovery rate ($\gamma(t)$). Globally, almost 218 countries have been affected by COVID-19. This research aims to compare SIR and SEIR estimation models using the number of cases, both infection, and recovery between 24 February 2020 and 28 May 2020, and to forecast and model the COVID-19 epidemic using the best model.

Python programming language is used for simulating the SEIR-based estimation model. Python provides a vast number of predefined modules or tools for a wide variety of applications. The Python tools or modules mainly used for modeling the COVID-19 outbreak are Matplotlib, math, xlsxwriter, xrd, and sklearm [33]. Matplotlib is a Python plotting library mainly used to visualize static, animated, or interactive figures. All the graphs used here are plotted using Matplotlib. The generated estimation model is evaluated using different measures such as RSS, RMSE, and $R^2$. RMSE and $R^2$ measures are estimated using sklearm module, which provides an effective platform for machine learning. The sklearm module is constructed from SciPy, Matplotlib, and NumPy. The development environment used for developing the python-based SEIR model is Python’s Integrated Development and Learning Environment (IDLE). The data required for modeling the COVID-19 outbreak in Kuwait is collected mainly from authorized sources such as Kuwait Government’s official websites [7] related to COVID-19 and the WHO [3]. Fig. 1 depicts the collected information of confirmed daily infection and recovery in Kuwait.

The actual values of infection rates ($\beta$) and recovery rates ($\gamma$) are calculated using the collected data per day (time) and is demonstrated in Figs. 2 and 3. The recovery and infection rates, a fraction of the population already infected, represent the percentage of newly or daily recovered and infected populations. Consider an example: a recovery rate of 0.15 points out that 15% of the currently infected population at time $t - \Delta t$ is recovered at time $t$. Eqs. (32) and (33) estimate $\beta$ and $\gamma$ values for any time $t$.

$$\beta(t) = \frac{I(t) - I(t - \Delta t)}{I \Delta t}$$ \hspace{1cm} (32)

$$\gamma(t) = \frac{R(t) - R(t - \Delta t)}{I \Delta t}$$ \hspace{1cm} (33)

The latent or incubation period is between 1 and 14 days. It is difficult to find the exact value for the incubation period. So, the study is conducted with various values of the latent period and hence the incubation rate. The estimated values of $a$ are 1/4, 1/5, 1/6, 1/12, and 1/13, with various exposed population values (47, 80, 94, and 477).

Eqs. (34)–(37) estimates the value of the main four compartments ($S, E, I$, and $R$), of the SEIR estimation model at any time $t$ as a sum of these values at time $(t − 1)$ and rate of change of these values calculated using Eqs. (18)–(21) respectively [34].

$$S(t) = S(t - 1) + \frac{dS}{dt}$$ \hspace{1cm} (34)

$$E(t) = E(t - 1) + \frac{dE}{dt}$$ \hspace{1cm} (35)

$$I(t) = I(t - 1) + \frac{dI}{dt}$$ \hspace{1cm} (36)

$$R(t) = R(t - 1) + \frac{dR}{dt}$$ \hspace{1cm} (37)

The cumulative infection and recovery counts are calculated using the collected information about actual infection and recovery cases in Kuwait. These values are studied and compared with the forecasted values of both infection and recovery using $R^2$ score, RSS, and RMSE measures. The initial values of all compartments are set based on the confirmed cases from the first day of the outbreak reported in Kuwait. The total population is calculated here as 4,776,000. The entire population is considered to be susceptible. The initial values of the exposure cannot be predicted precisely. This study is conducted based on some assumed values for the exposed population; 47, 80, 94, and 477. Based on these assumed values, the susceptible population also varies accordingly, 4,775,948, 4,775,915, 4,775,901, and 4,775,518. The initial values of the infected population and the recovered population are 5 and 0, respectively.

