Kuwait J.Sci., Special Issue. 2021,pp(1-30)

Data Analysis and Forecasting of COVID-19 Pandemic in Kuwait Based on Daily Observation and Basic Reproduction Number Dynamics

Kayode Oshinubi[1] , Fahimah Al-Awadhi[2] ,
Mustapha Rachdi[1] , Jacques Demongeot[1,*]
[1] *Laboratory AGEIS EA 7407, Team Tools for e-Gnosis Medical and Labcom CNRS/UGA/OrangeLabs Telecom4Health, Faculty of Medicine, University Grenoble Alpes (UGA), 38700 La Tronche, France; Kayode.Oshinubi@univ-grenoble-alpes.fr, Mustapha.Rachdi@univ-grenoble-alpes.fr, Jacques.Demongeot@univ-grenoble-alpes.fr*
[2] *Dept. of Statistics and Operations research, Faculty of Science, Kuwait University, P.O Box 5969, Safat 13060, Kuwait; mapfahaa@hotmail.com*
**Corresponding author: Jacques.Demongeot@univ-grenoble-alpes.fr*

Abstract

Coronavirus (COVID-19) has continued to be a global threat to public health. When the coronavirus pandemic began early in 2020, experts wondered if there would be waves of cases, a pattern seen in other virus pandemics. The overall pattern so far has been one of increasing cases of COVID-19 followed by a decline, and we observed a second wave of increased cases and yet we are still exploring this pandemic. Hence, updating the prediction model for the new cases of COVID-19 for different waves is essential to monitor the spreading of the virus and control the disease. Time series models have extensively been considered as the convenient methods to predict the prevalence or spreading rate of the disease. This study, therefore, aimed to apply the Autoregressive Integrated Moving Average (ARIMA) modelling approach for predicting new cases of coronavirus (COVID-19). We propose a deterministic method to predict the basic reproduction number R_o of first and second wave transition of COVID-19 cases in Kuwait and also to forecast the daily new cases and deaths of the pandemic in the country. Forecasting has been done using ARIMA model, Exponential smoothing model, Holt's method, Prophet forecasting model and machine learning models like log-linear, polynomial and support vector regressions. The results presented aligned with other methods used to predict R_o in first and second waves and the forecasting clearly shows the trend of the pandemic in Kuwait. The deterministic prediction of R_o is a good forecasting tool available during the exponential phase of the contagion, which shows an increasing trend during the beginning of the first and second waves of the pandemic in Kuwait. The results show that support vector regression has achieved the best performance for prediction while a simple exponential model without trend gives good optimal results for forecasting of Kuwait COVID-19 data.

Keywords: ARIMA; COVID-19; forecasting; machine learning; pandemic

1. Introduction

The world suffered a lot of pandemics and diseases throughout the history of mankind. One of these pandemics was the Coronavirus (COVID-19), which arose in the last year 2020 and has been classified as a global public health emergency. This pandemic is thought to be among the most horrific outbreaks of disease throughout all human history. In January, 2020 WHO declared the coronavirus is responsible for respiratory illness in Wuhan China and the first case in Kuwait was reported on February, 24th 2020. Since then the daily new cases increased exponentially until May, 2020 when the first wave started to decline. The same exponential dynamics has been observed between January and March, 2021. The forecast of new cases and deaths recorded daily is crucial so that health experts and citizens can be guided in order to avoid an escalation of the pandemic. Since the emergence of COVID-19 pandemic in Kuwait, the country has experienced a high stationary number of new cases until a slight decline in December, 2020. After the first wave, the new cases increased exponentially from January, 2021 until March, 2020 followed by a stationary phase analogue to that of 2020 but with a higher average number of daily new cases.

The COVID-19 pandemic has resulted in a lot of fatality across the globe, but the tenth of April 2021, the statistics by worldometer® (Worldometer 2021) indicated that Kuwait cumulated number of cases is 244,325, with a number of deaths equal to 1,393 (323 per million), corresponding to a fatality rate of 0.58% , placing Kuwait in 83rd place (in the ranking of decreasing fatality rate) of developed countries. Recovery rate was 93.57%, critical cases treated in ICU were $224 - 0.09\%$ of total cases, daily cases receiving treatment was $14,305 - 5.85\%$ of total cases, and cumulated confirmed cases proportion equals 56,568 per million.

In recent time, there has been a lot of research on the COVID-19 pandemic in different fields ranging from statistics, epidemiology, mathematics, biology, medicine, etc. and these fields have looked at various aspects of COVID-19 pandemic modelling in the areas of reported and unreported cases, prediction of basic reproduction number R_o, lockdown and more recently introduction of vaccines. For example, in (Demongeot *et al.* 2020) authors worked on the spread parameters of the new COVID-19 cases dynamic and concluded on how temperature indicates the cases in 21 countries. They proposed the ARIMA model to analyse incidence patterns and estimate short-term forecasts for retro-predicting the first wave of COVID-19 outbreak. (Seligmann *et al.* 2020) worked on inverted covariate effects for first versus mutated second wave COVID-19 and how confinements hasten viral evolution toward greater contagiousness, (Demongeot *et al.* 2021) proposed a new method for calculating the daily reproduction number during the contagiousness period of an individual.

Many researchers so far have applied data-driven statistical models like Autoregressive Integrated Moving Average (ARIMA) for prediction of future trends of the infectious disease as one wave using data of different countries, for example, (Abenvenuto *et al.* 2020), (Anastassopoulou *et al.* 2020) modelled China Covid 19, (Grasselli *et al.* 2020), (Russo *et al.* 2020) proposed a time series model for Italy, (Massonnaud *et al.* 2020) used France Covid 19 data, (Wise *et al.* 2020) proposed a statistical model for USA covid data, (Fanelli & Piazza 2020) used China, Italy and France data and (Gupta & K Pal 2020) India data. The technique of time series analysis has been widely applied, for its reliability and quick implementation by various stakeholders. Machine learning model is another effective technique that can be applied using different models such as log-linear and support vector regression. Both the waves of Covid 19 will be considered separately for better accuracy. Sensitivity analysis for advanced four compartment mathematical model was explained by (Munir *et al.* 2020).

In this paper, we propose a deterministic method to predict the basic reproduction number R_o of first and second wave transition of COVID-19 cases in Kuwait for the daily new cases and the number of the deaths of the pandemic in the country. We will apply several statistics tools for modelling epidemic data such as ARIMA model, Exponential smoothing model, Holt's method, Prophet forecasting model and machine learning models. Among machine learning models we will use log-linear, polynomial and support vector regressions to compare. The main objective and motivation of this paper is to present the best statistical model to describe the daily count of new cases and deaths due to COVID-19 infections. We, therefore, used the best model that we think will predict the number of new cases and a model for mortality due to the COVID-19 infections. The novelty of this work is to decompose the Kuwait Covid-19 data into two main waves and model each wave separately using different tools along with the classical time series model that depends on the basic reproduction number R_o. A deterministic method is applied to estimateR_o, followed by many different additive models to predict the time series such as Exponential regression, log-linear and Polynomial with different degrees. In addition, we applied several smoothing methods, Prophet and Support vector methods. A comparison is made between these different techniques for the best fit. These different tools have been applied to Kuwait data and our models is supported by validity and accuracy tests. We also presented the sensitivity analysis for parameters used for the ARIMA model and exponential models. The remainder of this article is structured in the following manner. In Section 2 we describe the methods used in this paper for processing data and present the deterministic modelling of daily new cases observed in order to predict R_o. In Section 3, we give results and visualisation of the machine learning tools applied on the pattern of the daily new cases and deaths in Kuwait and demonstrate the efficiency of the models in comparison to others. Finally, Section 4 contains conclusions conducted from the paper and major findings illustrated from our work.

2. Materials and Methods

Looking at the pattern of COVID-19 data of Kuwait, one realises that it consists of two main waves. Each wave lasts approximately three months. We took 100 days both in the first wave and second wave depending on the available data. First wave new cases were considered from 25/02/2020 to 03/06/2020 while second wave new cases were considered from 15/10/2020 to 22/01/2021. For the first wave, we used daily deaths data from 04/04/2020 to 12/07/2020, while for the second wave we used those from 15/10/2020 to 22/01/2021. We calculated the slopes from the log-linear regression analysis using exponential model $y = ae^{bx}$, where y is the daily number of new cases, x the number of days, b the slope and $loga$ a constant in the log format $logy = loga + bx$. We also calculated the initial negative autocorrelation slope of the epidemic spread averaged on six days. A deterministic model was used to predict R_o of both the first and second wave of the pandemic in Kuwait. Simulation, data visualisation and computation were done in R and Python environment.

2.1 Methods

2.1.1 Time series modelling

Time series modelling has been introduced by N. Wiener for prediction and forecasting (Wiener 1949) and also by (Granger & Newbold 1986), (Majid *et al.* 2020) and (Box & Jenkins 1976) . Its parametric approach assumes that after subtracting the trend (increasing or decreasing) of Covid-19 new cases dynamics, we get an underlying stationary stochastic process $N(j)$, at day j, j = 1,2, ... which can be described by a small number of parameters using the autoregressive ARIMA model:

$$N(j) = \sum_{i=1,\dots,r} a(i)N(j-i) + W(j), \tag{1}$$

where W is a random residue, whose variance is to minimise. The autocorrelation curve is obtained by calculating the correlation $A(k)$ between $N(j)$ and the $N(j-k)$'s (j belonging to a moving time window) by using the formula:

$$A(k) = \frac{E[N(j)N(j-k)] - E(N(j))E(N(j-k))}{\sigma(N(j)\sigma(N(j-k))}. \tag{2}$$

where E denotes the expectation and σ is the standard deviation.The autocorrelation function A allows examining the serial dependence of the $N(j)$'s. We used the classical ARIMA (p, d, q) model, where p means the order of auto-regression, d the degree of trend difference and q the order of moving average.

2.1.2 Linear, Polynomial and Support Vector Regression

Log-linear regression model use some historic data between variables X and Y depending on X, and consider a linear relationship between $logY$ and X, while polynomial regression models use a similar approach but the dependent variable Y is modelled as a polynomial of degree n in X. Support vector regression is a supervised machine learning model which draws a hyperplane relating the data points and creates a boundary of possible data points (high and low) in future. Support vector regression traditionally has huge forecasting abilities.

