TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NUMBER: 075MSICE001

PREDICTION OF COVID-19 CASES IN NEPAL USING THE COMBINATION OF EPIDEMIOLOGICAL AND TIME SERIES MODELS

Anita Sharma

075MSICE001

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
LALITPUR, NEPAL

AUGUST, 2021

COPYRIGHT©

The author has agreed that the library, Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk Campus, may make this thesis freely available for inspection. Moreover the author has agreed that the permission for extensive copying of this thesis work for scholarly purpose may be granted by the professor, who supervised the thesis work recorded herein or, in their absence, by the Head of the Department, wherein this thesis was done. It is understood that the recognition will be given to the author of this thesis and to the Department of Electronics and Computer Engineering, Pulchowk Campus in any use of the material of this thesis. Copying of publication or other use of this thesis for financial gain without approval of the Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk Campus and author's written permission is prohibited.

Request for permission to copy or to make any use of the material in this thesis in whole or part should be addressed to:

Head

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk Campus

Pulchowk, Lalitpur, Nepal

DECLARATION

I declare that the work hereby submitted for Master of Science in Information and Communication Engineering (MSICE) at IOE, Pulchowk Campus entitled **"Prediction of COVID-19 cases in Nepal using the combination of Epidemiological and Time series models"** is my own work and has not been previously submitted by me at any university for any academic award.

I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

ANITA SHARMA

075MSICE001

AUGUST, 2021

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PULCHOWK CAMPUS
DEPARTMENT OF ELECTRONICS AND COMMUNICATION AND COMPUTER ENGINEERING

The undersigned certify that the major project entitled "**Prediction of COVID-19 cases in Nepal using the combination of Epidemiological and Time series models**" submitted by **Miss Anita Sharma** to the Department of Electronics & Communication and Computer Engineering in partial fulfillment of requirement for the degree of Master of Science in Information and Communication Engineering. The project was carried out under special supervision and within the time frame prescribed by the syllabus and has been accepted as a bonafide record of work independently carried out by her in the department.

…………………………………………………

Supervisor, Dr. Aman Shakya

Department of Electronics and Communication Engineering

……………………………….

External Examiner, Manoj Ghimire

……………………………….

Committee Chairperson, Dr. Basanta Joshi

Coordinator, Information and Communication Engineering

Department of Electronics and Communication Engineering

…………………………….

Date

ABSTRACT

This work analyze the official data of coronavirus (Infected, Recovered and Death) and predict the evolution of the epidemic in Nepal. The generalized SEIR model has been applied with hybrid of ETS-ARIMA time series model for the time series analysis and predictions of evolution of Covid-19 cases (Quarantined, Recovered and Deaths). The prediction has been made for 30 days using the past data of thirteen months.

The prediction made by generalized SEIR model has been corrected using two time series models, ETS and ARIMA model. The estimation error of generalized SEIR model is fed to ETS model to predict the error. Then, the predicted error by ETS model is added to the prediction made by generalized SEIR model. Now, the remaining error is again fed to ARIMA model to predict the error. The predicted error by ARIMA model is added to the prediction made by generalized SEIR model to get final prediction. Use of generalized SEIR model along with ETS and ARIMA model improve the time series prediction of coronavirus spread in case of Nepal as compared to generalized SEIR model. Also, the SEIR-ETS-ARIMA model reduce the estimation error as compared to SEIRD-ARIMA model. Improvement in all quality measures, MAE, MSE, RMSE and MAPE, has been observed.

Keywords:
COVID-19, SEIR, ETS, ARIMA, SEIRD-ARIMA

ACKNOWLEDGEMENT

At the outset, I am very thankful to the Department of Electronics and Computer Engineering, IOE, Pulchowk Campus for providing us very insightful suggestions at the beginning of the project title selection phase.

I would like to extend my sincere gratitude to Dr. Aman Shakya, Dr. Nanda Bikram Adhikari and Dr. Basanta Joshi for their guidance and advice that aided me in the preparation of this project. I would always appreciate your suggestions for the improvement of this project.

Sincerely,
Anita Sharma
075MSICE001

CONTENTS

TABLES INDEX

FIGURES INDEX

ABBREVIATION

AIC	Akaike Information Criterion
ARIMA	Auto- Regressive Integrated Moving Average
COVID19	Corona Virus Disease 2019
ETS	Error Trend Seasonal
HEOC	Health Emergency Operation Centre
HWA	Holt Winter's Additive
ICU	Intensive Care Unit
KNN	K-Nearest Neighbour
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MERS	Middle East Respiratory Syndrome
MSLE	Mean Square Logarithm Error
NAR	Non-linear Auto Regressive
NN	Neural Network
NRMSE	Normalize Root Mean Square Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Auto- Regressive Integrated Moving Average
SARS-COV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SEIR	Susceptible-Exposed-Infective-Recovered
SES	Simple Exponential Smoothing
SIR	Susceptible-Infective-Recovered
SSM	Statistical SARIMAX model
SVR	Support Vector Regression
USD	United State Dollar
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

1.1 Background:

The COVID-19, first reported in Wuhan, China spread in nearly every country on the planet with more than 140 million global infection (April, 2021). COVID-19 was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020 [1].

Thailand was the first country to report the COVID-19 case outside of China while Nepal becomes the first in South Asian country. From banning international air travel to monitoring Nepal-China and Nepal-India ground crossing Points of Entry, from strict lockdown to partially executed lockdown, the Government of Nepal tries different strategy to prevent the outbreak. The intervention with vaccination has just been started.

COVID-19 is a disease that spreads rapidly and endangers the health of many people within a short period of time. COVID-19 is caused by a new form of coronavirus belonging to the coronavirus family along with MERS and SARS, which can spread to humans [2]. Fever, shortness of breath, cough, losing smell and taste and diarrhea are some common symptoms the infected person could show. COVID-19 has a two-week or longer incubation time [3]. In its latent period, the disease can still be contagious. The virus can be transmitted from person to person via respiratory droplets and close contact.

While Nepal is trying to put off the possible pandemic in the country but due to lack of clarity on strategy to be followed, its response is not showing such effective results. The trajectory of COVID-19 for Active, Recovered and Deaths was predicted for the cases of Nepal using generalized SEIR model in previous project work. The model we proposed here predicts the coronavirus spread in Nepal for next 30 days, expecting less erroneous than the previous work. The policy maker can use this prediction to find the number of quarantined bed, number of ICU and ventilators required for the next 30 days.

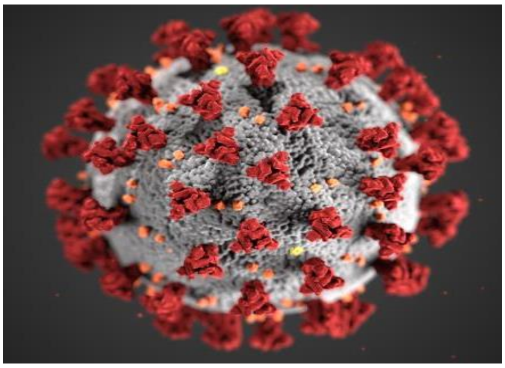


Figure 1.1 Ultrastructural morphology exhibited by coronavirus

Source: Centers for Disease Control and Prevention (CDC), US

1.2Problem Statement:

Maher Ala'raj, et al used SEIRD model to simulate COVID-19 outbreak in US and the prediction error is reduced by using ARIMA time series model [28]. This model consists of two parts: the modified SEIRD model and ARIMA models. The model fit SEIRD model parameters against historical values of infected, recovered and deceased population. Residuals of the first model for infected, recovered, and deceased populations are then corrected using ARIMA models. However, the hybrid model wouldn't handle the seasonality factor present in the data. To incorporate the seasonality factor, another time series model, the Error-Trend-Seasonality (ETS) model is added in between generalized SEIR and ARIMA model.

The epidemiological model incorporates parameters which describe the nature of coronavirus. The time series model, the ARIMA and ETS models can predict the time series evolution of the disease, but they have very few parameters which cannot represent the coronavirus spread in real scenario. The generalized SEIR model can give us the tentative idea on how the outbreak will go in future, but it cannot exactly predict the number of cases. Knowing the nature of coronavirus spread in Nepal, the generalized SEIR, ETS and ARIMA models separately would not efficiently predict the future value of COVID-19 cases. To address the limitations of epidemiological model as well as time series model, a new hybrid model of three layer, the SEIR-ETS-ARIMA is proposed. The proposed model also introduce the time and intervention dependency in the infection rate.

We have used three different models to address following issues.

i. Generalized SEIR model: To incorporate the parameters that describe the nature of coronavirus like protection rate, infection rate (mobility, population demography, and intervention), latent time of the virus, average quarantine time, recovery rate, deaths rate

ii. ETS model: To address any trend and/or seasonality present in the reported data

iii. ARIMA model: To incorporates random disruptions

1.3 Objectives

- To reduce the prediction error made by generalized SEIR model by using ETS and ARIMA statistical models
- To evaluate the SEIR-ETS-ARIMA hybrid model using active, recovered and deaths case of Nepal and validate the SEIR-ETS-ARIMA hybrid model by comparing with SEIRD-ARIMA hybrid model

1.4 Scope and limitation of work

The model can be used to predict the possibility of a second peak or to predict the eventual seasonal peaks. The work addresses all possible parameters that explains the nature of COVID-19 and other parameters that would affect the spread of the virus. The generalized SEIR model incorporates parameters like protection rate, infection rate (mobility, population demography, and intervention), latent time of the virus, average quarantine time, recovery rate, deaths rate. The time series model, the ARIMA and ETS incorporates parameters like time, auto-regression, moving average, trend coefficient, seasonality coefficient and no. of periods in seasonal cycle.

However, the work will not cover the case of new variant of virus separately. This work will not include the effect of vaccination over time. The total population will be considered constant, which means the natural birth and deaths are not considered in the proposed model.

CHAPTER 2: LITERATURE REVIEW

Mathematics in biology have made great contribution in modelling of epidemiological diseases like smallpox [4], tuberculosis transmission [5], Ebola [6], SARS pandemic [7] and the list goes on. After the outbreak of novel coronavirus in December 2019 in Wuhan, different studies is being carried out to find the nature of spread of the disease. Since the found intervention method (the vaccination) is under test till date, it is more important than ever to understand the current epidemiological models for disease spread, mortality, and recovery. The widely used compartmental model (SEIR model) have been used along with many variations to model the nature of coronavirus [8] [9] [10] [11] [12]. Improved version of SEIR model was designed by Shaobo. He et al by dividing Infective compartment into two class: the infectious without intervention and infectious with intervention, and considering the Quarantined and Hospitalized compartment to represent real scenario. They applied Particle Swarm Optimization algorithm to approximate the model's parameters. The model is applied to show SARS-COV-2 virus spread in Hubei, China [11]. Alberto Godio et al apply a generalized SEIR model, use PSO to fit the model parameters, and linked model equations to vary the infection rate for COVID-19 outbreak in Italy and its different region to enhance the accuracy of predictions for 30 days [12]. Although by using a very complex equations to represent the scenario of COVID-19 spread along with heuristic machine learning algorithm like PSO, they could only predict trend of the spread, not the exact number of cases.

Time series statistical models are extensively been used in forecasting since years. Oleg Ostashchuk used the ARIMA model to predict IBM stock price (in USD) [13]. C. A. Jofipasi et al forecast weather in the Aceh Besar District, Indonesia, using the ETS model [14]. In 2017, In Wuhan, China, the seasonal ARIMA model was used to forecast the occurrence of Hand-Foot-Mouth disease [15]. Similarly, infectious diseases like tuberculosis and Dengue fever were forecasted using ARIMA models [16] [17]. Leila Ismail, et al had an extensive case studies of 187 countries. They suggest best time series model with least RMSE and MAPE for each country for given dataset of COVID-19 [18].