A fraction between the rates of infection and recovery is referred to as basic reproduction number $R_0$. The transmission rate of an outbreak is determined based on the value of the $R_0$. An epidemic’s growth is determined by $R_0$: either it may or may not form an outbreak in that country or the global population. The value of $R_0$ is less than 1, then it will not become an outbreak and diminish suddenly. Otherwise, it will be emerging exponentially and severely affected a significant percentage of the total population [22,35]. The value of $R_0$ is used to identify the number of infections to be expected during the initial stages of the epidemic [35]. The infected person makes $\beta$ contacts on average and he or she is recovered within $1/\gamma$ days based on $R_0$ [36]. This research is performed based on different values of $R_0$ and $a$.

**Comparison of SIR and SEIR**

A comparison between the SIR and SEIR models was conducted for infection and recovery cases. Figs. 4–7 clearly explain how the estimated models fit with the actual values of infection and recovery. For the estimation of both infection and recovery, the SEIR model outperforms the SIR model.

The emerging COVID-19 outbreak is modeled using SIR and SEIR models for various $R_0$ values. The estimated values that were obtained using the SEIR model are close to the actual reported numbers, espe-
socially for the recovery cases.

In examining the evaluation measures for both models on infection and recovery as shown in Figs. 8–10; the SEIR model has resulted in a more accurate estimate of the values for both infection and recovery. The SEIR model with parameters $E_0 = 80$ and $\alpha = 0.2$ has surpassed the SIR model used for modeling the outbreak.

Based on the comparison performed in section “Comparison of SIR and SEIR”, the SEIR estimation model outperforms the SIR model for predicting both infection and recovery. Hence, the SEIR model is selected for forecasting the future values of infection, recovery, peak dates, and peak values. The actual values of both infection and recovery cases are collected from recognized sources like official websites of the Kuwait Government [7].

Figs. 11 and 12 illustrate the actual values of daily infection and recovery as collected from the government-authorized sources from 29 May 2020 to 1 December 2020. The daily infection count is gradually decreasing due to the government’s successful preventive measures and social distancing guidelines that were followed by the residents.

Figs. 13 and 14 depict the daily infection and recovery rates from 29 May 2020 to 1 December 2020. The infection rate, and hence the daily count, has decreased. Moreover, due to the successful control of COVID-19, the recovery rate has increased. The SEIR model has shown better results overall for predicting both infection and recovery cases. The basic reproduction number $R_0$ is decreased from the initial values, which implies that the government’s preventive measures are successful.

Results and discussion

The SEIR-based estimation is performed based on the confirmed cases between 24 February 2020 and 1 December 2020. The first five COVID-19 cases in Kuwait were reported on 24 February 2020. The values of infection and recovery rates are changed over time. In the early stages, the infection rates and the number of infections increased slowly and then decreased. The recovery rate and the number of recovered cases increased slowly. The main problem that was raised when using the SEIR model for modeling COVID-19 in Kuwait, is the lack of methods for measuring the initial exposed population’s exact value. So, this research assumes the initial exposed population and incubation rate.

This research is performed with various values for $R_0$, $E_0$, and $\alpha$. The various values for $E_0$ are 47, 80, 94, and 477 and for $\alpha$ 1/4, 1/5, 1/6, 1/12, and 1/13. From these values, $E_0 = 80$ and $\alpha = 0.2$ gave better performance than other values. The forecasted cumulative infection and recovery using the SEIR model with $E_0 = 80$ and $\alpha = 0.2$ based on different $R_0$ values is illustrated in Figs. 15 and 16.

The preventive measures such as quarantine, curfew, travel, and entry restrictions issued by the Kuwait Government to control the growth of COVID-19 almost came into fruition. The number of confirmed cases decreased gradually, and recovered cases increased. Social distancing and other precautionary measures reduce the spread of the virus.

The containment measures issued by the Kuwait Government had a positive impact on daily infection count. The precautionary measures, such as a full or partial curfew, travel ban, social distancing, wearing mask and gloves, using hand sanitizer, etc., had positive feedback in decreasing infection count. The daily infection count is gradually decreasing. The preventive measures controlled the growth of COVID-19 in Kuwait to some extent.