2.1.3 The Prophet Forecasting Model, Exponential smoothing model and Holt's model

The Prophet forecasting modelling uses a decomposable time series model with three main model components which are trend, seasonality and holidays. It is described by the model equation below:

$$y(j) = g(t) + s(j) + h(j) + \epsilon_j. \tag{3}$$

where $g(j)$ is a piecewise linear or logistic growth curve for modelling represents non-periodic changes in time series, $s(j)$ represent periodic changes like weekly, yearly or seasonal, $h(j)$ is the effect of holidays with irregular schedules and ϵ_t. is the error term which accounts for any unusual changes or noise not accommodated by the model (Rafferty 2021). This tool was also used on another set of data by (Abioye *et al.* 2021). Exponential smoothing methods use weighted averages of past observations to forecast new values. It combines error, trend and seasonal components in a smoothing calculation. Each term can be combined either additively, multiplicatively or be left out of the model. Holt's-winter method is also called triple exponential smoothing and it is used in order to apply exponential smoothing to the seasonal components in addition to level and trend. Holt's linear trend method takes into account the trend of a given data set. The method map the trend accurately without any assumptions (Chatfield 2000).

2.2 Deterministic modelling

According to (Demongeot *et al.* 2021), let R_o denotes the basic reproduction number among the Kuwait population. We can estimate the distribution V (whose coefficients are denoted $V_j = R_j/R_o$) of the daily reproduction numbers R_j along the contagious period of an individual, by remarking that the number X_j of new infectious cases at day j, equal to $X_j = I(j)–I(j-1)$, where $I(j)$ is the cumulated number of infectious at day j, verifies the discrete convolution equation:

$$X_j = \sum_{k=1,r} R_k X_{j-k},$$

giving in continuous time:

$$X(t) = \int_1^r R(s)X(t-s)ds, \tag{4}$$

where r is the duration of the contagion period, estimated by $1/(k+\mu)$,where k is the recovering rate and μ is the death rate in SIR equations:

$$dS/dt = -\nu SI$$

$$dI/dt = \nu SI - (k+\mu)I, \tag{5}$$

where S and I are respectively the size of the susceptible and infectious populations.

If r and S can be considered as constant during the first exponential phase of the pandemic, we can also assume that the distribution V is constant and then, V can be estimated by solving the linear system:

$$X_j = \sum_{k=1,r} R_k X_{j-k}, X_{j-1} = \sum_{k=1,r} R_k X_{j-1-k}, \cdots,$$

$$X_{j-r+1} = \sum_{k=1,r} R_k X_{j-r-k},$$

which can be written as $X = MR$, hence giving

$R = M^{-1}X$, where:

$$M = \begin{bmatrix} X_{j-1}, & X_{j-2}, & \cdots, & X_{j-r} \\ X_{j-k}, & X_{j-k-1}, & \cdots, & X_{j-k-r+1} \\ & & \cdots & \\ X_{j-r}, & X_{j-r-1}, & \cdots, & X_{j-2r+1} \end{bmatrix} \tag{6}$$

Equation (6) can be solved numerically, if the pandemic is observed during a time greater than $1/(k+\mu)$. Then, the entropy of $V = R/R_o$ can be calculated, as the Kolmogorov-Sinaï entropy of the Markovian Delbrück scheme ruling the $X_j's$ and giving new parameters for characterising pandemic dynamics, namely for quantifying its robustness and stability ((Rhodes & Demetrius 2010), (Demongeot & Demetrius 2015)).

If there are negative V_j's, it is still possible to define an index of proximity to uniformity of V by considering the entropy of the distribution W defined by $W_j = [(V_j - min\{V_k \leq 0\})/\sum_{i=1,r} W_i]$.

3. Results

3.1 Application of Deterministic modelling

3.1.1 Start of the pandemic in Kuwait

Suppose we use the daily new infectious cases given in worldometer from the beginning of the pandemic, then we can calculate M^{-1} for the period from February 25th to March 1st 2020, by choosing 3 days for the duration of the infectiousness period and the following raw data for the new infected cases are $X_1 = 6$ the 25th February, $X_2 = 15$ the 26th, $X_3 = 17$ the 27th, $X_4 = 2$ the 28th, $X_5 = 0$ the 29th and $X_6 = 1$ the 1st March, hence giving for the matrix M:

$$M = \begin{bmatrix} X_5 & X_4 & X_3 \\ X_4 & X_3 & X_2 \\ X_3 & X_2 & X_1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 17 \\ 2 & 17 & 15 \\ 17 & 15 & 6 \end{bmatrix}$$

Then, we have:

$M^{-1} =$

$$\begin{bmatrix} 0.03140158 & -0.06203727 & 0.06612203 \\ -0.06203727 & 0.07378095 & -0.00868011 \\ 0.06612203 & -0.00868011 & 0.00102119 \end{bmatrix}$$

Because

$$X = \begin{bmatrix} X_6 \\ X_5 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix},$$

we get $R = M^{-1}X$, where:

$$R_1 = 0.16364565$$

$$R_2 = -0.0793975$$

$$R_3 = 0.06816441$$

The graphical representation of the R_j's is given in Figure 1 and the average transmission rate $R_o \approx 0.2$, value close to that calculated directly, with a maximal daily reproduction rate the first day of the infectiousness period. Because of the negativity of R_2, we have to calculate the entropy H of the distribution $W = [(R_1 - R_2)/(R_1 - 2R_2 + R_3), 0, (R_3 - R_2)/(R_1 - 2R_2 + R_3)] = (0.622, 0, 0.378)$, which equals 0.663.

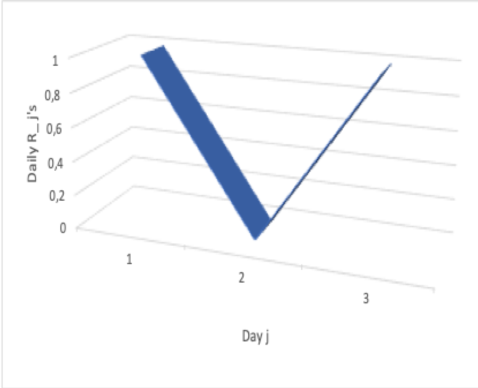


Fig. 1. V-shape of the evolution of the daily R_j's along the infectious 3-day period of an individual, calculated for the period from February 25 to March 1 2020.

3.1.2 First exponential phase of the pandemic in Kuwait

If we use the daily new infectious cases given in (Worldometer 2021) during the exponential phase of the first wave, we can calculate M^{-1} for the period from October 20 to October 25 2020, by choosing 3 days for the duration of the infectiousness period and the following raw data for the new infected cases are $X_1 = 886$ the 20th October, $X_2 = 813$ the 21th, $X_3 = 889$ the 22th, $X_4 = 812$ the 23th, $X_5 = 695$ the 24th and $X_6 = 708$ the 25th, hence giving for the matrix M :

$$M = \begin{bmatrix} 695 & 812 & 889 \\ 812 & 889 & 813 \\ 889 & 813 & 886 \end{bmatrix}$$

Then, we have:

$M^{-1} =$

$$\begin{bmatrix} 0.00507334 & -0.00013316 & 0.00521271 \\ -0.00013316 & 0.00699023 & -0.00628068 \\ 0.00521271 & -0.00628068 & 0.00166151 \end{bmatrix}$$

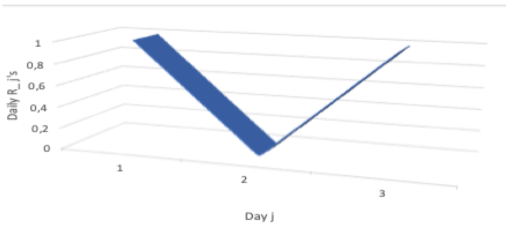
Because

$$X = \begin{bmatrix} X_6 \\ X_5 \\ X_4 \end{bmatrix} = \begin{bmatrix} 708 \\ 695 \\ 812 \end{bmatrix},$$

we get $R = M^{-1}X$, where:

$$R_1 = 0.54824733$$
$$R_2 = -0.33597567$$
$$R_3 = 0.67466856$$

The graphical representation of the R_j's for the second wave is given in Figure 2. The average transmission rate R_o verifies $R_o \approx 0.9$, value close to that calculated directly, with a maximal daily reproduction rate the first day of the infectiousness period. Because of the negativity of R_2, we have to calculate the entropy H of the distribution $W = [(R_1 - R_2)/(R_1 - 2R_2 + R_3), 0, (R_3 - R_2)/(R_1 - 2R_2 + R_3)] = (0.467, 0, 0.533)$, which is equal to $H = 0.691$.



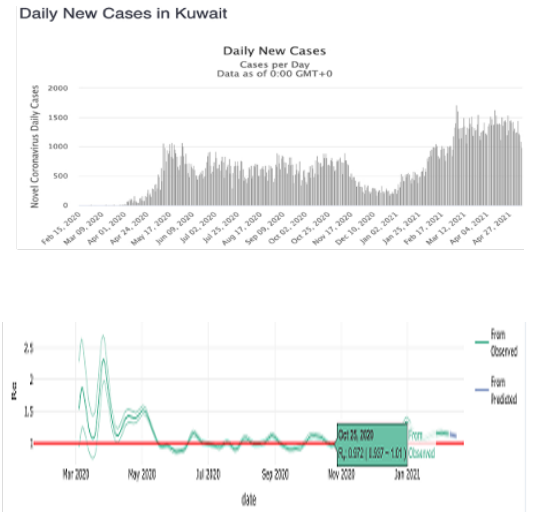


Fig. 2. V-shape of the evolution of the daily R_j's along the infectious 3-day period of an individual, daily new cases in Kuwait between October 20th to October 25th 2020 in worldometer and estimation of the average transmission rate R_o for the 20th and 25th October 2020 with its 95% confidence interval (in green).