After COVID-19 spread all over the world, along with epidemiological model, various time series model are being used in different research to predict the cases of COVID-19 in respective region. Ovidiu-Dumitru Ilie et al forecast the spread of coronavirus in nine different countries using ARIMA model. They use non-seasonal ARIMA (p,d,q) model [19]. Ram Kumar Singh et al. create a spatial map of the COVID-19 cumulative data for over 170 countries and territories. The spatial map is used to determine the severity of COVID-19 infections in the top 15 countries and continents [20]. Using ARIMA, X. Duan and Xi. Zhang model and forecast irregularly patterned covid-19 outbreaks using data from Japan and South Korea. For the ARIMA model, the Box-Jenkins method is used for model detection, estimation, diagnostic testing, and forecasting. On some nonstationary time series, the differencing transformation was used to achieve stationarity [21]. K.E. Arun Kumar, et. al. use ARIMA and SARIMA model separately to forecast the scenario of COVID-19 cases. They use RMSE, MAE, MAPE to select the best model and AIC, BIC to evaluate the model [22].

S. Makridakis et al compares various time series models and machine learning model to make prediction, and suggest that ETS and ARIMA are best to make time series prediction with less error [23]. Zhang GP use hybrid ARIMA-NN model for time series forecasting [24]. In time series forecasting of tourist travel, Aslanargun A, et al compared ARIMA, neural networks, and hybrid models [25]. Yu L, et al apply hybrid of SARIMA and NARNN to forecast the cases of HFMD in Shenzhen, China [26]. Aman Swaraj, et al used a hybrid model ARIMA-NAR in COVID-19 data of India, and suggest that using hybrid model significantly reduce RMSE, MAE and MAPE [27]. Maher Ala'raj, et al used SEIRD model to simulate COVID-19 outbreak in US and the prediction error is reduced by using ARIMA time series model. The hybrid model was used for short and long term forecast of the disease [28].

Table 2.1 Evolution of SEIR model

S.N.	Author/s	Year	Epidemiological Model
1.	Daniel Bernoulli	1760	Use of a simple mathematical method to evaluate the effectiveness (in terms of an improvement in life expectancy) of the technique of variolation to protect against smallpox infection
2.	Ronald Ross	1908	Pioneer model for transmission dynamics of malaria in continuous time framework, the SIR model
3.	Kermack and McKendrick	1927	SIR model for closed population with threshold density population
4.	Stavros N. Busenberg, Kenneth L. Cooke	1979	Effect of Incubation period in SIR model, delayed SIR or SEIR model
5.	Roy M. Anderson	1991	Partition of transmission coefficient into two different components – one representing the likelihood of transmission between a susceptible and an infected, and the other denoting the probability of contact between individuals in different groups

6.	Michael Y. Li et. al.	1999	SEIR model that incorporates exponential natural birth and death, as well as disease-caused death so that the total population size may vary in time
7.	Lekone et. al.	2006	Probabilistic approach focused on a stochastic discrete-time approximation to the SEIR method integrating control intervention to model Ebola epidemics
8.	Mukkai S. Krishnamoorthy et. al.	2010	Hybrid model for disease spread which discuss about local and global spread of the SARS pandemic. Local spread is largely correlated with population density and global spread is due to people's mobility
9.	Nuri Ozalp and Elif Demirci	2011	Fractional order SEIR model with transmission in a non-constant population. This solved the limitation of integer-order differential equations
10.	Syahrini et. al.	2017	Susceptible compartment is further divided into two – people with vaccination and without vaccination for tuberculosis transmission
11.	Zhou Tang et. al.	2020	Modify classical SEIR for coronavirus considering latent period is infectious in closed population
12.	Kiran Raj Pandey et. al.	2020	Age-structured SEIR model to investigate the effects of COVID-19 control intervention and finding of an active case
13.	Shaobo. He et. al.	2020	Infective compartment are divided into two compartments, the infectious without intervention and infectious with intervention, and considering the Quarantined and Hospitalized compartment
14.	Alberto Godio et. al.	2020	Generalized SEIR model, PSO is used to fit the model parameters

Table 2.2: Use of different statistical time series models and their variations

Author/s	Model	Application
Oleg Ostashchuk	ARIMA	Predict IBM stock price (in USD)
C. A. Jofipasi, et al	ETS	Forecast weather in the Aceh Besar District, Indonesia
Y. PENG, et al	SARIMA	Forecast the occurence of hand-foot-mouth disease in Wuhan, China.
Zhang X.	SARIMA	Typhoid fever forecasting

Ovidiu-Dumitru Ilie, et al	ARIMA	Forecast the spread of coronavirus in nine different countries.
X. Duan and Xi. Zhang	ARIMA, Box-Jenkins method, differencing transformation	Forecast covid-19 outbreaks using the data from Japan and South Korea.
K.E. Arun Kumar, et al	ARIMA, SARIMA, RMSE, MAE, MAPE, AIC, BIC	Forecast the dynamics of COVID-19 cases.
S. Makridakis, et al	ETS, ARIMA, SARIMA, SES, Holt, SVR, RNN, KNN, LSTM, etc.	Compares various time series models and machine learning model to make prediction, and suggest that ETS and ARIMA are best to make time series prediction
Leila Ismail, et al	SA, SMA, LT, QT, ST, DT, ARIMA, LSTM, HWA, SSM	case study of COVID-19 dynamics on 187 countries

Table 2.3 Hybrid models

Zhang GP	Hybrid ARIMA-NN	Time series forecasting of sunspot data
Aslanargun A, et al	ARIMA, neural networks and hybrid models	Time series in forecasting tourist travel
Yu L, et al	Hybrid SARIMA-NARNN	Forecast the cases of HFMD in Shenzhen, China
Aman Swaraj, et al	Hybrid ARIMA-NAR	Forecast COVID-19 data of India, and suggest that using hybrid model significantly reduce RMSE, MAE and MAPE
Maher Ala’raj, et al	Hybrid SEIRD-ARIMA	fit SEIRD model parameters for COVID-19 cases, residuals of SEIRD model are then corrected using ARIMA models, provide long and short-term forecasts with 95% confidence intervals

CHAPTER 3: METHODOLOGY

3.1 Proposed model:

For time series analysis of the COVID-19 outbreak in Nepal, our model employed the generalized SEIR model which is then corrected with an ETS-ARIMA hybrid statistical model, for the prediction of the infection, recovery and death case for next 10 days. The generalized SEIR model incorporates parameters like protection rate, infection rate (mobility, population demography, and intervention), latent time of the virus, average quarantine time, recovery rate and deaths rate. The epidemiological model give us the tentative idea on how the outbreak will go in future, but it cannot exactly predict the number of cases. ARIMA model incorporates trends, regular changes and even random disruptions and the ETS model comprises error, trends and seasonality. Thus these three models on stack will be appropriate for the prediction of COVID-19 cases. AIC estimator, ACF and PACF plots has been used to select appropriate ARIMA (p,d,q) and ETS decomposition method has been used to select appropriate ETS model.

A training set and a testing set are created from the dataset. To evaluate the model's performance, we train it on the training set and make predictions on the test set. The evaluation is done by calculating RMSE, MAE and MAPE between test set and predicted value. Then the model has been used to forecast future values. In our case the test set is the data of last 30 days, all previous data is in training set and the prediction is made for next 30 days.

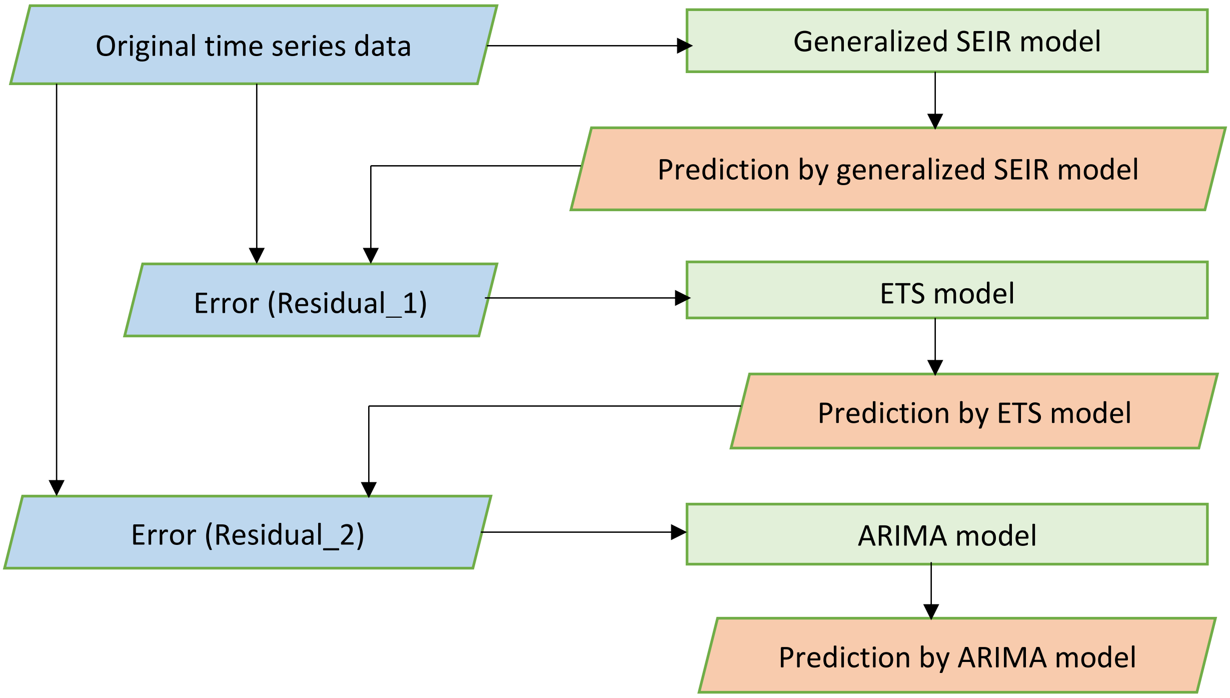


Figure 3.1 General Overview of Proposed Model

The general overview of proposed model is shown in figure 3.1. Using the original time series data, the generalized SEIR model make a prediction. The difference between the original time series data and the prediction made by generalized SEIR model is the first residual (Residual_1). Using the Residual_1 data, which is again a time series data, the ETS model makes another prediction. However, this prediction is the prediction of the error (Residual_1). The first correction (or the prediction made by the combination of the generalized SEIR and the ETS model) is made by adding the prediction made by generalized SEIR model with the prediction of Residual_1.

The difference between the original time series data and the prediction made by the combination of the generalized SEIR and the ETS model, is the second residual (or the Residual_2). Using the Residual_2 data, (a time series data) the ARIMA model makes another prediction. This prediction is the prediction of the error (Residual_2). The second correction (or the prediction made by the combination of the generalized SEIR, the ETS model and the ARIMA model) is made by adding the prediction made by generalized SEIR model with the prediction of Residual_2. The second correction is the output of SEIR-ETS-ARIMA model.

Figure 3.2 shows the detail representation of workflow of proposed model.