The study is performed on various values $E_0$ and $\alpha$ using different $R_0$ values. Based on the evaluation performed on various values of initial exposed population and incubation rate, $E_0 = 80$ and $\alpha = 0.2$ gives better results. Figs. 17–20 illustrate the forecasted growth of infection and recovery for various values of $E_0$ and $\alpha$ using various $R_0$ values.

Conclusion

Analyzing and forecasting the spread of an outbreak, while it is happening, is essential to help authorities determine necessary precautions and containment measures for controlling the transmission of the disease. Several mathematical and compartmental models have been commonly used in epidemiology research to model an epidemic’s spread including HIV, Ebola, and COVID-19. Factors such as population size and purpose of prediction can affect a model’s suitability. The main focus of this study was to analyze and compare both SIR and SEIR models and find the more suitable one among them for forecasting future values, while taking into consideration the impact of preventive measures implemented by the Kuwait Government. In this research, we evaluated and compared both models’ performance and selected the SEIR model as the more suitable model based on the data collected for the period of 24 February 2020 to 1 December 2020 from Kuwait government-authorized sources. The Python programming language was used for simulating the SEIR model on various values of basic reproduction number $R_0$, initial exposed population $E_0$, and incubation rate $\alpha$. The results showed that the SEIR model is fitted with the infection and recovery cases for values of $R_0$ ranging from 3 to 4, an incubation rate of 0.2, and an initial exposed population of 80. In our evaluation of the estimation models, we used accuracy measures like $R^2$, RMSE, and RSS. It should be noted that the data collected for analyzing and estimating the spread does not consider any external factors that might influence the number of infection and recovery cases. The results of our study have shown that containment measures like travel restrictions and lockdowns were proven to control the spread of COVID-19. Other precautionary measures such as social distancing and wearing masks have helped in curbing the spread of COVID-19.

Declaration of Competing Interest

We, the authors, declare that we have no conflict of interest in this research study.

Appendix

Tables 1–5 show an evaluation measure of the SEIR-based estimation model using various $E_0$ and $\alpha$ values. Based on presented evaluation, the SEIR model with $E_0 = 80$ and $\alpha = 0.2$ gives a better prediction and the estimated values are best fitted with the actual values.