3.1.3 Second exponential phase of the pandemic in Kuwait

If we use the daily new infectious cases given in (Worldometer 2021) during the exponential phase of the second wave, we can calculate M^{-1} for the period from December 30 2020 to January 4 2021, by choosing 3 days for the duration of the infectiousness period and the following raw data for the new infected cases are $X_1 = 205$ the 30th December, $X_2 = 286$ the 31th, $X_3 = 285$ the 1st January, $X_4 = 205$ the 2nd, $X_5 = 269$ the 3rd and $X_6 = 372$ the 4th, hence giving for the matrix M :

$$M = \begin{bmatrix} 269 & 205 & 285 \\ 205 & 285 & 286 \\ 285 & 286 & 205 \end{bmatrix}$$

Then, we have:

$M^{-1} =$

$$\begin{bmatrix} 0.00504561 & -0.00852449 & 0.00487808 \\ -0.00852449 & 0.00563046 & 0.00399594 \\ 0.00487808 & 0.00399594 & -0.00747849 \end{bmatrix}$$

Because

$$X = \begin{bmatrix} X_6 \\ X_5 \\ X_4 \end{bmatrix} = \begin{bmatrix} 372 \\ 269 \\ 205 \end{bmatrix},$$

we get $R = M^{-1}X$, where:

$$R_1 = 0.58388424$$

$$R_2 = -0.83734838$$

$$R_3 = 1.35646161$$

Then, we can give a graphical representation of the R_j's (Figure 3). The average transmission rate R_o equals about 1.1, value close to that calculated directly, with a maximal daily reproduction rate the first day of the infectiousness period. Because of the negativity of R_2, we have to calculate the entropy H of the distribution $W = [(R_1 - R_2)/(R_1 - 2R_2 + R_3), 0, (R_3 - R_2)/(R_1 - 2R_2 + R_3)] = (0.393, 0, 0.607)$, which is equal to $H = 0.691$, the same value than for the first wave.

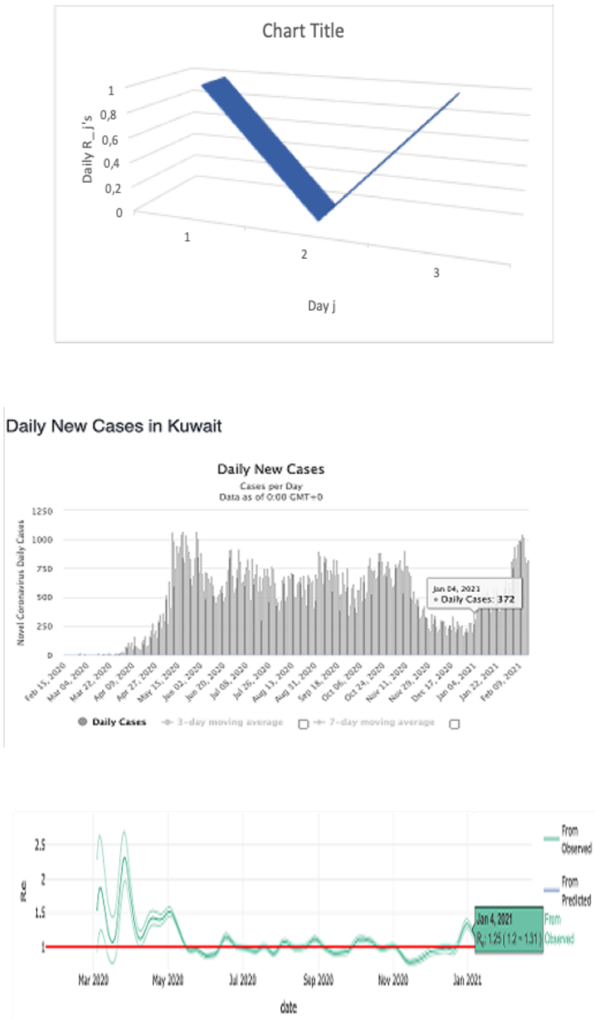


Fig. 3. V-shape of the evolution of the daily R_j's along the infectious 3-day period of an individual, daily new cases in Kuwait between December 30 2020 and January 4 2021 in worldometer and estimation of the average transmission rate R_o for December 30th 2020 and January 4th 2021 with its 95% confidence interval (in green).

3.2 Regression curves and slopes for first wave start and second wave transition

In this Section we applied the regression model on the daily new cases for both waves. As expected, the 100 first days of the first wave start show a positive slope (0.0687) and the 100 first days of the second wave transition show a negative slope (-0.0094).

For first wave start, we obtain the exponential model: $y = 0.8e^{0.07x}$, based on the following results: Log-linear regression: slope = 0.0686852051631, intercept=0.800565672854, r = 0.939152595486, p-value $= 2.079801359e^{-45}$, standard error = 0.002591135543915, R-squared = 0.882008, Root MSE = 0.711400960815.

Using polynomial regression of order 4 we get: R-squared = 0.787421, Root MSE = 0.586403581239, p-value$< 2.2e^{-16}$, standard error = 0.6069, F-statistics = 256.7. Using Support vector regression we get: R-squared = 0.98382, Root MSE = 0.263424533991, slope = 0.0128466, intercept= 3.697591, p-value$< 4.33e^{-15}$.

For second wave transition, the exponential model is $y = 6.5e^{-0.009x}$, based on the following results: Log-linear Regression: slope = -0.0093674219195, intercept = 6.5198283885, r = -0.57106504118, p-value = 5.50386624132e-10, standard error = 0.00136023603, R-squared = 0.326115, Root MSE = 0.388700386316, Polynomial of order four Regression: R-squared = 0.0927367, Root MSE = 0.1818719990, p-value$< 2.2e^{-16}$, standard error = 0.186, F-statistics = 137.2, Support Vector Regression: R-squared = 0.9680774158, Root MSE = 0.0846001608, slope = -0.2570435, intercept= -1.370541, p-value = 0.2775. Figures 4 show different regressions for the first and transition to second waves of Covid-19 in Kuwait. All coefficients for the first wave and transition to second wave for both log-linear and polynomial regression are significant with p-value less than 0.001. The residuals of the log-linear regression were examined and it was discovered that the median for both the cases were close to zero with median = 0.06031 for the first wave and median = 0.07629 for transition to the second wave. The normality of the residual was tested using Jarque-Bera test and with high p-value we fail to reject the null hypothesis that the skewness and kurtosis of the residuals are significantly equal to zero. Also, the median of the residual for both first wave and transition to second wave of polynomial regression of order four is close to zero with median = 0.0271 and median = 0.00669 respectively. For support vector regression, we tested the normality of the residual using the Jarque-Bera test and it was discovered that it's skewness and kurtosis are significantly equal to zero since the p-value is large. The mathematical formulation of the support vector regression where b and w are the coefficients is given as:

$$y = f(x, w) =< w, x >= \sum_{j=1}^{m} w_j x_j + b, \tag{7}$$

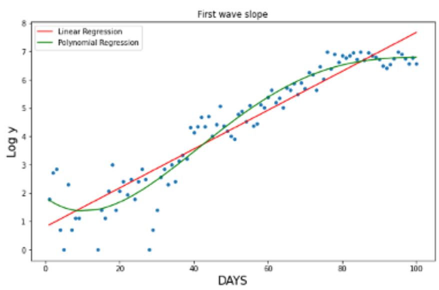
with $w_j > 0$ support points minimize prediction error where $\|w\|$ is the magnitude of the normal vector

$$min_w \frac{1}{2}\|w\|^2, \tag{8}$$

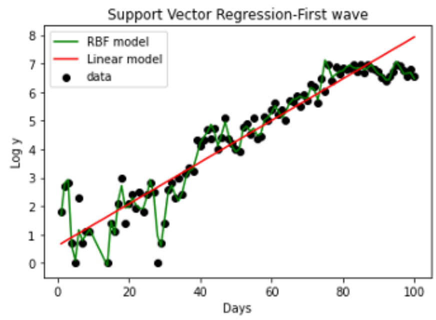
with constraints

$$\mid y_j - w_j x_j \mid \leq \epsilon. \tag{9}$$

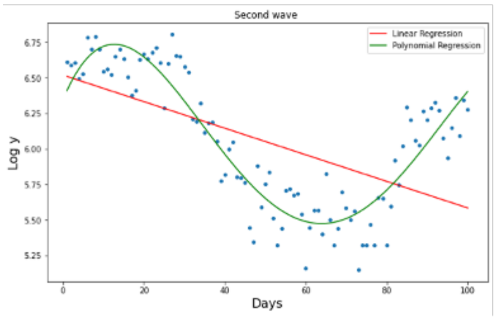
The error term is instead handled in the constraints, where we set the absolute error less than or equal to a specified margin, called the maximum error (ϵ). The general parameters for the support vector we used are cost = 1, $\gamma = 1$ and $\epsilon = 0.1$.



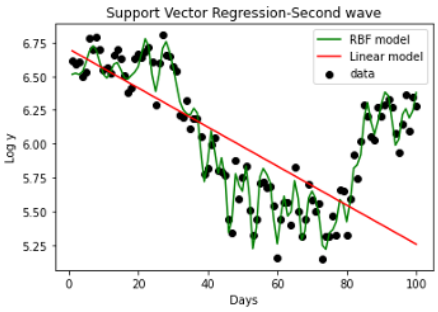
(a)



(b)



(c)

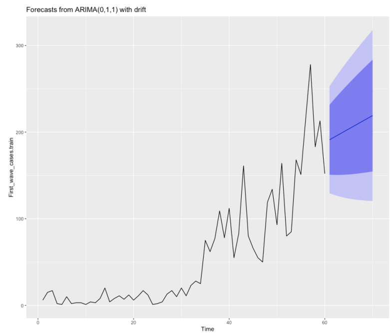


(d)

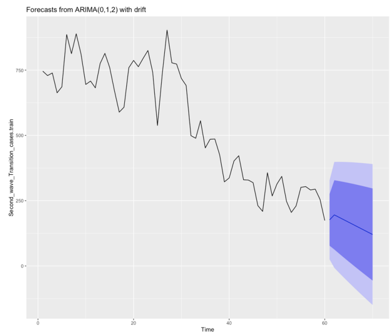
Fig. 4. (a) Log-linear and polynomial regression for first wave cases. (b) Support vector regression for first wave cases. (c) Log-linear and polynomial regression for second wave transition cases. (d) Support vector regression for second wave transition cases. The origin of time corresponds to the corresponding wave start.