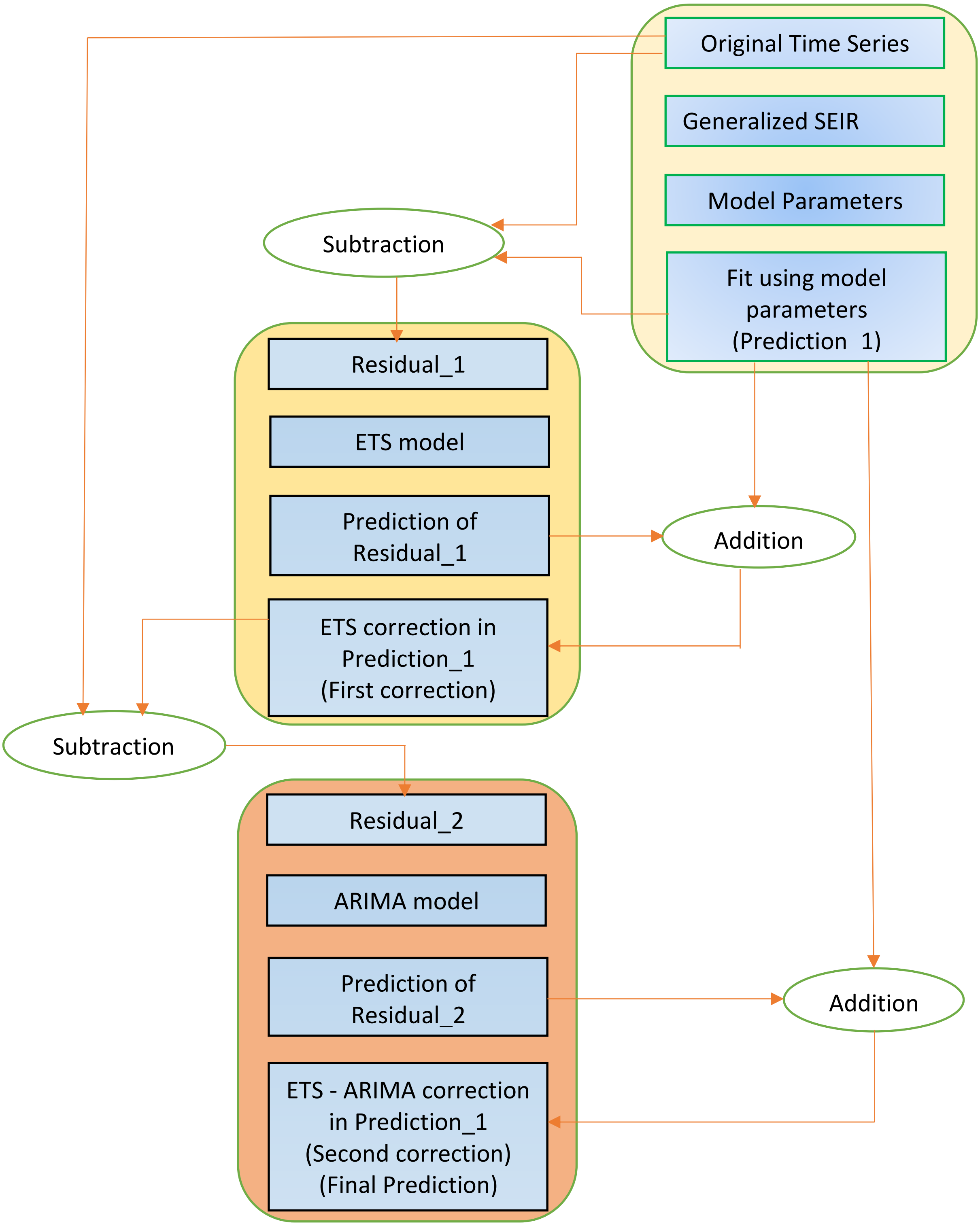


Figure 3.2: Graphical representation of proposed model

The proposed model use three layer of models, the generalized SEIR model, the ETS model and the ARIMA model. The infection rate in generalized SEIR model is a parameter which depends on number of other factors such as mobility, time and population density. The ETS model mainly handles the seasonality and trends on the evolution of COVID-19. The ARIMA model further make correction on the prediction made by generalized SEIR model and ETS model on stack. The ARIMA model basically corrects the prediction by following the nature of time series data. The hybrid of epidemiological and time series model minimizes the limitation of each of the models as compared to when used separately.

The Prediction_1 in the block diagram is the prediction of generalized SEIR model. Prediction_1 is subtracted from original time series data to get the first residual (Residual_1). Then, the ETS model is used to make correction in the prediction made by the generalized SEIR model. The input for the ETS model is the residual from the generalized SEIR model, i.e, Residual_1. The ETS model makes the prediction of Residual_1, which is denoted as "Prediction of Residual_1" in the block diagram. The first correction is made by adding "Prediction of Residual_1" with "Prediction_1". "ETS correction in Prediction_1" denotes the first correction; the corrected prediction by using generalized SEIR model with the ETS model.

The ARIMA model is used to make second correction. The residual of SEIR-ETS hybrid model is used as input for the ARIMA model. "ETS correction in Prediction_1" is subtracted from the original time series data to get second residual (Residual_2). The ARIMA model is implied on Residual_2 to make prediction of Residual_2, i.e, "Prediction of Residual_2". The second correction is made by adding "Prediction of Residual_2" with the prediction made by generalized SEIR model, i.e, "Prediction_1".

The epidemiological model cannot explain the time series nature of the data. The time series model doesn't incorporate the unique nature of coronavirus spread. Using three layer hybrid model minimizes the limitation of both epidemiological model and the time series model. The generalized SEIR model incorporates parameters like protection rate, infection rate (mobility, population demography, and intervention), latent time of the virus, average quarantine time, recovery rate and deaths rate. The epidemiological model give us the tentative idea on how the outbreak will go in future, but it cannot exactly predict the number of cases. ARIMA model incorporates trends, regular changes and even random disruptions and the ETS model comprises error, trends and seasonality. Thus these three models on stack is appropriate for the prediction of COVID-19 cases.

3.2 Generalized SEIR model:

The SEIR model in its classical form, models complex interaction of number of population between four different conditions, the susceptible (S), exposed (E), infective (I), and recovered (R). The generalized SEIR model adds new compartments Quarantined and Insusceptible and consider the key epidemic parameters for COVID-19 like the latent time, quarantine time and basic reproduction number.

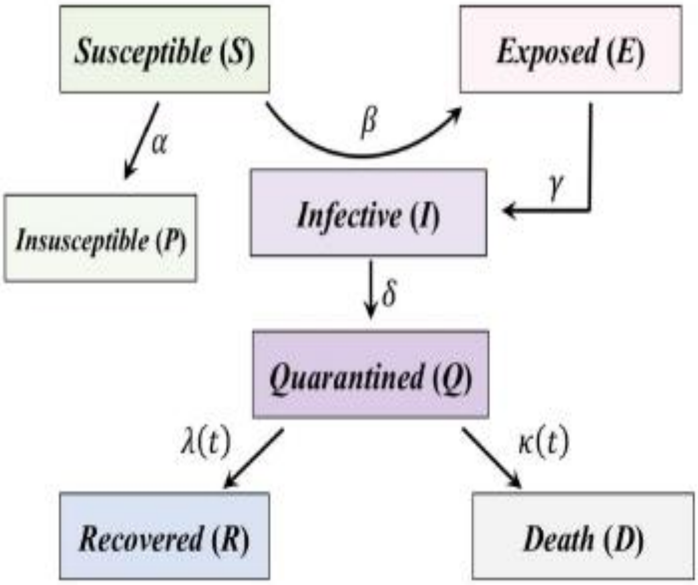


Figure 3.3: The generalized SEIR epidemic model for COVID-19

where,

S: susceptible cases

E: exposed cases

I: infective cases

Q: quarantined cases

R: recovered cases

D: death cases

P: insusceptible cases

The differential equations of generalized SEIR model is given below:

$$\left.\begin{aligned}
\frac{dS(t)}{dt} &= -\beta I(t).\frac{S(t)}{N} - \alpha S(t)\\
\frac{dE(t)}{dt} &= \beta I(t).\frac{S(t)}{N} - rE(t)\\
\frac{dI(t)}{dt} &= rE(t) - \delta I(t)\\
\frac{dQ(t)}{dt} &= \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t)\\
\frac{dR(t)}{dt} &= \lambda(t)Q(t)\\
\frac{dD(t)}{dt} &= \kappa(t)Q(t)\\
\frac{dP(t)}{dt} &= \alpha S(t)
\end{aligned}\right\}\quad \text{Equation 1}$$

The output of the model are α, β, r, δ, λ and κ parameters. These parameters are called problem unknowns. Here,

α : recovery rate. S* α gives number of people transferred from susceptible to the protected class each day.

β : infection rate. Rate of infection from an infective person.

r : 1/r is the average latent time . The time difference between exposure to infection and experiencing symptoms

δ : 1/ δ is the average time a person with symptoms need to be quarantined.

λ : recovery rate. Recovery rate is time dependent parameter.

κ: death rate. Death rate is time dependent parameter.

N: Total number of population

Since the health system can improve its capability to treat people over time, λ and κ are time-dependent parameter. λ and κ are required to fit an exponential function because as time increases, the death rate should be closer to zero and the recovery rate converges towards a constant value.

$$\left.\begin{aligned}
&Lambda\ Function\ 1 = \frac{a(1)}{1 + \mathrm{e}(a(2) * (t - a(3)))}\\
&Lambda\ Function\ 2 = \mathrm{a}(1) + \exp(-a(2) * (t + a(3)))
\end{aligned}\right\}\quad \text{Equation 2}$$

$$\left.\begin{aligned}
&Kappa\ Function\ 1 = \frac{a(1)}{\mathrm{e}(a(2) * (t - a(3)))} + \exp(-a(2) * (t - a(3))))\\
&Kappa\ Function\ 2 = a(1) * \exp((-a(2) * (t - a(3)))\hat{}2)\\
&Kappa\ Function\ 3 = \mathrm{a}(1) + \exp(-a(2) * (t + a(3))))
\end{aligned}\right\} \text{Equation 3}$$

Where,
For lambda function,
a(1): the final asymptotic value of the cure rate.
a(2): the rate of adaptation to the emergency
a(3): constant
For kappa function,
a(1): the initial value of the mortality rate
a(2): changed mortality rate with time.
a(3): constant

Originality:

We proposed the infection rate (β) to be dependent on time, mobility and intervention. Furthermore, the mobility depends on population density. Higher the population density higher will be the mobility, and mobility of people can transmit the disease from person to person. Also, the application of the intervention may lead to reduce infection rate. . β fits an exponential function because as time increases, the infection rate should be closer to zero if intervention is applied and converge towards a constant value if intervention is weak or is not applied.

$$\left.\begin{aligned}
&Beta\ Function\ 1 = k1.k2[\frac{a(1)}{\mathrm{e}\big(a(2) * \big(t - a(3)\big)\big)} + \exp(-a(2) * (t - a(3))))]\\
&Beta\ Function\ 2 = k1.k2[\frac{a(1)}{1 + \mathrm{e}\big(a(2) * \big(t - a(3)\big)\big)}]
\end{aligned}\right\} \text{Equation 4}$$

Where, k1 is correlation coefficient between population density and covid-19 case

k2 is correlation coefficient between intervention measure and covid-19 case

a(1) : the initial value of the infection rate

a(2) : changed infection rate with time.

a(3): constant

3.3 ARIMA model:

Auto- Regressive Integrated Moving Average (ARIMA) is basically constituent of three different section, AR, I and MA. ARIMA model is written as ARIMA (p,d,q) where p stands for the order of Auto-regression, d for difference and q for Moving-average. Akaike information criterion (AIC) determines the best value for p, d and q.

AR model: The previous time series observation is used to forecast the future value. The number of previous observations used to establish the AR model's order.

$$Y(t) = \phi 1 Y(t-1) + \phi 2 Y(t-2) + \cdots + \phi n Y(t-n) + \varepsilon(t) \longrightarrow \text{Equation 5}$$

Where, Φ is parameter that indicate the auto-regression, t is time, Y(t) is observed value at time t; ε(t) is value of a random shock as a function of t and n is past value.

Integrated: Any time series data that has to be modeled must be stationary which means that the statistical properties like mean, variance, seasonality, and so on are nearly constant over time. We should convert the dataset to a stationary series if it is not stationary. To make it stationary, a difference operation is performed. Differencing with previous d value indicate order d integration.

To confirm seasonality and stationarity, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) will be used. ACF determines whether the previous values in a series are related to the next one, whereas PACF highlights the degree of correlation between a variable and its lag.

$$delta, \Delta = Y(t) - Y(t-1) \longrightarrow \text{Equation 6}$$

Where, Y(t) is observed value at time t and Y(t-1) is observed value at previous time t-1.