References

[1] Roda W, Varughese M, Han D, Li M. Why is it difficult to accurately predict the covid-19 epidemic? Infect Disease Model 2020;5:03.
[2] Hamzah FR, Lau C, Nazri H, Ligot D, Lee G, Tan C, et al. Coronatracker: worldwide covid-19 outbreak data analysis and prediction. Bull World Health Organ 2020;1:32.
[3] WHO. Coronavirus disease (covid-19) situation report. Available online: http://www.who.int/emergencies/diseases/novel-coronavirus-2019, 05 2020, Accessed on: 28-05-2020.
[4] Unhale S, Bilal Q, Sanap S, Thakrre S, Wadatkar S, Bairagi R, et al. A review on corona virus (covid-19). Int J Pharmaceut Life Sci 2020;6:109–15.
[5] Zhang X, Ma R, Wang L. Predicting turning point, duration and attack rate of covid-19 outbreaks in major western countries. Chaos Solitons Fract 2020;109825.
[6] Arsath Abbasi A, Sathesh S. On predicting the novel covid-19 human infections by using infectious disease modelling method in the indian state of tamil nadu during 2020. medRxiv, 2020.
[7] CAIT. Corona virus covid-19 updates. Available online: https://corona.e.gov.kw/En/, 05 2020, Accessed on: 28-05-2020.
[8] PACI official website. URL: http://stat.paci.gov.kw/englishreports/, accessed: 2020-06-01.
[9] Kelendar H, Ma M. Hospital bed occupancy and utilization: Is kuwait on the right track. J Hospital Med Manage 2019;5(2):3.
[10] Krauer B. Compart Models Epidemiol, vol. 1945, ch. 2. Berlin, Heidelberg: Springer; 04 2008., p. 19-75.
[11] Shi P, Cao S, Feng P. Seir transmission dynamics model of 2019 ncov coronavirus with considering the weak infectious ability and changes in latency duration, medRxiv, 2020.
[12] Smith N, Trauer J, Gambhir M, Richards J, Maude R, Keith J, Flegg J. Agent-based models of malaria transmission: A systematic review. Malaria J 2018;17:12.
[13] McCullagh, P, Nelder JA. Generalized Linear Models. CRC Press, 1989, vol. 2, ch. 2, p. 31–52.
[14] Batista M. Estimation of the final size of the coronavirus epidemic by the logistic model (update 3), medRxiv, 2020, Accessed on: 28-05-2020.
[15] Weiss HH. The sir model and the foundations of public health. Mater Matemat 2013:0001–17.
[16] Bhandari P. Analysis of prediction models in spread of ebola virus disease, Ph.D. dissertation, DEAKIN University Australia, 05 2019.
[17] Little TD. The Oxford Handbook of Quantitative Methods, vol. 2, ch. 3. Oxford University Press; 2013. p. 26–51.
[18] Adhikari R, Agrawal R. An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613, 01 2013.
[19] Perasso A. An introduction to the basic reproduction number in mathematical epidemiology. ESAIM: Proc Surv. 2018;62:123–138.
[20] Mandal M, Mandal S. Covid-19 pandemic scenario in india compared to china and the united states. Biomech Model Mechanobiol 2020;04.
[21] Huang Y, Yang L, Dai H, Tian F, Chen K. Epidemic situation and forecasting of covid-19 in and outside china. Bull World Health Organ 2020;10.
[22] Lutz M. Learning Python: Powerful Object-Oriented Programming. 4th ed. Media: O’Reilly; 2009.
[23] M. Alenezi et al. Mohammed N. Alenezi is an assistant professor at the Public Authority for Applied Education and Training, College of Business Studies, Computer Science & Information Systems (CSIS) Department. He received his B. E. in Computer Engineering from the University of Kuwait in 2001. He obtained his master’s degree in information networks from the University of Essex, UK, in 2008 with Distinction. He received his Ph.D. in Network Security from the University of Essex, UK, in 2014. His main research interests are network security, network protocols, data science & analytics, web development, ethical hackings and has published in these areas.
[24] Alabdulrazzaq as an Assistant Professor in the Computer Science and Information Systems Department at the College of Business Studies a part of the Public Authority for Applied Education and Training. Alabdulrazzaq received a PhD. in Computer Science and Software Engineering from Auburn University, a Master of Science in Computer Science and Software Engineering from the George Washington University, and a Bachelor’s of Science in Computer Science from Kuwait University. His research interests include secure software development, software process, software quality assurance and agile software development.
[25] Alhusaini received his bachelor degree in Computer Engineering from Kuwait University, a Master and Ph.D. degrees in Computer Engineering from the University of Southern California (USC). He is an assistant professor in the Computer Engineering Department at Kuwait University, and currently serving as the deputy director general for the Central Agency for Information Technology. He also worked as the director of Kuwait Cultural Office at the Kuwait Embassy in Australia. He received several awards over his career, including the Distinguished Teaching Award and the Best Electronic Course Award from Kuwait University, and the Outstanding Academic Achievement Award from the University of Southern California, USA . Dr. Alhusaini served as a consultant and participated in several projects and committees in the field of information technology for public and private sectors.
[26] Abdullah F. Al-Anzi is a Computer Engineer that graduated from the American University of Kuwait and pursuing his masters degree at Kuwait University. He is an active member in the IEEE student branch, a member of the Kuwaiti Delegation in the Public Youth Authority, and a supervisor and instructor in Maker’s Academy in Kuwait. He is currently a full stack application developer in The Shared Electronic Banking Company, KNET. His research interests includes data analytics, cyber security and system modeling.