3.3 ARIMA result for first wave start and second wave transition

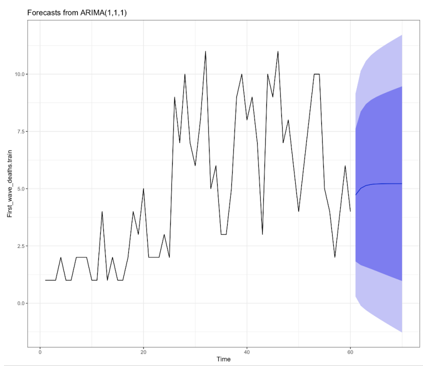
In this Section we applied the time series model on both the daily new cases and deaths for the first and second waves. We used the 'auto' function in R software for ARIMA model to fit the best model for the data in consideration with degree of freedom 8 and model degree of freedom 2 with 10 lags. We compared the results given by this function with other parameters for (p,d,q) as defined in Section 2.1 and we discovered that the parameters values given by this function were efficient and gave a better result as others gave higher Root MSE and p-value. We trained 60% of the data and forecast 100 days while the rest was tested to show the validation and accuracy of the model. We observed that while the Root MSE of the tested data was high, by checking the residual using Ljung-Box test, the Q^* is large with p-value < 0.05 which shows that the autocorrelation did not come from white noise. We present the results as follows and the visualisation in Figure 5: for first wave new cases the ARIMA parameters used are (0,1,1) with drift, Root MSE for training = 30.7, Root MSE for test = 533.5, $Q^* = 33$ and p-value= 0.00007. For second wave transition new cases the ARIMA parameters used are (0,1,2) with drift, Root MSE for training = 74.5, Root MSE for test = 402.3, $Q^* = 15$ and p-value= 0.04. Using the Death data we apply the ARIMA model and got: for first wave deaths the ARIMA parameters used are (1,1,1), Root MSE for training = 2.2, Root MSE for test = 2.37, $Q^* = 16$ and p-value= 0.04. For the second wave transition deaths the ARIMA parameters used are (0,1,0) and this is the only case where the p-value was high despite low Root MSE: the model degree of freedom is zero, Root MSE for training = 2.21, Root MSE for test = 0.894, $Q^* = 13$ and p-value= 0.2. We conclude that the ARIMA model best fits the first wave deaths with the best performance.



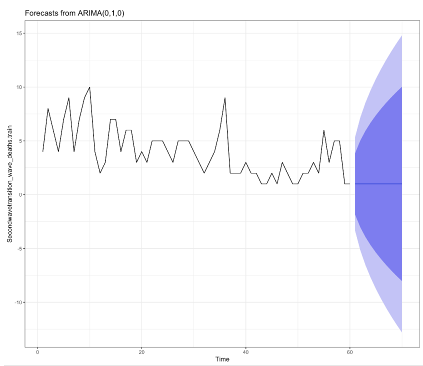
(a)



(b)



(c)

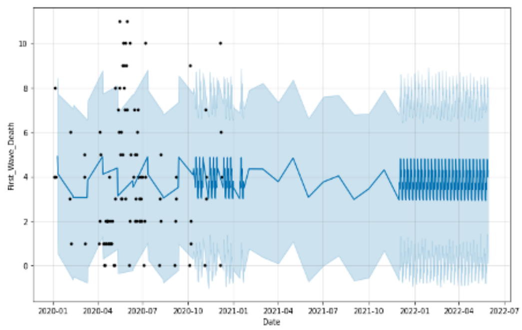


(d)

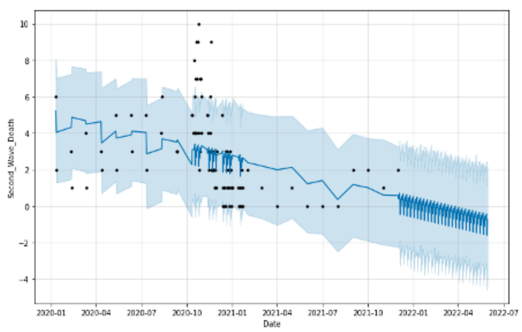
Fig. 5. (a) ARIMA forecast for first wave new cases, (b) ARIMA forecast for second wave transition cases, (c) ARIMA forecast for first wave deaths and (d) ARIMA forecast for second wave transition deaths.

3.4 Prophet® package Forecasting®

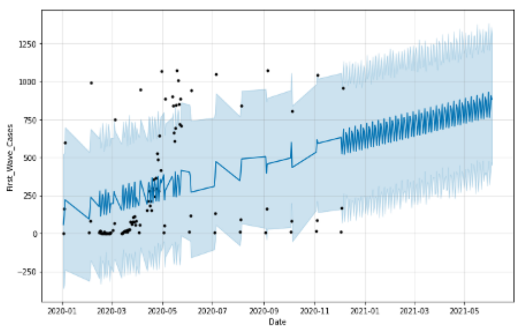
Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends fit yearly, weekly, and daily data seasonal and holiday effects. We observed that the death data has a better performance than the new case data. We present the visualisation and 100 days forecasting of the results in the Figures concerning (a) Prophet forecast for first wave deaths, (b) Prophet forecast for second wave transition deaths, (c) Prophet forecast for first wave cases, (d) Prophet forecast for second wave transition cases, (e) Trend plots for first wave deaths, (f) Trend plots for second wave transition deaths, (g) Trend plots for first wave cases and (h) Trend plots for second wave transition cases. Also we present in Tables 1 and 2 the predicted values for Kuwait daily new cases and deaths for October, 2020 along with predicted range. It is observed that some of the numerical values generated using this model are close to the observed values of the COVID-19 pandemic in Kuwait. The mean square error is given as follows: first wave cases = 79247.71, second wave transition cases = 40288.81. The MSE for first wave deaths = 9.04 and second wave transition deaths = 3.93.



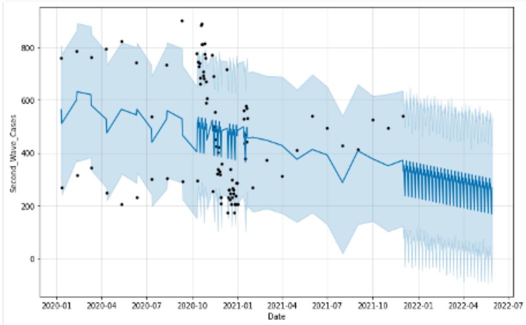
(a)



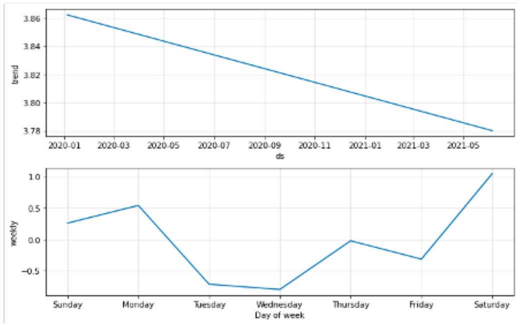
(b)



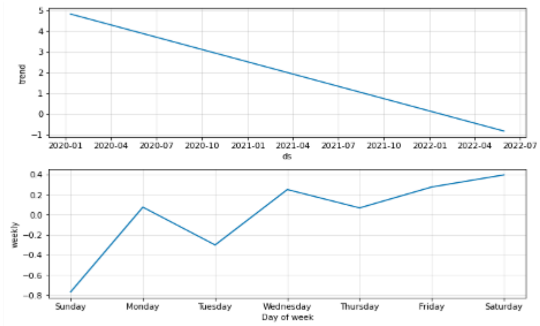
(c)


Second_Wave_Cases
Date

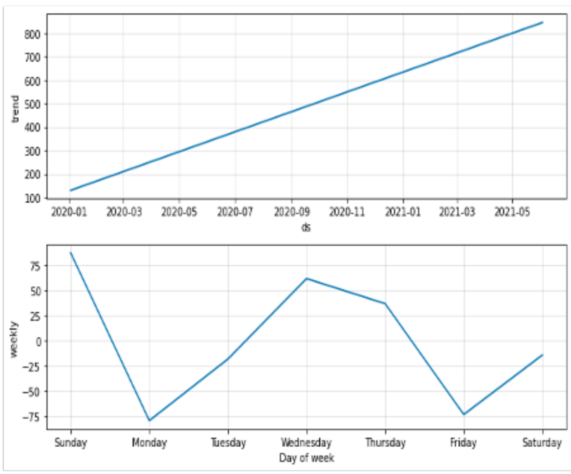
(d)


trend
ds
weekly
Day of week

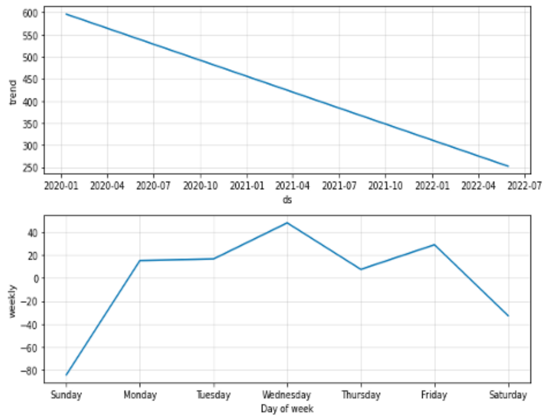
(e)


trend
ds
weekly
Day of week

(f)


trend
ds
weekly
Day of week

(g)

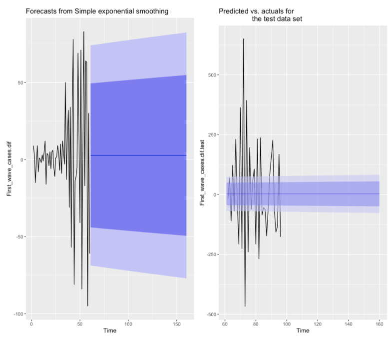


(h)

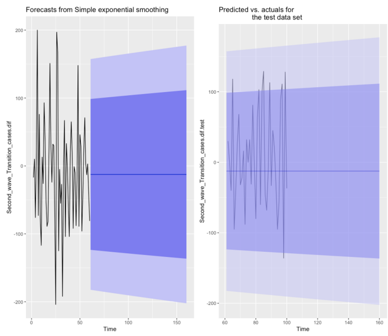
Fig. 6. (a) Prophet forecast for first wave deaths, (b) Prophet forecast for second wave transition deaths, (c) Prophet forecast for first wave cases, (d) Prophet forecast for second wave transition cases, (e) Trend plots for first wave deaths, (f) Trend plots for second wave transition deaths, (g) Trend for first wave cases and (h) Trend plot for second wave transition cases.