MA model: The previous errors is used to make the future prediction. The number of previous errors used, determines the order of the MA model.

$$Y(t) = \theta 1 \varepsilon(t-1) + \theta 2 \varepsilon(t-2) + \cdots + \theta n \varepsilon(t-n) + \varepsilon(t) \longrightarrow \text{Equation 7}$$

Where, θ is parameter that indicate the moving average, t is time, Y(t) is observed value at time t, ε(t) is value of a random shock as a function of t and n is past value.

ARMA model: The ARMA model is a hybrid of AR and MA models. ARMA model expresses the current and previous values as well as their residuals in linear form. It is expressed as ARMA (p,q).

$$\left.\begin{aligned} Y(t) &= \propto + [\, \phi 1 Y(t-1) + \phi 2 Y(t-2) + \cdots + \phi n Y(t-n)] \\ &-[\theta 1 \varepsilon(t-1) + \theta 2 \varepsilon(t-2) + \cdots + \theta n \varepsilon(t-n)] + \varepsilon(t) \end{aligned}\right\} \text{Equation 8}$$

Where, α is constant, Φ and θ are parameters that indicate auto-regression and the moving average respectively, t is time, Y(t) is observed value at a time t, ε(t) is value of the random shock dependent by t and n is past value.

In ARIMA model, firstly the dataset should be made stationary if it is not using equation 2 (higher order can be achieved by using equation 2 repeatedly, if required). Then apply ARMA model in equation 4. The order of ARIMA will be determined by using model selection method.

ARIMA model selection:

The Akaike Information Criterion (AIC) will be used for the model selection. For the given data and given sets of models, the AIC estimates the quality of each model. AIC score compares different models and determine the best model among given models and dataset. Lower the AIC score, better is the model.

In time series analysis, the most recent data is the most valuable data. But this data is often stuck in the test set and validation set. Therefore traditional train-validation-test method of model selection cannot select the best model. We can train a model on all the data and use the AIC for improved model selection.

$$AIC = -2\ln(L) + 2k \quad \longrightarrow \quad \text{Equation 9}$$

Where L is likelihood and k is the number of parameters

For the given model, log-likelihood measures how likely one is to their observed data. The best-fit model has the maximum likelihood. AIC is low for models with high log-likelihoods. For models with higher parameter complexity, a penalty term 2k is added.

AIC score is the probabilistic ranking of the models that are likely to reduce the information loss. After calculating the AIC score for each possible ARIMA model by varying p, d and q, the probability that the ith model reduces the information loss can be calculated as,

$$p = \exp(\frac{AICmin - AICi}{2}) \quad \longrightarrow \quad \text{Equation 10}$$

Where, AIC$_{min}$ is the lowest AIC score.

Lower value of the p indicates the better model.

3.4 ETS model:

The ETS models comprises of three component, error component (E), trend component (T), and seasonal component (S).

$$\left.\begin{aligned}
&\textbf{Forecast Error: } E(t) = Y(t-1) - Z(t-1)\\
&\textbf{Trend: } T(t) = \gamma\big(Z(t) - Z(t-1)\big) + (1-\gamma)T(t-1)\\
&\textbf{Seasonality: } S(t) = \delta\frac{Y(t)}{Z(t)} + (1-\delta)S(t-s)
\end{aligned}\right\} \quad \text{Equation 11}$$

Where, Y(t) is observed value at time t, Z(t) is estimated value at time t.

T(t) is trend term at time t, γ is trend coefficient.

I(t) is seasonal term, s = number of periods in seasonal cycles, δ is seasonality coefficient. $\frac{Y(t)}{Z(t)}$ capture seasonal effects.

The trend (T) and seasonal (S) components of ETS model is shown as below:

Trend Component (T)	Seasonal component (S)		
	None (N)	Additive (A)	Multiplicative (M)
None (N)	N, N	N, A	N, M
Additive (A)	A, N	A, A	A, M
Multiplicative (M)	M, N	M, A	M, M

The combination of ETS models:

Additive Error (A)	A, N, N	A, N, A	A, N, M	A, A, N	A, A, A	A, A, M	A, M, N	A, M, A	A, M, M
Multiplicative Error (M)	M, N, N	M, N, A	M, N, M	M, A, N	M, A, A	M, A, M	M, M, N	M, M, A	M, M, M

Among 18 ETS models, the best model will be selected.

3.5 Evaluation measures:

The error will be calculated as the misfit between observed and estimated values. Then the error will be normalized by dividing by the range of observed values. The normalized error is squared, and mean of squared error is calculated. Finally the squared root of mean square error is calculated as NRMSE.

Mathematically,

$$\left.\begin{aligned} & error = Observed\ value - Estimated\ value \\ & Normalized\ Error = \frac{error}{\max(Observed\ value) - \min(Observed\ value)} \\ & Root\ Mean\ Square\ Error = \sqrt{\frac{1}{N}\sum_{1=1}^{N}(Normalized\ Error)^{\wedge}2} \end{aligned}\right\}\ \text{Equation 12}$$

Average error (percentage) is calculated as,

$$Average\ error = \frac{RMSE}{Mean\ value} * 100\% \longrightarrow \text{Equation 13}$$

The mean absolute error,

$$MAE = \frac{1}{N}\sum_{t=1}^{n}|error| \longrightarrow \text{Equation 14}$$

The Mean Absolute Percentage error,

$$MAPE = \frac{100\%}{N}\sum_{t=1}^{n}\left|\frac{error}{observed\ value}\right| \longrightarrow \text{Equation 15}$$

Where, N is number of time points.

The Mean Squared Error (also known as Mean Squared Deviation),

$$MSE = \tfrac{1}{N}\textstyle\sum_{t=1}^{n}(error)^{\wedge}2 \longrightarrow \text{Equation 16}$$

The model which has the lowest value of RMSE, MAE, MSE and MAPE is the best model.

3.6 Dataset:

- COVID-19 data: COVID-19 data has been collected from dashboard by the John Hopkins University in the USA, the World Health Organization (WHO) and Health Emergency Operation Center (HEOC) Nepal. The data consists of number of daily positive, recovered and death case.
- Data from 1 May 2020 to 8 June 2021 (404 days) is taken for analysis.
- The dataset consists of data of number of Confirmed case (cumulative number of all positive cases), number of Recovered case (cumulative number of recovered cases) and number of Deaths case (cumulative number of deaths cases)
- The number of Active case is determined by removing the number of recovered case and deaths case. The number of active case is the total number new positive cases reported within 24 hours plus number of those who are not recovered yet or died out of coronavirus after infected.
- Out of 1212 data (404 for each case), 1122 data (374 for each case) are used for training the model, while 90 (30 for each case) are used to validate the model.

Table 3.1: Sample statistics of COVID-19 outbreak in Nepal

Case Date	Active	Recovered	Deaths
1 May 2020	43	16	0
1 Jan 2021	6048	253107	1864
8 Jun 2021	82736	504530	8098

CHAPTER 4: RESULTS AND DISCUSSION

Ensemble of different model in prediction is quite common to find to reduce the generalization error (difference between the error of the training data and the one of the test data). In the predictive modelling, reducing the error between observed data and estimated data has always been an issue. Using the combination of two models has found to be one of the approach to reduce the prediction error [Table 2.3].

In the prediction of evolution of epidemiology, the compartment [SEIR] model has widely been used since years [Table 2.1]. It is obvious that the spread of epidemic is dependent on the nature of the virus (infection rate, latent time, etc.), also it is a function of time component, and hence, the use of time series model in the prediction of epidemic is not novel [Table 2.2].

Maher Ala'raj, et al used SEIRD model to simulate COVID-19 outbreak in US and the prediction error is reduced by using ARIMA time series model [28]. This model consists of two parts: the modified SEIRD model and ARIMA models. The model fit SEIRD model parameters against historical values of infected, recovered and deceased population. Residuals of the first model for infected, recovered, and deceased populations are then corrected using ARIMA models. However, the hybrid model wouldn't handle the seasonality factor present in the data. To incorporate the seasonality factor, another time series model, the Error-Trend-Seasonality (ETS) model is added in between generalized SEIR and ARIMA model.

The hybrid of generalized SEIR, ETS and ARIMA model significantly reduce the error between observed data and its estimation as compared to SEIRD-ARIMA model, and hence improve the future prediction.

4.1 Results of SEIR model:

The optimized values of model parameter α, β, ɾ, δ, λ and κ for three time period has been calculated. The optimized values of model parameter α, β, ɾ, δ, λ and κ for no lockdown period are found to be 0.0185, 4.9992, 0.0177, 0.4297, [0.0851, 0.0062, 49.9999] and [3.7130*10^{-4}, 0.2238, 70.1418] respectively. The optimized values of model parameter α, β, ɾ, δ, λ and κ for lockdown period are found to be 0.2646, 1.2262, 0.6166, 0.0302, [0.1151, 0.0343, 5.3275] and [0.0050, 0.0101, 1.8263*10^{2}] respectively. The optimized values of model parameter α, β, ɾ, δ, λ and κ for partial lockdown period are found to be 0.0689, 1.1895, 0.1888, 0.0779, [0.2346, 0.0294, 87.0693] and [0.0031, 0.0075, 1.2882*10^2] respectively. The optimized parameters and initial condition of generalized SEIR model is shown in table below.

Table 4.1: Optimized values of parameters of generalized SEIR model

Parameter	Initial values	Optimized values (Lockdown) [2020/05/01 to 2020/09/31]	Optimized values (Partial Lockdown) [2021/04/01 to 2021/06/08]	Optimized values (No Lockdown) [2020/10/01 to 2021/03/31]
α	0.06	0.0185	0.0689	0.2646
β	1.0	4.9992	1.1895	1.2262
r	0.2	0.0177	0.1888	0.6166
δ	0.1	0.4297	0.0779	0.0302
λ	[0.01, 0.001, 10]	[0.0851, 0.0062, 49.9999]	[0.2346, 0.0294, 87.0693]	[0.1151, 0.0343, 5.3275]
κ	[0.001, 0.001, 10]	[3.7130*10^{-4}, 0.2238, 70.1418]	[0.0031, 0.0075, 1.2882*10^2]	[0.0050, 0.0101, 1.8263*10^2]

The model parameters (optimized values) are used to fit the generalized SEIR model described by equation 1. Figure 4.1(a), figure 4.2(a) and figure 4.3(a) shows the prediction by using generalized SEIR model for active, recovered and deaths cases respectively. Figure 4.1(b), 4.5(b) and 4.6(b) shows the estimation error of generalized SEIR model for each cases.