3.5 Exponential smoothing method

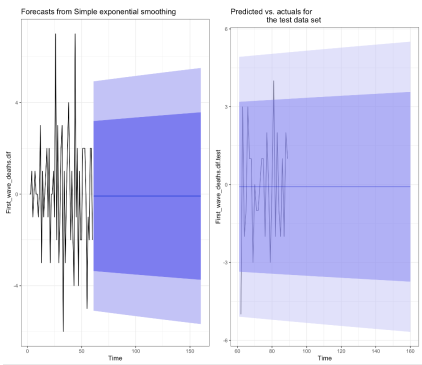
Simple exponential model (SES or ANN) is a good forecasting tool and in our case we used $\alpha = 0.2$ as one of our parameters while other parameters varies depending on the case in consideration. The degree of freedom (df) is 8 while the model degree of freedom is 2 with 10 lags. Also, 60% of the data was trained with 100 days forecasting while 40% was tested. We observed that while training and testing the model, the best exponential model is when the trend is removed from the test set as it can be seen in Figure 7 even though the test Root MSE is quite large but the p-value is low. Other exponential models like diff.SES (removing trend from SES), simple exponential smoothing with multiplicative errors (MNN) and simple exponential smoothing with additive errors (MAN) did not give better performance when we compared because their Root MSE were extremely larger than the simple exponential model without trend. The median of the residual is zero for all cases except first wave cases whose median is -55, which shows that residual is normally distributed. We present the parameters and result for each case as follows: for first wave cases the parameters used are $l = 0.48$ and $\sigma = 39$, Root MSE for training = 38.3, Root MSE for test = 219.9, $Q* = 78$ and p-value= $1e^{-13}$; for second wave transition cases the parameters used are $l = 0$ and $\sigma = 92.3$, Root MSE for training = 90.7, Root MSE for test = 76.2, $Q* = 30$ and p-value= $2e^{-04}$; for first wave deaths the parameters used are $l = 0.0106$ and $\sigma = 2.73$, Root MSE for training = 2.69, Root MSE for test = 2.07, $Q* = 29$ and p-value= $4e^{-04}$ and for second wave transition deaths the parameters used are $l = 0.4891$ and $\sigma = 2.49$, Root MSE for training = 2.45, Root MSE for test = 1.05, $Q* = 14$ and p-value= 0.09. We can conclude that the simple exponential smoothing model without trend works better for first wave deaths.



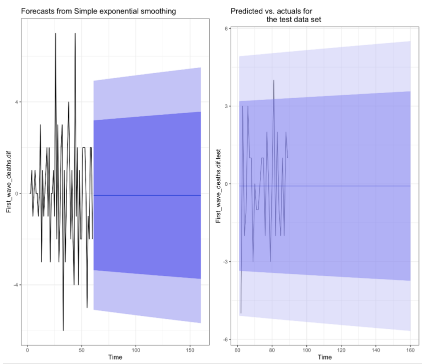
(a)



(b)



(c)



(d)

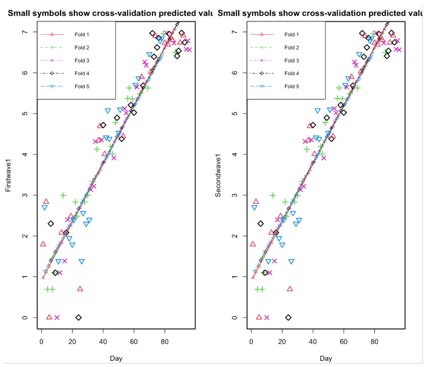
Fig. 7. (a) Simple exponential smoothing for first wave cases, (b) Simple exponential smoothing for second wave transition cases (c) Simple exponential smoothing for first wave deaths and (d) Simple exponential smoothing for second wave transition deaths.

3.6 Performance, accuracy and validation of the regression models

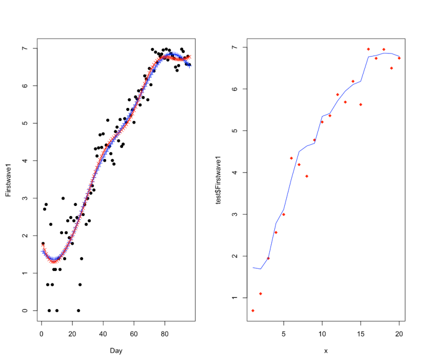
In order to know the performance of our regression models we trained 80% of the data and test 20% percent of the data and also did cross validation to be sure of the accuracy. The predicted and the observed values are very close to the result presented for all the regression models used in this article. For Log linear model we present the cross validation result in Figure 8a whose average mean square errors for the 5 portion folds are 0.5027791 for first wave on the left and 0.1533665 for second wave transition on the right. We observed high correlations between the tested and the predicted values (R-squared = 0.9278587 for first wave cases and R-squared = 0.5312499 for second wave transition cases) for both cases.

For polynomial regression of order four, we present the performance of the test model as follows: for first wave cases, multiple R-squared= 0.9437, p-value= $3.448e^{-09}$, relative standard error is 0.5152 and the residual median is 0.03408, value close to zero which shows that the model performs very well; for second wave transition cases, multiple R-squared= 0.9039, p-value= $1.829e^{-07}$, relative standard error is 0.194 and the residual median is 0.00386, value close to zero which shows that the model performs optimally.

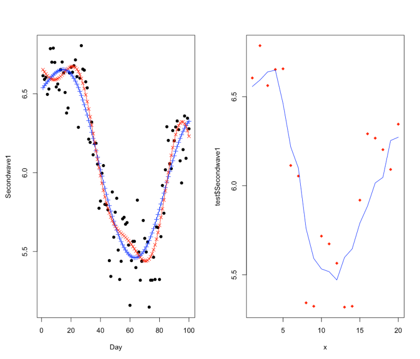
Lastly for support vector regression we present the optimum model with parameters ($\epsilon = 0$ for both cases, and cost = 4 and cost = 10) respectively, for the first and second wave transition cases in Figure 8d and 8e with mean square error values of 0.3840138 and 0.03 respectively using 10 folds cross validation. Root MSE for the first and second wave transition cases are 0.5757138 and 0.16 respectively, slope values of -0.0452387 and - 0.1503307 respectively, p-value $= 1.554e^{-15}$ for first wave cases, p-value = 0.293 for second wave transition cases and intercept values 3.751997 and 3.85671 respectively. We present a comparison of Root MSE for support vector model, tuned support vector model, constructed support vector model and also the test model in Figure 8b for first wave cases and Figure 8c for second wave transition cases. The left hand side of both figures is the tuned support vector model while the right hand side is the test model prediction. The test model performance is presented as follow: first wave cases: R-squared = 0.8450823 and Root MSE = 0.784234; second wave transition cases: R-squared= 0.8516843 and Root MSE = 0.1668129.



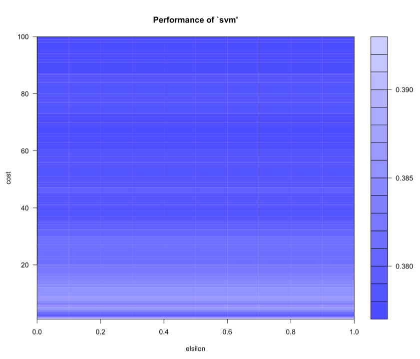
(a)



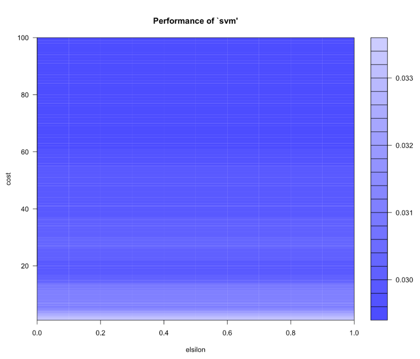
(b)



(c)



(d)

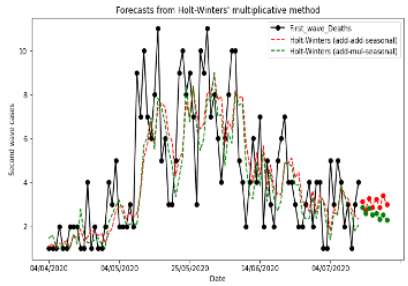


(e)

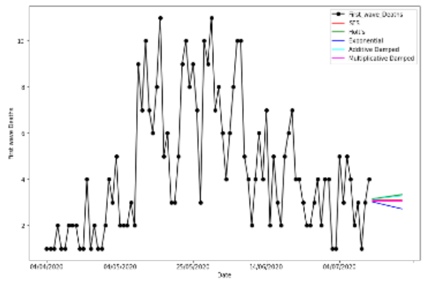
Fig. 8. (a) Cross validation result for Log linear model for new cases; (b) Comparison of the tuned support vector model and support vector regression model with the test prediction for first wave new cases; (c) Comparison of the tuned support vector model and support vector regression model with the test prediction for second wave transition new cases; (d) Optimum support vector model mean square error visualisation for first wave new cases and (e) Optimum support vector model mean square error visualisation for second wave transition new cases.

3.7 Holt's method

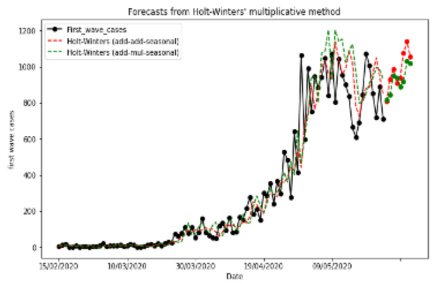
We have been able to demonstrate different types of exponential smoothing models with their comparison visualisation (see Figures 9b, 9d, 9f and 9h) to give significance for recent observations and produce accurate forecasts of 100 days while Holt's model (AAN) as shown in Figures 9a, 9c, 9e and 9g gives the trend and level of a time series and is computationally more efficient than double moving average. Holt's-Winters' model considers randomness using efficient smoothing process and is computationally efficient too. Holt's linear method with additive errors did not give better performance when we compared with the optimal Holt's model because the Root MSE was extremely larger. In subsequent Section we will give a precise evaluation for the training and test sets. We present the visualisation of the results in Figure 10.