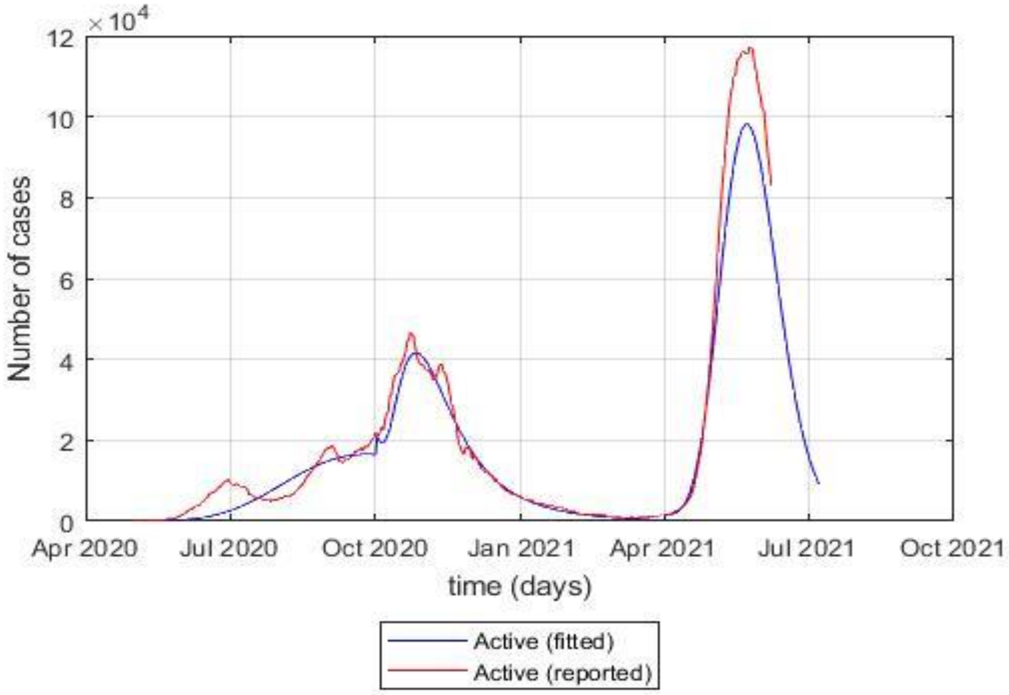


Figure 4.1(a) Prediction of active case using Generalized SEIR model

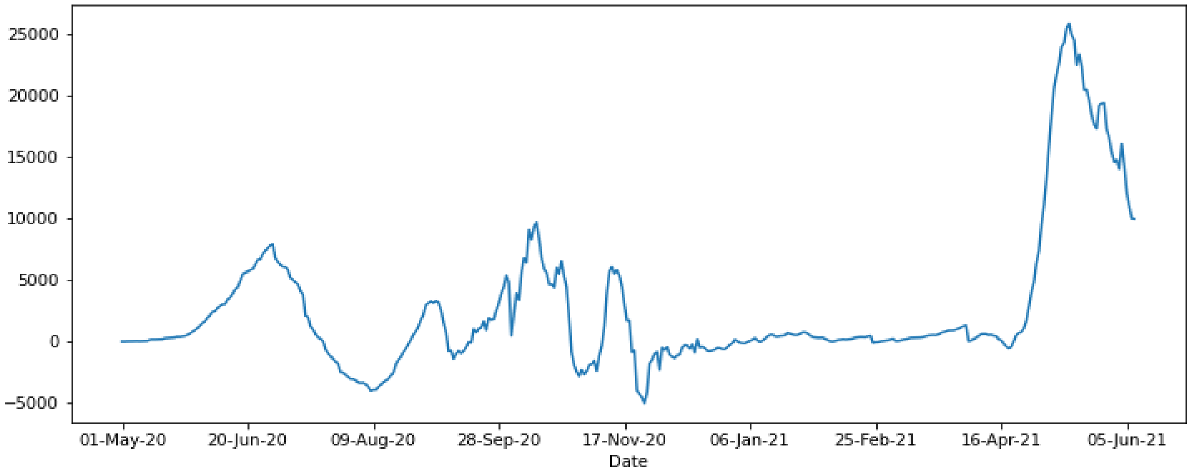


Figure 4.1(b) Estimation error of generalized SEIR model (Active case)

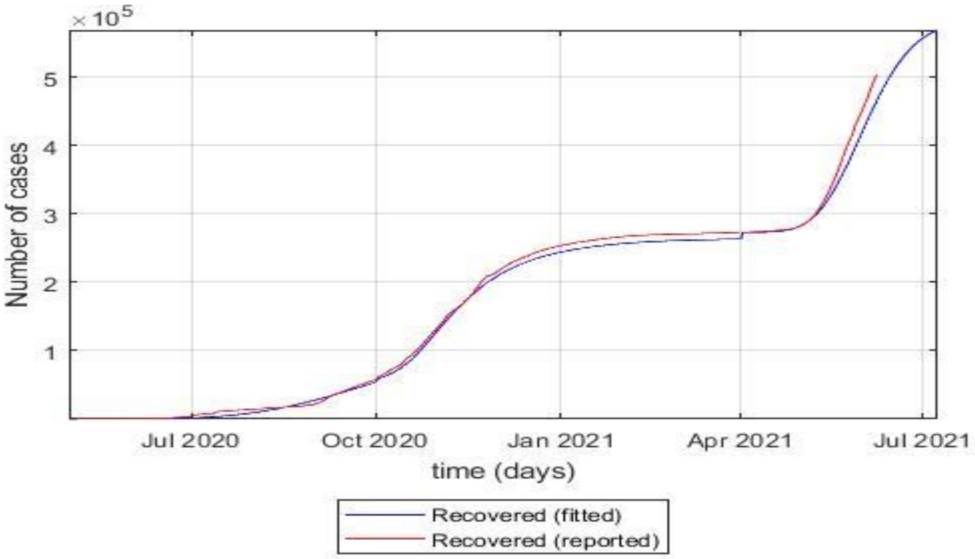


Figure 4.2(a) Prediction of recovered case using Generalized SEIR model

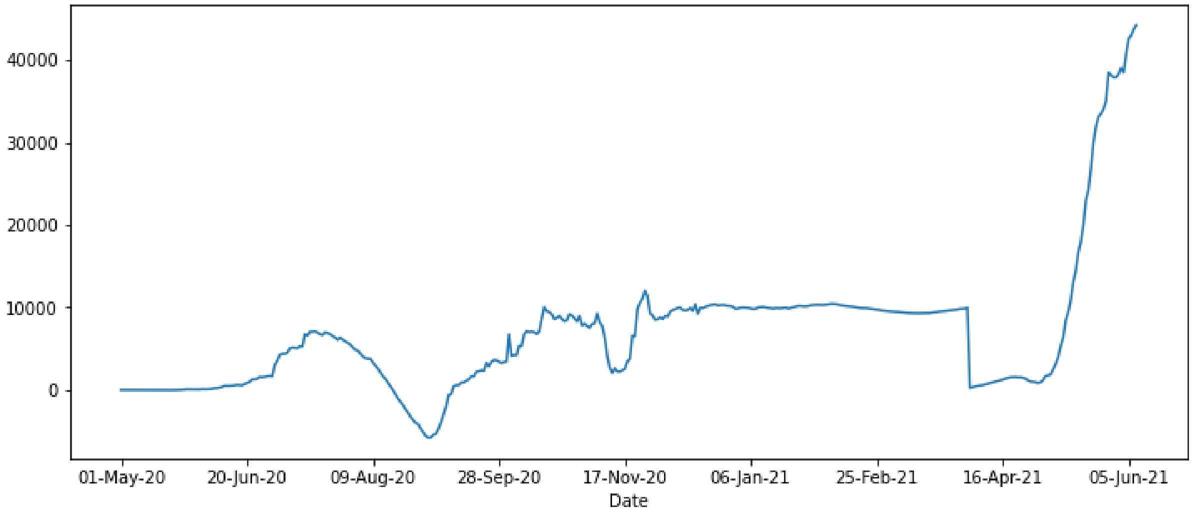


Figure 4.2(b) Estimation error of generalized SEIR model (Recovered case)

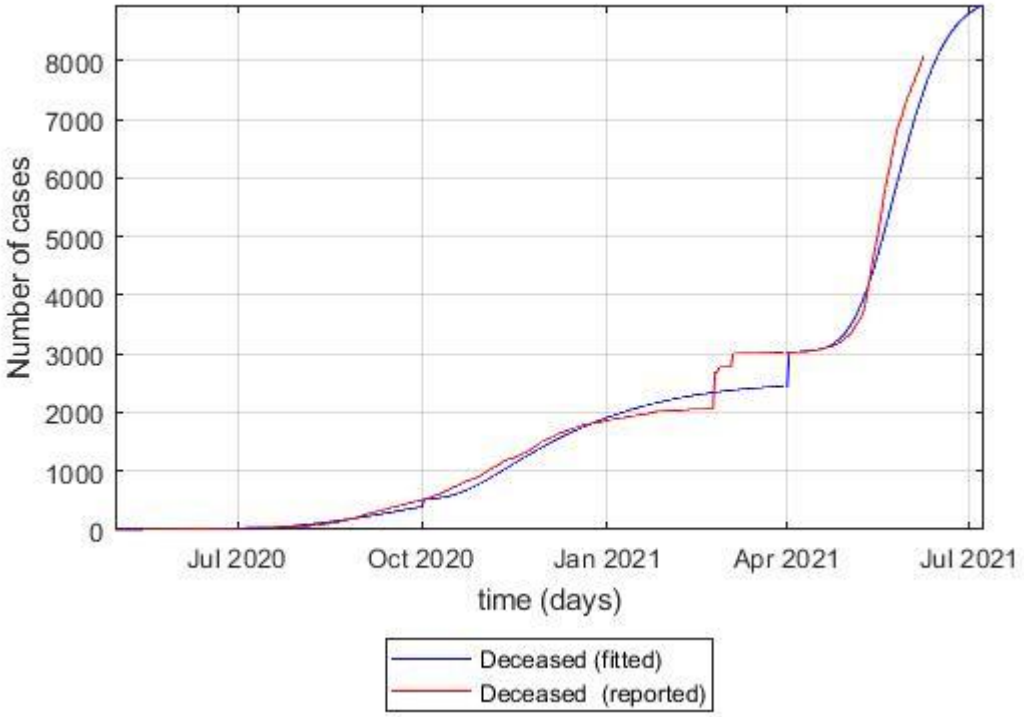


Figure 4.3(a) Prediction of deaths case using Generalized SEIR model

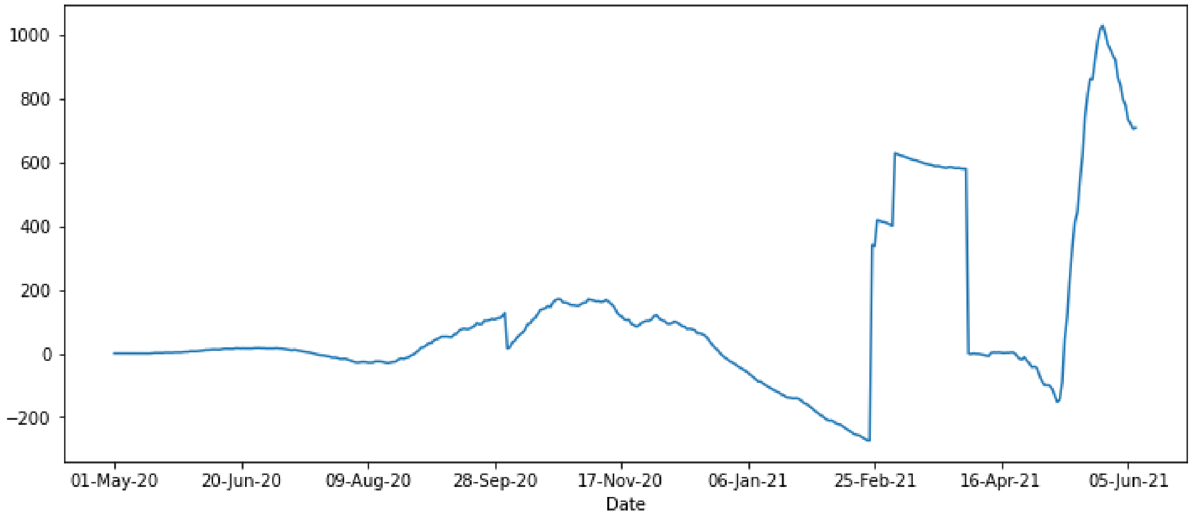


Figure 4.3(b) Estimation error of generalized SEIR model (Deaths case)

The estimation error of generalized SEIR model is then used as input to the ETS model. Table 4.2 shows the appropriate ETS model for each cases. Seasonal period for active, recovered and deaths cases are found to be 30, 50 and 15 days respectively. Appropriate ETS model for active, recovered and deaths cases are found to be (M,M,M), (M,M,M) and (M,A,M) respectively, where M stands for multiplicative and A for Additive model.

Table 4.2 Appropriate ETS models with seasonal period

Case	Appropriate ETS model	Seasonal period (days)
Active	M,M,M	30
Recovered	M,M,M	50
Deaths	M,A,M	15

The estimation error of SEIR-ETS model is then used as input to the ARIMA model. The appropriate ARIMA model is selected by observing ACF and PACF plot. Appropriate ARIMA model for each cases is shown in table 4.3.

Table 4.3: ARIMA model for each cases

Cases	Appropriate ARIMA model	p-value
Active	(2,0,0)	0.03531017
Recovered	(3,1,2)	$3.946795*10^{-11}$
Deaths	-	$2.172609*10^{-5}$

4.2 Prediction using SEIR-ETS-ARIMA model:

The prediction after using generalized SEIR-ETS-ARIMA model for each case is shown in figure 4.10, figure 4.11 and figure 4.12. The prediction of active case shows that there is decrease in cases for next 15 days (9 June, 2020 to 23 June, 2020) and slight increase in case after that for 7 days (24 June, 2020 to 30 June, 2020) and then increasing trend then after. The prediction for recovered and deaths cases shows increasing graph. Recovered cases would increase from 520 thousands to 600 thousands plus in next month. The deaths case would increase from 8 thousands to nearly 10 thousands in next month (8 June, 2020 to 8 July, 2020).

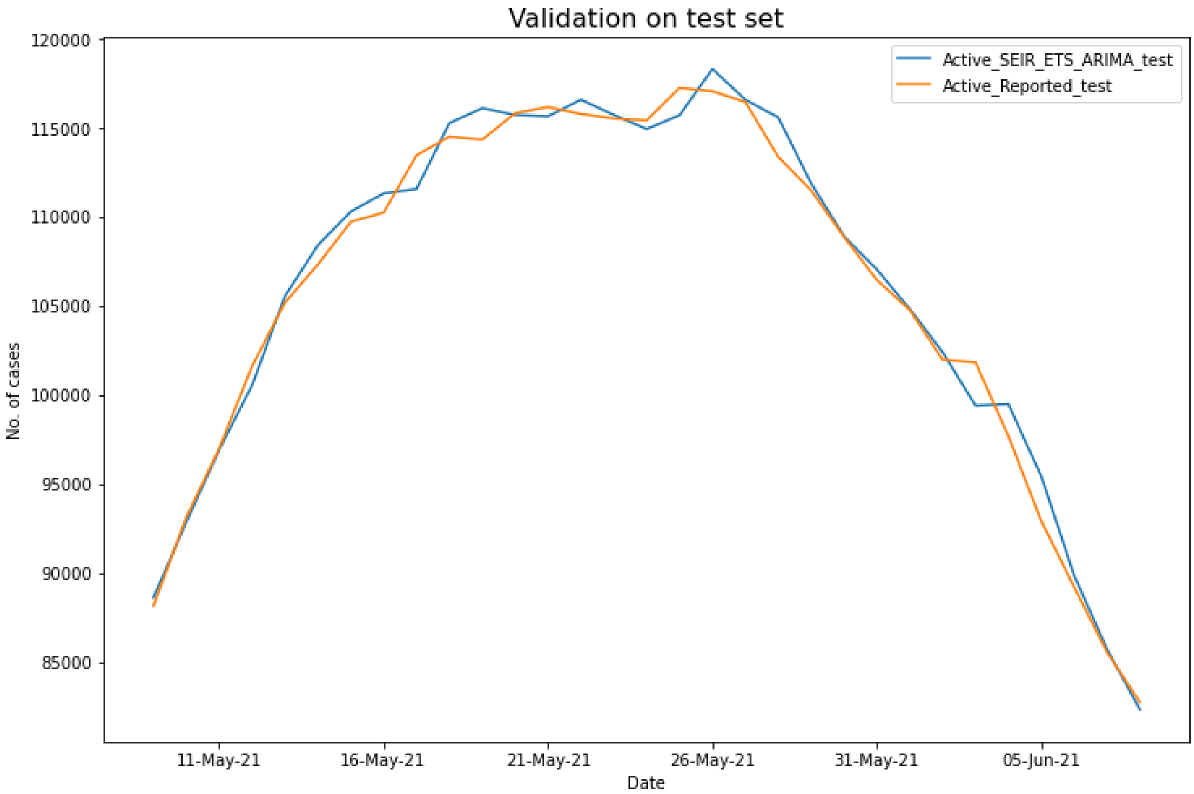


Figure 4.4 (a): Generalized SEIR-ETS-ARIMA model prediction for active case

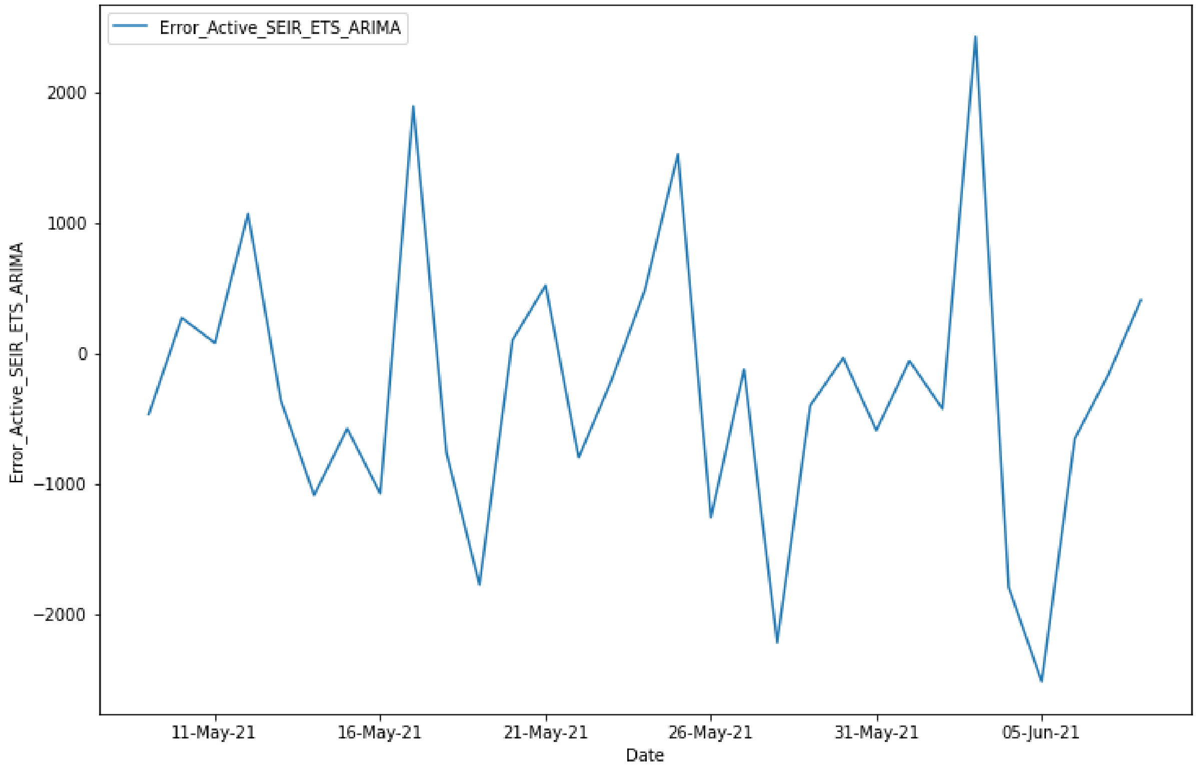


Figure 4.4 (b): Error of Generalized SEIR-ETS-ARIMA model prediction for active case

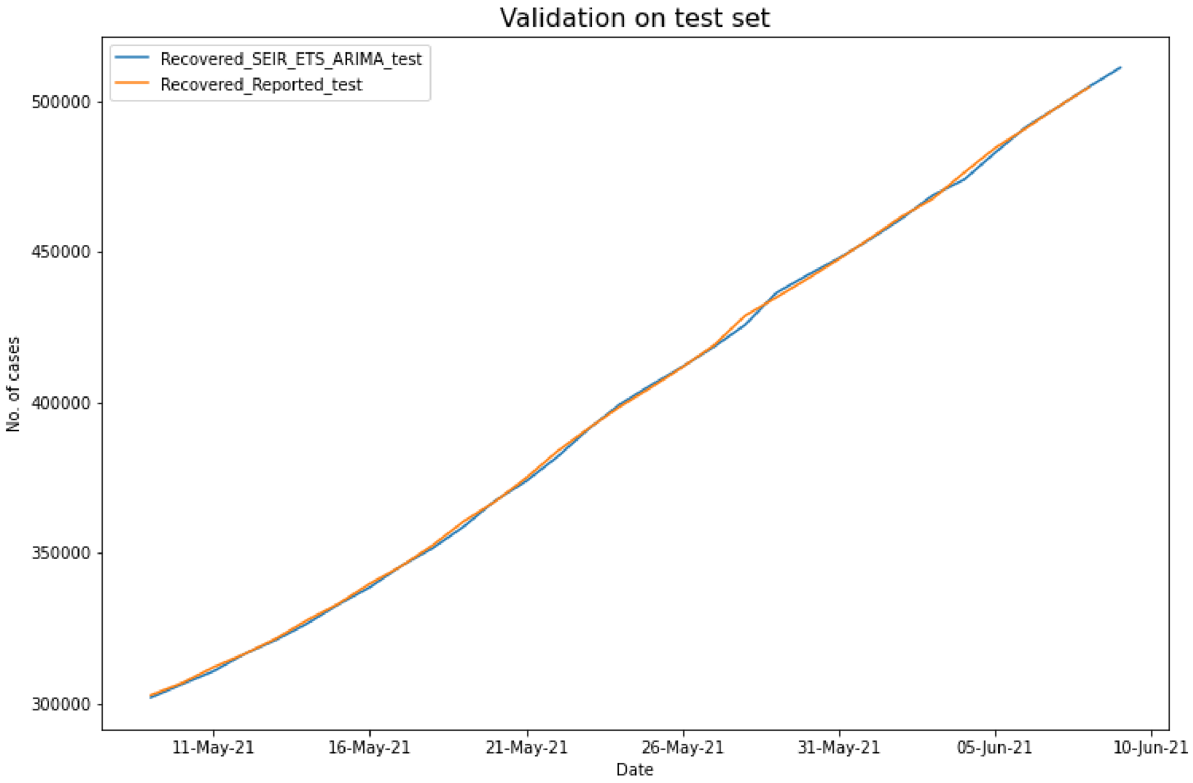


Figure 4.5(a): Generalized SEIR-ETS-ARIMA model prediction for recovered case

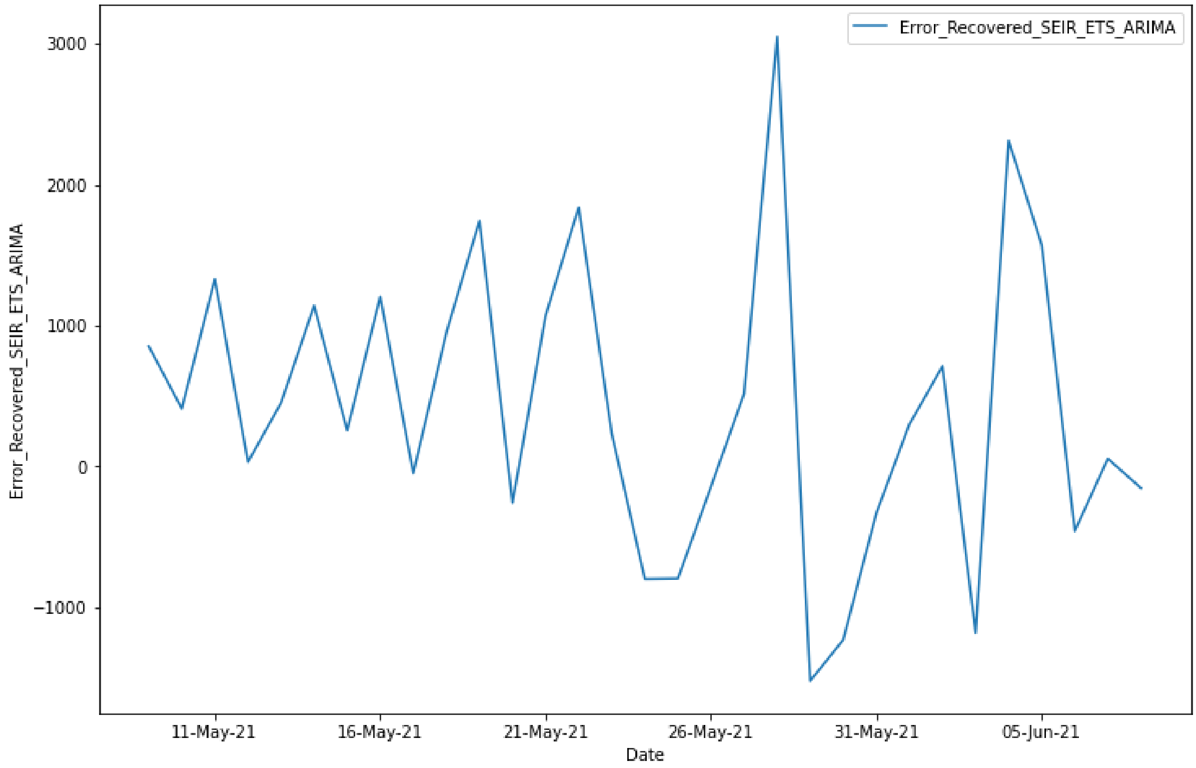


Figure 4.5(b): Error of Generalized SEIR-ETS-ARIMA model prediction for recovered case