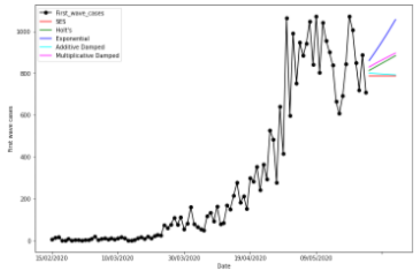
(a)



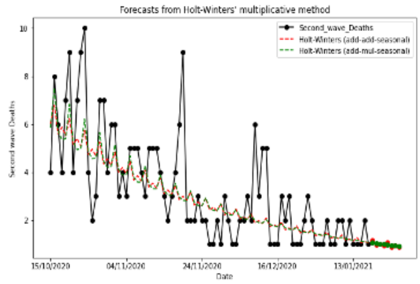
(b)



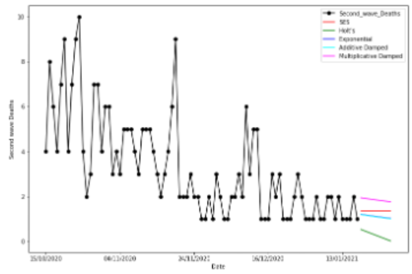
(c)


First_wave_cases
SES
Holt's
Exponential
Additive Damped
Multiplicative Damped
Date

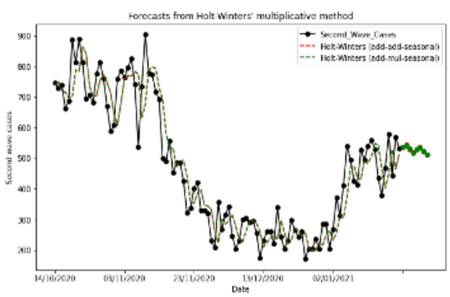
(d)


Forecasts from Holt-Winters' multiplicative method
Second_wave_Deaths
Holt-Winters (add-add-seasonal)
Holt-Winters (add-mul-seasonal)
Second wave Deaths
15/10/2020
04/11/2020
24/11/2020
16/12/2020
13/01/2021
Date

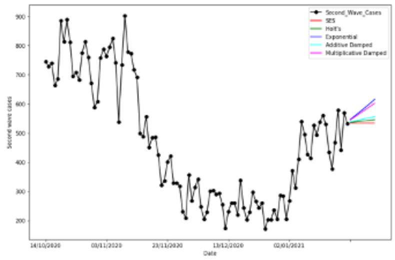
(e)


Second_wave_Deaths
SES
Holt's
Exponential
Additive Damped
Multiplicative Damped

(f)


Forecasts from Holt-Winters' multiplicative method
Second_Wave_Cases
Holt-Winters (add-add-seasonal)
Holt-Winters (add-mul-seasonal)
Second wave cases
14/10/2020
03/11/2020
23/11/2020
13/12/2020
02/01/2021
Date

(g)



(h)

Fig. 9. (a) Holt's method first wave deaths, (b) Comparison between different methods for first wave deaths, (c) Holt's method first wave cases, (d) Comparison between different methods for first wave cases, (e) Holt's method second wave transition deaths, (f) Comparison between different methods for second wave transition deaths, (g) Holt's method second wave transition cases and (h) Comparison between different methods for second wave transition cases.

3.8 Performance, accuracy and validation of the exponential models

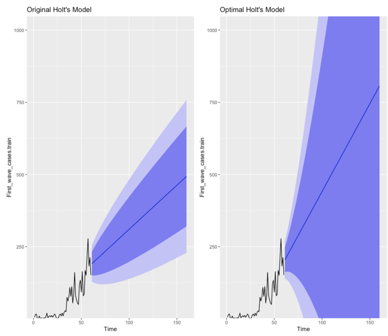
We give an explicit analysis of the performance and accuracy of the simple exponential without trend and Holt's model. Figures 10e to 10h gives the comparison between optimal Root MSE's of simple exponential model without trend and Holt's model. We trained 60% of the model while others were tested and 10 lags were used for the modelling. Figures 10a to 10d show visualisation of comparison between the holts model and the optimal Holt's model

- for first wave cases : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 1.3$ and $\sigma = 36.4$, with result of Root MSE for training = 35.8, Root MSE for test = 219.4, Holt's model parameters are $\alpha = 0.04095, l = 8.9848, b = 3.0532, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 31.8$, with result of Root MSE for training = 30.7, Root MSE for test = 534.4, $Q^* = 32$ and p-value= $1e^{-05}$, Holt's optimal model result of Root MSE for training = 21.3, Root MSE for test = 409.3;

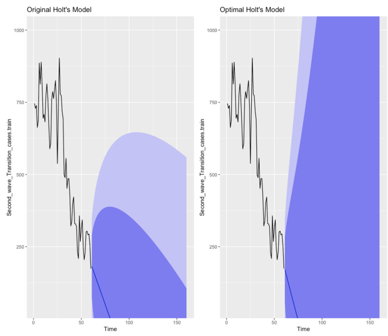
- for second wave transition cases : optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = -5.0588$, df= 8, model df = 2 and $\sigma = 86.7$, with result of Root MSE for training = 85.2, Root MSE for test = 73.2, $Q^* = 30$ and p-value= $2e^{-04}$, Holt's model parameters are $\alpha = 0.7798, l = 755.866, b = -9.4126, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 84.9$, with result of Root MSE for training = 82, Root MSE for test = 422, $Q^* = 35$ and p-value= $4e^{-06}$, Holt's optimal model result of Root MSE for training = 84.6, Root MSE for test = 504.5;

- for first wave deaths : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 0.107$, df= 8, model df = 2 and $\sigma = 2.55$, with result of Root MSE for training = 2.51, Root MSE for test = 2.06, $Q^* = 27$ and p-value= $6e^{-04}$, Holt's model parameters are $\alpha = 0.5016, l = 0.6829, b = 0.0621, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 2.38$, with result of Root MSE for training = 2.3, Root MSE for test = 271, $Q^* = 27$ and p-value= $1e^{-04}$, Holt's optimal model result of Root MSE for training = 2.40, Root MSE for test = 2.46;

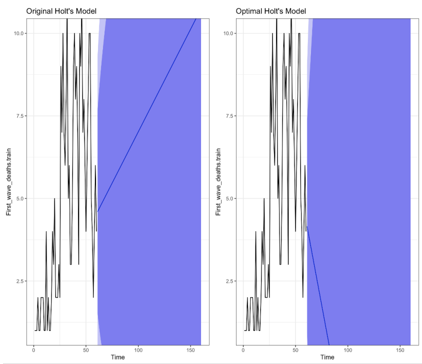
- for second wave transition deaths : Optimal model for simple exponential model without trend parameters are $\alpha = 0.05, l = 0.0422$, df= 8, model df = 2 and $\sigma = 2.32$, with result of Root MSE for training = 2.28, Root MSE for test = 0.94, $Q^* = 13$ and p-value= 0.1, Holt's model parameters are $\alpha = 0.0004, l = 6.8362, b = -0.0891, \beta = 0.0004$, df= 6, model df = 4 and $\sigma = 1.92$, with result of Root MSE for training = 1.85, Root MSE for test = 1.50, $Q^* = 8$ and p-value= 0.2, Holt's optimal model result of Root MSE for training = 1.79, Root MSE for test = 1.79.


Original Holt's Model
Optimal Holt's Model
Time
Time

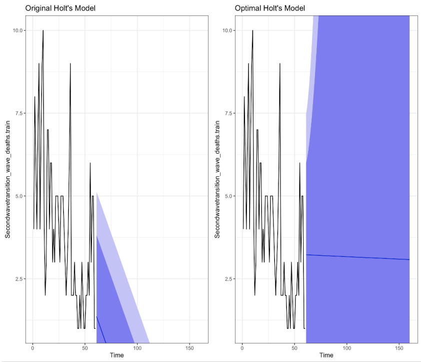
(a)


Original Holt's Model
Optimal Holt's Model
Time
Time

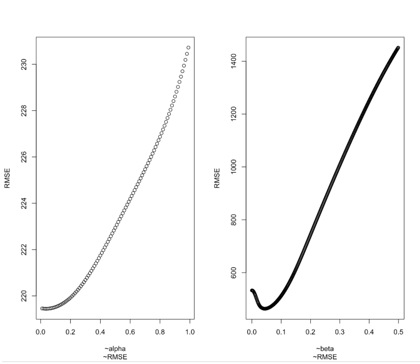
(b)


Original Holt's Model
Optimal Holt's Model
Time
Time

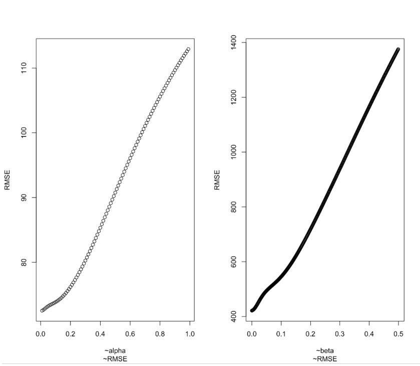
(c)


Original Holt's Model
Optimal Holt's Model
Secondwavetransition_wave_deaths.train
Time

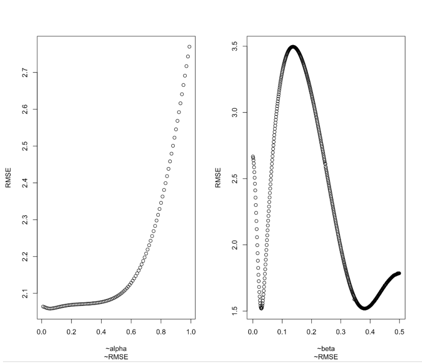
(d)


RMSE
~alpha
~RMSE
~beta
~RMSE

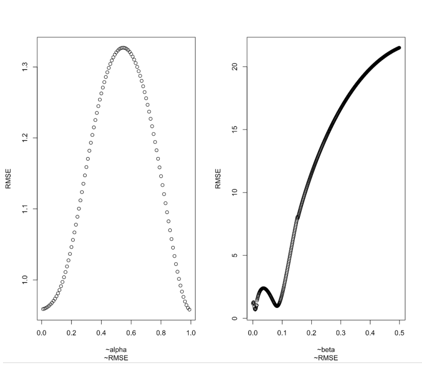
(e)