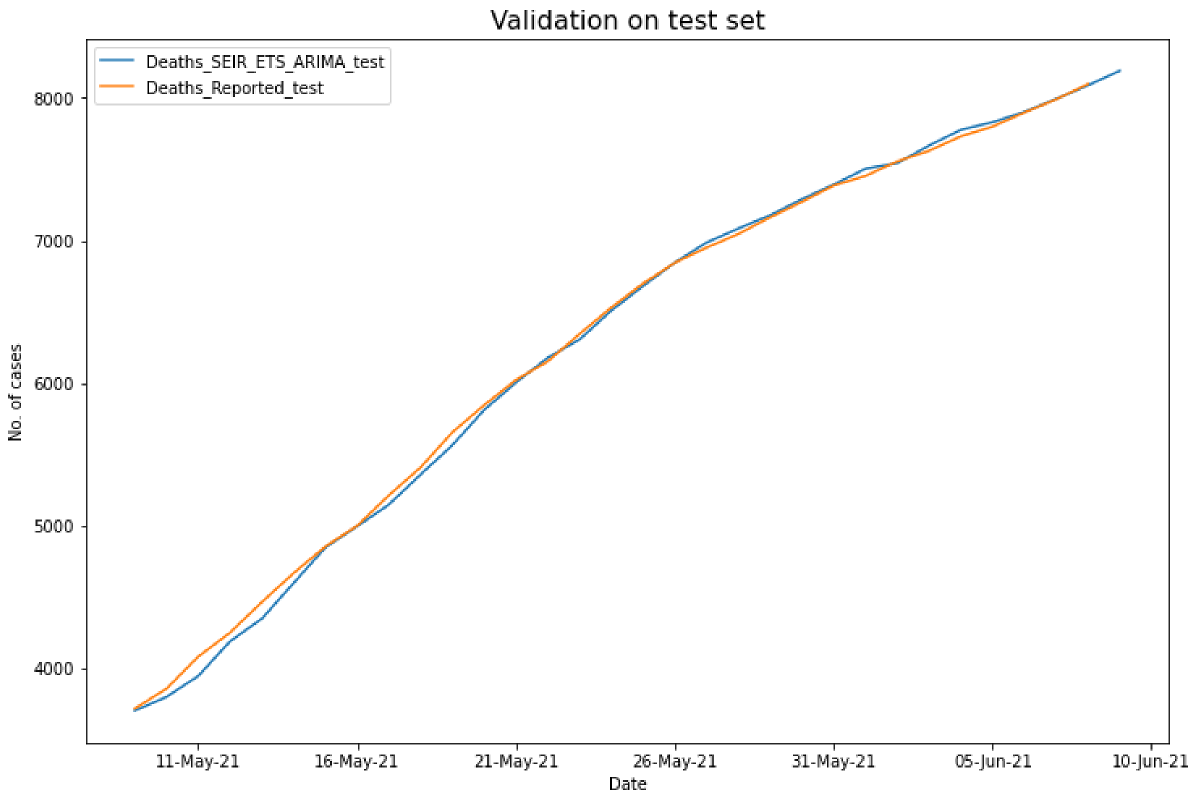


Figure 4.6(a): Generalized SEIR-ETS-ARIMA model prediction for deaths case

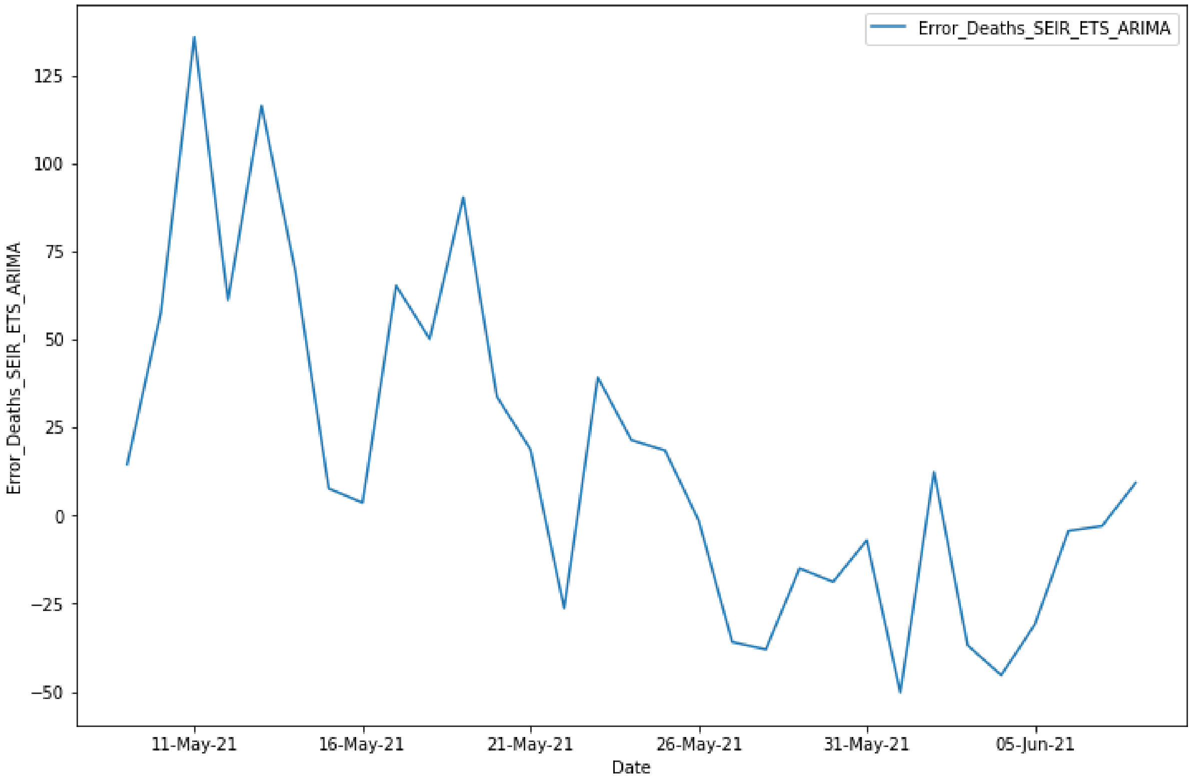


Figure 4.6(b): Error of Generalized SEIR-ETS-ARIMA model prediction for deaths case

4.3 Improvement compared to SEIR model:

Table 4.4, Table 4.5, Table 4.6 and Table 4.7 shows various quality measures for using Generalized SEIR model and on using SEIR-ETS-ARIMA model. We can see that SEIR-ETS-ARIMA model improves in all quality measures than in generalized SEIR model.

Table 4.4: MAE and Normalized MAE on using Generalized SEIR model versus using SEIR-ETS-ARIMA model

Cases	Active	Improvement %	Recovered	Improvement %	Deaths	Improvement %
MAE (SEIR-ETS-ARIMA)	$2.811*10^{-5}$	75.39%	$1.41*10^{-5}$	94.21%	$5.51*10^{-7}$	89.71%
MAE (Generalized SEIR)	0.00011		0.000243		$5.35*10^{-6}$	
Normalized MAE (SEIR-ETS-ARIMA)	0.024425	16.45%	0.004312	70.15%	0.00838	57.73%
Normalized MAE (Generalized SEIR)	0.029236		0.014441		0.01984	

Table 4.5: MSE and Normalized MSE on using Generalized SEIR model versus using SEIR-ETS-ARIMA model

Cases	Active	Improvement %	Recovered	Improvement %	Deaths	Improvement %
MSE (SEIR-ETS-ARIMA)	$4.87*10^{-10}$	98.92%	$6.14*10^{-10}$	99.52%	$2.29*10^{-12}$	97.41%
MSE (Generalized SEIR)	$4.49*10^{-08}$		$1.2876*10^{-7}$		$8.83*10^{-11}$	
Normalized MSE (SEIR-ETS-ARIMA)	0.00104	64.66%	$13.103*10^{-5}$	93.18%	0.000125	89.71%
Normalized MSE (Generalized SEIR)	0.002939		0.000455		0.0012115	

Table 4.6: RMSE and Normalized RMSE on using Generalized SEIR model versus using SEIR-ETS-ARIMA model

Cases	Active	Improvement %	Recovered	Improvement %	Deaths	Improvement %
RMSE (SEIR-ETS-ARIMA)	1112.77	82.49%	1123.772	89.56%	48.87397	82.66%
RMSE (Generalized SEIR)	6354.897		10764.8003		281.8702	
Normalized RMSE (SEIR-ETS-ARIMA)	0.03223	40.55%	0.00557	73.89%	0.011164	67.93%
Normalized RMSE (Generalized SEIR)	0.054214		0.021337		0.03481	

Table 4.7: MAPE and Normalized MAPE on using Generalized SEIR model versus using SEIR-ETS-ARIMA model

Cases	Active	Improvement %	Recovered	Improvement %	Deaths	Improvement %
MAPE (SEIR-ETS-ARIMA)	0.798 %	96.96%	0.22 %	98.49%	0.691 %	96.66%
MAPE (Generalized SEIR)	26.221%		14.573%		20.72%	

The MAE for active, recovered and deaths case are 2.81*10^{-5}, 1.41*10^{-5} and 5.51*10^{-7} respectively. The Normalized MAE for active, recovered and deaths cases are 0.024425, 0.004312 and 0.00838 respectively. The MSE for active, recovered and deaths case are 4.87*10^{-10}, 6.14*10^{-10} and 2.29*10^{-12} respectively. The Normalized MSE for active, recovered and deaths cases are 0.00104, 13.103*10^{-5} and 0.000125 respectively. The Root Mean Square Error (RMSE) for active, recovered and deaths cases are 1112.77, 1123.772 and 48.87397 respectively. The Normalized RMSE for active, recovered and deaths cases are found to be 0.03223, 0.00557 and 0.011164 respectively. The Mean Absolute Percentage Error (MAPE) for active, recovered and deaths cases are 0.798 %, 0.22 % and 0.691 % respectively. Value of every quality measures for each cases has been reduced by new model (SEIR-ETS-ARIMA) as compared to generalized SEIR model.

The greatest reduction in MAE with respect to generalized model is achieved for recovered case (94.21% each), followed by deaths case (89.71%) and active case (75.39%). Normalized MAE has been improved by 16.45%, 70.15% and 57.73% for active, recovered and deaths cases using SEIR-ETS-ARIMA model than using generalized SEIR model. The MSE has been improved by around 98% each cases using SEIR-ETS-ARIMA model. Similarly, normalized MSE has been improved by 64.66%, 93.18% and 89.71% for active, recovered and deaths cases using SEIR-ETS-ARIMA model. RMSE has been improved by 82.49%, 89.56% and 82.66% for active, recovered and deaths cases using SEIR-ETS-ARIMA model than using generalized SEIR model. Normalized RMSE has been improved by 40.55%, 73.89% and 67.93% for active, recovered and deaths cases using SEIR-ETS-ARIMA model than using generalized SEIR model. The MAPE has been improved by around 97% for each cases using SEIR-ETS-ARIMA model than using generalized SEIR model. We can observe that the improvement in quality is less for active case as compared to other two cases.