RMSE
~alpha
~RMSE
~beta
~RMSE

(f)



(g)



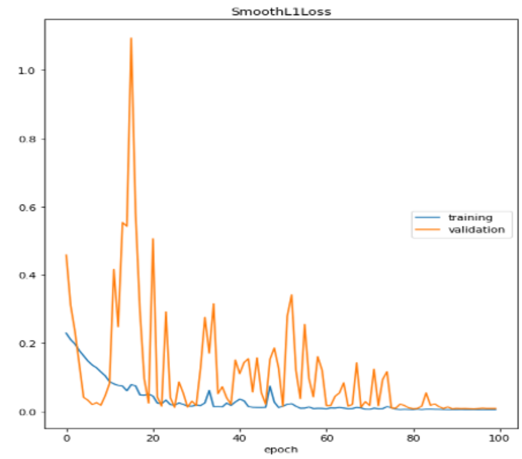
(h)

Fig. 10. (a) Comparison of Holt's model and optimal Holt's model forecast for first wave cases, (b) Comparison of Holt's model and optimal Holt's model forecast for second wave transition cases, (c) Comparison of Holt's model and optimal Holt's model forecast for first wave deaths, (d) Comparison of Holt's model and optimal Holt's model forecast for second wave transition deaths, (e) Comparison of simple exponential model without trend optimal Root MSE and Holt's model optimal Root MSE for first wave cases, (f) Comparison of simple exponential model without trend optimal Root MSE and Holt's model optimal Root MSE for second wave transition cases, (g) Comparison of simple exponential model without trend optimal Root MSE and Holt's model optimal Root MSE for first wave deaths and (h) Comparison of simple exponential model without trend optimal Root MSE and Holt's model optimal Root MSE for second wave transition deaths.

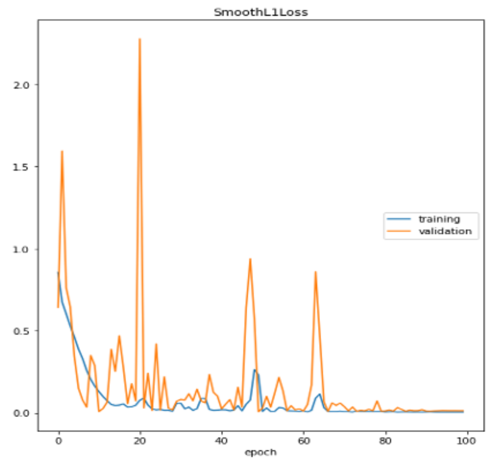
To conclude this Section we give a critical look at the Root MSE comparison in Figures 10e to 10h because it guides the choice of parameters to be used in the model. In Figure 10e we observed that for simple exponential model without trend on the left, if the α value is increased beyond 0.2, the Root MSE value is on the increasing trend while for Holt's model on the right, it is best to use β of about 0.05 and after then the Root MSE is on the increasing trend. In Figure 10f we notice that for the simple exponential model without trend on the left the choice for α must be below 0.2 and after that the Root MSE continues to increase while for Holt's model on the right, β must be below 0.1 else the Root MSE value becomes large. The Figure 10g is a bit tricky because for the choice of α makes the values of Root MSE to be stationary till the point where α is 0.6 where the increasing trend begins for simple exponential model without trend on the left while for Holt's model on the right there was turning point when the value of β is 0.15 but the value gives minimum Root MSE when $\beta = 0.04$ and $\beta = 0.38$, which means that the choice of β for this case is very critical and must be precise. In Figure 10h, the simple exponential model on the left shows that there was a turning point at $\alpha = 0.5$ with the highest Root MSE and the least Root MSE is for values $\alpha = 0$ and $\alpha = 1.0$. For Holt's model the minimum Root MSE values is for $\beta = 0$ and $\beta = 0.1$, and any other values increases the Root MSE value.

3.9 Neural Prophet method

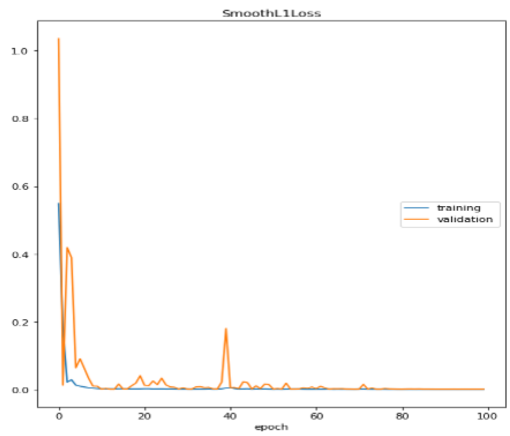
We have also presented in Figures 11a to 11d a loss plot for our neural network model which helps to know the performance of the model. It was observed that our loss plot shows a good fit and convergence of the model. The plot of training loss decreases to a point of stability and also the plot of validation loss decreases to a point of stability and has a small gap (generalisation gap) with the training loss. In Figures 11e and 11f we provide the visualisation of the neural forecast of 100 days for the first wave cases and second wave transition cases which aligns with the trend of results we have presented in other forecasting results in the previous Section, and also from the observed results from worldometer, we noted a decrease in daily cases from end of April to May 2021 and this also aligns with the result presented in Table 2.



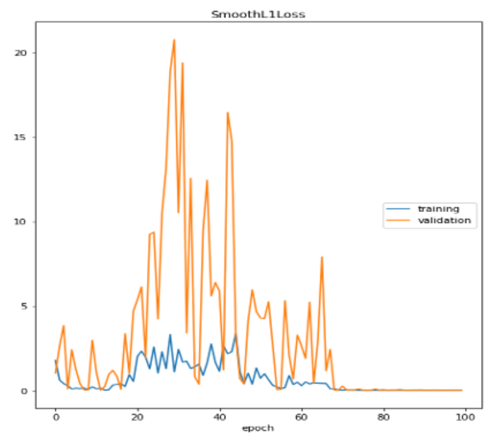
(a)



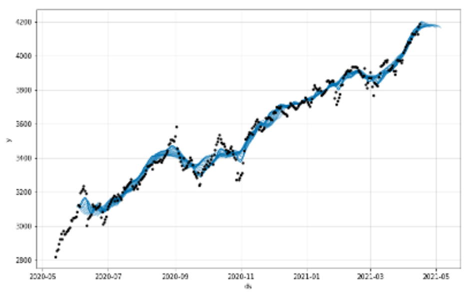
(b)


SmoothL1Loss
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
epoch
training
validation

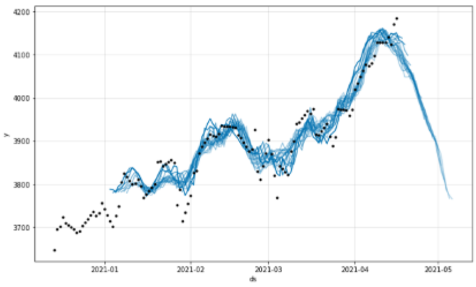
(c)


SmoothL1Loss
20
15
10
5
0
0
20
40
60
80
100
epoch
training
validation

(d)


4200
4000
3800
3600
3400
3200
3000
2800
y
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
ds

(e)



(f)

Fig. 11. (a) Loss plot for second wave transition deaths, (b) Loss plot for second wave transition cases, (c) Loss plot for first wave deaths, (d) Loss plot for first wave new cases, (e) Neural forecast for first wave cases and (f) Neural forecast for second wave transition cases.

We present the predicted values for Kuwait daily new cases and deaths alongside the observed values.

Table 1. Kuwait number of daily deaths predicted for October, 2020 (Note: PR means predicted range, OV means Observed value from worldometer and PV means predicted value).

Date	PR	OV	PV	In range?
11/10/20	$0 \sim 5$	3	2	YES
12/10/20	$0 \sim 6$	6	3	YES
15/10/20	$0 \sim 6$	8	3	NO
16/10/20	$0 \sim 6$	6	3	YES
17/10/20	$1 \sim 6$	4	3	YES
18/10/20	$0 \sim 5$	7	2	NO
19/10/20	$0 \sim 6$	9	3	NO
20/10/20	$0 \sim 5$	4	3	YES
21/10/20	$1 \sim 6$	7	3	NO
22/10/20	$0 \sim 6$	9	3	NO
23/10/20	$0 \sim 6$	10	3	NO
24/10/20	$1 \sim 5$	4	3	YES
25/10/20	$0 \sim 5$	2	2	YES
26/10/20	$0 \sim 6$	3	3	YES
27/10/20	$0 \sim 5$	7	3	NO
28/10/20	$0 \sim 6$	7	3	NO
29/10/20	$0 \sim 6$	4	3	YES
30/10/20	$0 \sim 6$	6	3	YES
31/10/20	$0 \sim 6$	6	3	YES

Table 2. Kuwait number of daily deaths and daily cases predicted for April and May, 2021 (Note: PRD means predicted range for daily deaths, PVD means predicted values for daily deaths, PRC predicted range for daily cases and PVC means predicted value for daily cases).

No	Date	PRD	PVD	PRC	PVC
1	17-04-21	1 ∼ 8	5	344 ∼ 1223	766
2	18-04-21	1 ∼ 8	4	467 ∼ 1303	869
3	19-04-21	1 ∼ 8	3	267 ∼ 1164	704
4	22-04-21	0 ∼ 7	4	380 ∼ 1276	824
5	23-04-21	0 ∼ 7	3	237 ∼ 1172	716
6	24-04-21	1 ∼ 9	5	345 ∼ 1241	776
7	25-04-21	0 ∼ 8	4	445 ∼ 1303	879
8	26-04-21	1 ∼ 8	4	268 ∼ 1131	714
9	27-04-21	0 ∼ 7	3	336 ∼ 1233	776
10	28-04-21	0 ∼ 7	3	397 ∼ 1280	857
11	29-04-21	0 ∼ 8	4	395 ∼ 1271	834
12	30-04-21	0 ∼ 7	3	281 ∼ 1166	725
13	01-05-21	1∼ 8	5	340 ∼ 1216	786
14	02-05-21	0∼ 8	4	439 ∼ 1311	888
15	03-05-21	0 ∼ 8	4	282 ∼ 1151	723
16	04-05-21	0 ∼ 7	3	369 ∼ 1228	786
17	05-05-21	0 ∼ 7	3	455 ∼ 1313	867
18	06-05-21	0 ∼ 8	4	387 ∼ 1274	844
19	07-05-21	0 ∼ 7	3	280 ∼ 1178	735
20	08-05-21	1 ∼ 8	5	357 ∼ 1215	795
21	09-05-21	0∼ 8	4	478 ∼ 1342	898
22	10-05-21	1 ∼ 8	4	291 ∼ 1155	733
23	11-05-21	0 ∼ 7	3	371 ∼ 1230	795
24	12-05-21	0 ∼ 6	3	457 ∼ 1345	877
25	13-05-21	0 ∼ 8	4	447 ∼ 1302	853
26	14-05-21	0 ∼ 7	3	330∼ 1211	744
27	15-05-21	1 ∼ 9	5	383 ∼ 1259	805
28	16-05-21	0 ∼ 8	4	448 ∼ 1369	908
29	17-05-21	1 ∼ 8	4	308 ∼ 1193	743
30	18-05-21	0 ∼ 7	3	381 ∼ 1245	805
31	19-05-21	0 ∼ 7	3	450 ∼ 1333	886
32	20-05-21	0∼ 7	4	441 ∼ 1291	863
33	21-05-21	0 ∼ 7	3	299 ∼ 1184	754
34	22-05-21	1 ∼ 9	5	332 ∼ 1284	815
35	23-05-21	0 ∼ 7	4	467 ∼ 1354	917
36	24-05-21	1 ∼ 8	4	279 ∼ 1184	753
37	25-05-21	0 ∼ 7	3	362 ∼ 1262	815
38	26-05-21	0 ∼ 7	3	494 ∼ 1332	896
39	27-05-21	0 ∼ 7	8	406 ∼ 1299	873
40	28-05-21	0 ∼ 7	3	340 ∼ 1213	764
41	29-05-21	1 ∼ 9	5	367 ∼ 1241	824
42	30-05-21	0 ∼ 7	4	510 ∼ 1375	927
43	31-05-21	1 ∼ 8	4	328 ∼ 1207	762

Sensitivity Analysis

Here, we present the sensitivity analysis of the parameters used in the modelling of our study. The support vector model performance is sensitive to the choice of the cost function, γ and ϵ parameters for the data set and that is why we used the idea of cross validation and optimal model to choose the best parameters that best fit the model.

From Table 3, we present the sensitivity parameters for the cases we considered. We discovered that for ARIMA models, all the cases are more sensitive to the choice of the order of the auto regression and moving average while it is least sensitive to the trend difference.

For exponential models, the data is very sensitive to the choice of σ and α but least sensitive to the number of lags and choices of β. The optimal algorithm was able to choose the best parameter that fit the model.

Table 3. The sensitive parameters for the analysis showing their Root MSE results (Note: FWC means first wave new cases, SWC means transition to second wave new cases, FWD means first wave deaths and SWD means transition to second wave deaths).

FWC	SWC	FWD	SWD
ARIMA (6,1,0) = 281.5	ARIMA (6,1,0) = 200.7	ARIMA (6,1,0) = 3.0	ARIMA (6,1,0) = 2.0
ARIMA (2,1,0) = 98.1	ARIMA (0,1,2) = 73.1	ARIMA (2,1,3) = 1.9	ARIMA (0,1,1) = 1.6
ARIMA (0,1,1) = 30.7	Holt's = 82.0	ARIMA (1,1,1) = 2.4	ARIMA (0,1,0) = 0.9
Holt's = 30.7	SES = 76.8	Holt's = 2.3	MNN = 1.7
SES = 31.4	diff.SES = 90.7	SES = 2.3	Holt's = 1.9
MAN = 111.0	-	MAN = 2.2	SES = 1.9
diff.SES = 38.3	ARIMA (0,1,2) with drift = 74.5	diff.SES = 2.7	diff.SES = 2.5

4. Discussion and conclusion

We have proposed a set of different regression methods in order to find the best ones in the Kuwait context, both for daily new cases and deaths, without a priori about the degree of non-linearity and the stochastic structure of noise behind the data. Surprisingly, we discovered that often the best regression method was the support vector one and that the stochasticity of the data at the start of the two waves was the same. This result confirmed our choice of comparing several regression methods (exponential regression being the most commonly, and often the only one, chosen) and showed that once the trend due to epidemic dynamics and its seasonality has been removed, the random factors explaining the variations compared to the deterministic model of Bernoulli-Ross-McKendrick (namely the uncertainties related to the counting of new cases and deaths) were expressed through a similar noise for the first and the second wave.

The two phases of the COVID-19 outbreak in Kuwait present differences:

- the first wave start shows an increase of the daily new cases (with a peak of 1000 daily new cases) during about 60 days with a slope of the exponential regression equal to 0.07, followed by an endemic phase with a high mean number of daily new cases (about 600) and with a delay of about 15 days an increase of deaths,

- the second wave transition shows a decrease of the new cases during about 50 days with a slope of the exponential regression equal to -0.009, followed by an increase during about 20 days. But these two phases show a certain homogeneity in their stochastic structure at their beginning, because they have the same level for the variation coefficient (around 0.5), both for daily new cases number and for death number, and for the autocorrelation initial slope (-0.031 for the first phase and -0.038 for the second).

In most of the examples studied in this article, the best regression for prediction is support vector regression and the best forecasting model is the simple exponential model without trend for Kuwait COVID-19 data. Also from our result we discovered that all models work better for the daily deaths case than for the daily new cases as well.

Data Availability

The code and data for replication of the analysis is available in the link below:
https://github.com/Honkay/Kuwait-time-series-project.git

Acknowledgements

The authors thank Professor Pierre Magal and Dr Hervé Seligmann for many fruitful discussions. The authors also wish to acknowledge the Petroleum Technology Development Fund (PTDF) Nigeria doctoral fellowship in collaboration with Campus France Africa Unit.

References

Worldometer. (2021) Available online: https://www.worldometers.info/coronavirus/ (accessed on 10th April 2021).

Demongeot, J., Flet-Berliac, Y., & Seligmann, H. (2020). Temperature decreases spread parameters of the new covid-19 cases dynamics. *Biology (Basel)*, **9:94**.

Seligmann, H. , Iggui, S. , Rachdi, M. ,Vuillerme, N., & Demongeot, J. (2020). Inverted covariate effects for mutated 2nd vs 1st wave Covid-19: high temperature spread biased for young. *Biology (Basel)*, **9:226**.

Demongeot, J., Oshinubi, K., Seligmann, H., & Thuderoz, F. (2021). Estimation of Daily Reproduction Rates in COVID-19 Outbreak. *MedRxiv*, doi.org/ 10.1101/2020.12.30.20249010.

Wiener, N. (1949).Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The MIT Press: Cambridge (Mass.), USA.

Rafferty, N. (2021). Forecasting Time Series Data with Facebook Prophet. Packt Publishing: Birmingham, UK.

Rhodes, C.J., & Demetrius, L. (2010). Evolutionary Entropy Determines Invasion Success in Emergent Epidemics. *PLoS ONE*, **5:701**.

Demongeot, J., & Demetrius, L. (2015). Complexity and Stability in Biological Systems. *Int. J. Bifurcation and Chaos*, **25:40013**.

Renkulab. (2021) Available online: https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/we213563a /taecdcpred (accessed on 22nd February 2021).

Chatfield, C. (2000). Time-Series Forecasting. Chapman & Hall, Boca Raton.

Abioye, A.I., Umoh, M.D., Peter, O.J., Edogbanya, F.A., Oguntolu, O., Oshinubi, K., & Amadiegwu, S. (2021). Forecasting of COVID-19 pandemic in Nigeria using real statistical data. *Commun Math Biol Neurosci*, **Article ID 2**.

Abenvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, **Article 105340**.

Anastassopoulou, C., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*, **15, Article e0230405.** https://doi.org/10.1371/journal.pone.0230405.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. *Holden day*, **D1981, 1057e1072**.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, **134, 1e12.** https://doi.org/10.1016/j.chaos.2020.109761.

Granger, K., & Newbold, J. (1986). Forecasting economic time series. USA: Academic Press.

Gupta, R., & K Pal, S. (2020). Trend analysis and forecasting of COVID-19 outbreak in India. *MedRxiv*, doi.org/10.1101/2020.03.26.20044511. preprint.

Grasselli, G., Pesenti, A., & Ciccozzi, M. (2020). Critical care utilization for the COVID-19 outbreak in lombardy, Italy: Early experience and forecast during an emergency response. *Journal of the American Medical Association*, **323(16), 1545e1546**.

Massonnaud, C., Roux, J., & Crepey, P. (2020). Forecasting short term hospital needs in France. *medRxiv*, doi: https://doi.org/10.1101/2020.03.16.20036939. preprint.

Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G. N., Campana, E. F., Toraldo, G., & Siettos, C. (2020). Tracing day-zero and forecasting the fade out of the COVID-19 outbreak in lombardy, Italy: A compartmental modelling and numerical optimisation approach. *PLoS One*, doi: 10.1371/journal.pone.0240649.

Wise, T., Zbozinek, T. D., Michelini, G., & Hagan, C. C. (2020). Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. *Royal open science*, https://royalsocietypublishing.org/doi/10.1098/rsos.200742.

Mohammed Abdul Majid, Omar Hafidz, M., & Noorani, M.S.M. (2020). Analysis and predictive validity of Kelantan River Flow using RQA and time series analysis *Kuwait Journal of Science*, **48:1**. doi: 10.48129/kjs.v48i1.8899.

Munir, M., Panni, U. S. K., Kausar, N. & Anjum, R. (2021). Sensitivity analysis of solute kinetics in a four compartmental model for hemodialysis patients. *Kuwait Journal of Science*, **48:3**. doi: 10.48129/kjs.v48i3.9085.

Submitted **:** 01/06/2021
Revised **:** 29/11/2021
Accepted **:** 01/12/2021
DOI **:** 10.48129/kjs.splcov.14501