4.4 Comparison with baseline paper:

4.4.1 Prediction on test set of baseline paper:

We have used the same data as used in the reference paper [28]. The dataset includes the daily number of cases of active, recovered and deaths incidence of United State from January 30, 2020 to September 16, 2020. Figure 4.7(a), 4.8(a) and 4.9(a) shows prediction of active, recovered and deaths case on test set of dataset of United State. Figure 4.7(b), 4.8(b) and 4.9(b) shows estimation error. Error for recovered and deaths cases largely reside around zero value. Also, for active case, the error is mostly around zero to few thousands value.

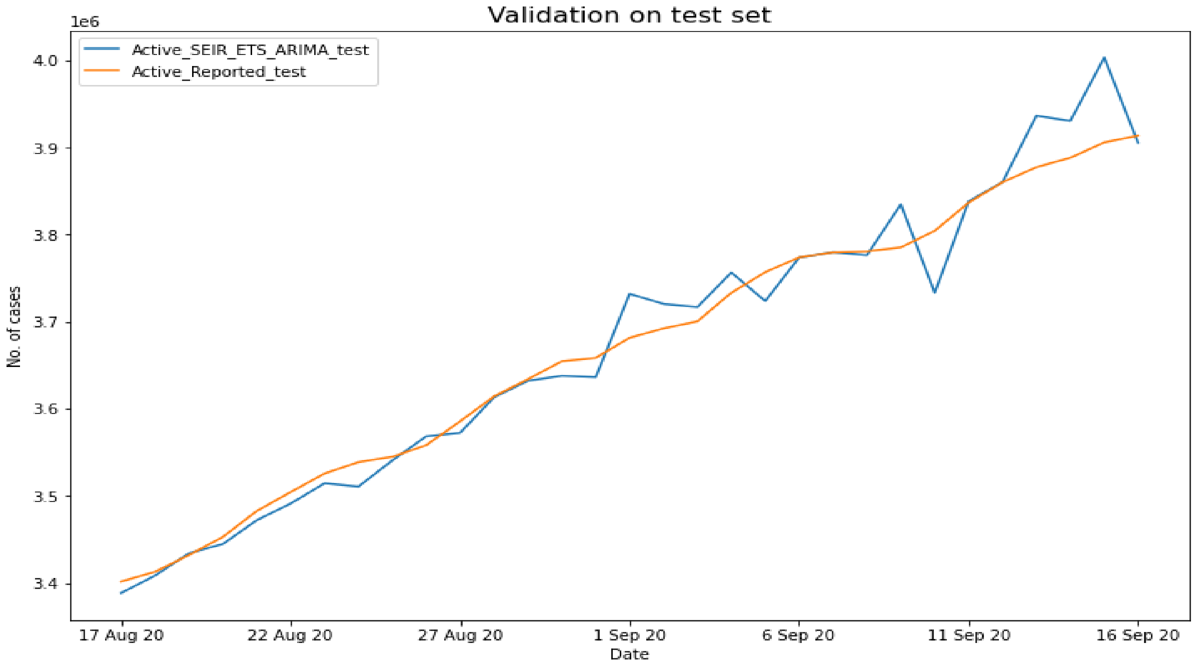


Figure 4.7(a) Active case prediction on test set

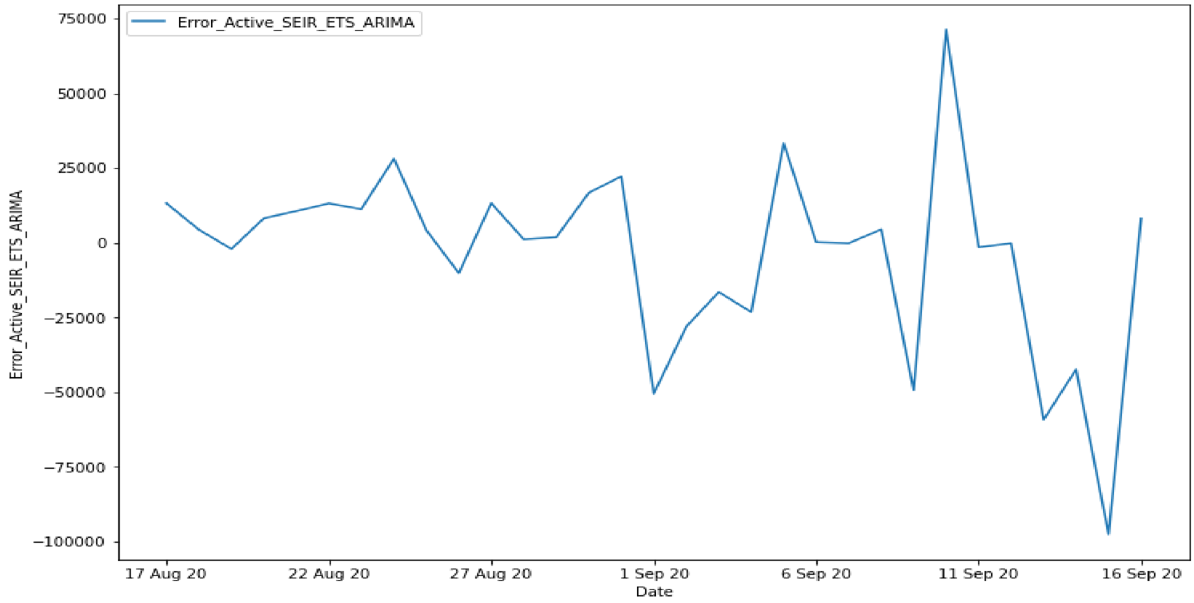


Figure 4.7(b) Error graph for active case (validation set)

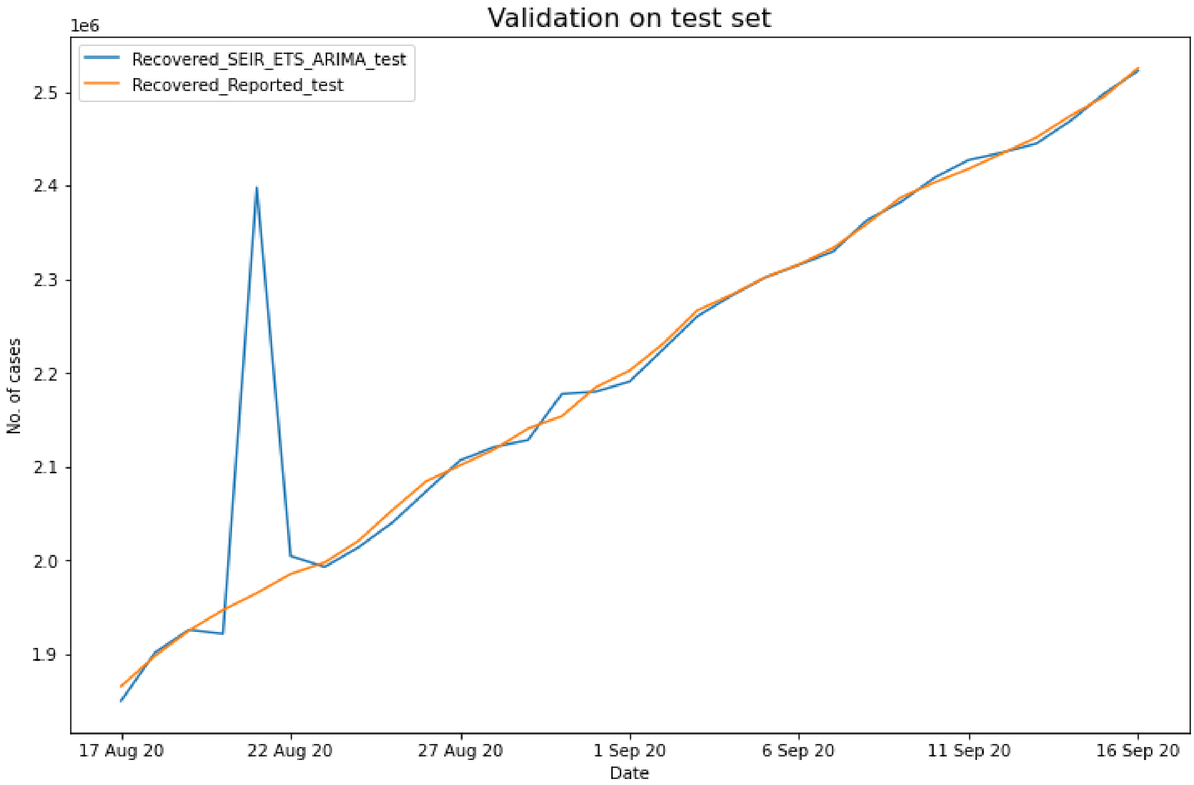


Figure 4.8(a) Recovered case prediction on test set

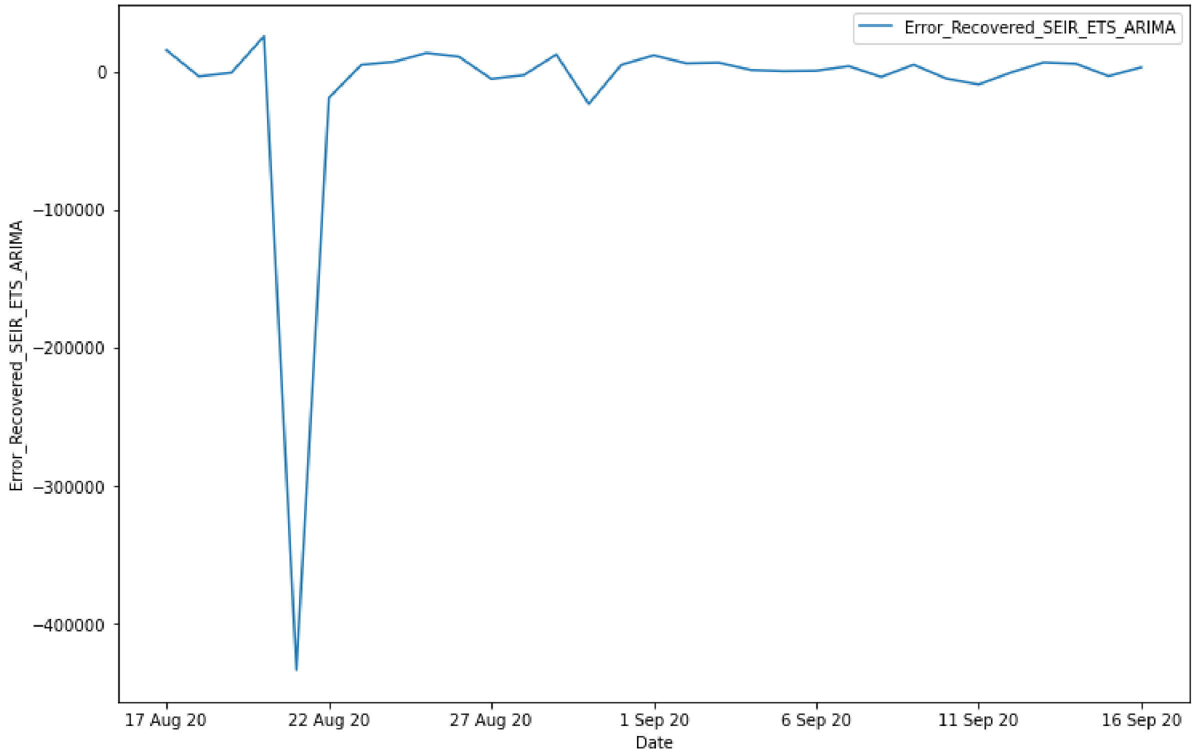


Figure 4.8(b) Error graph for recovered case (validation set)

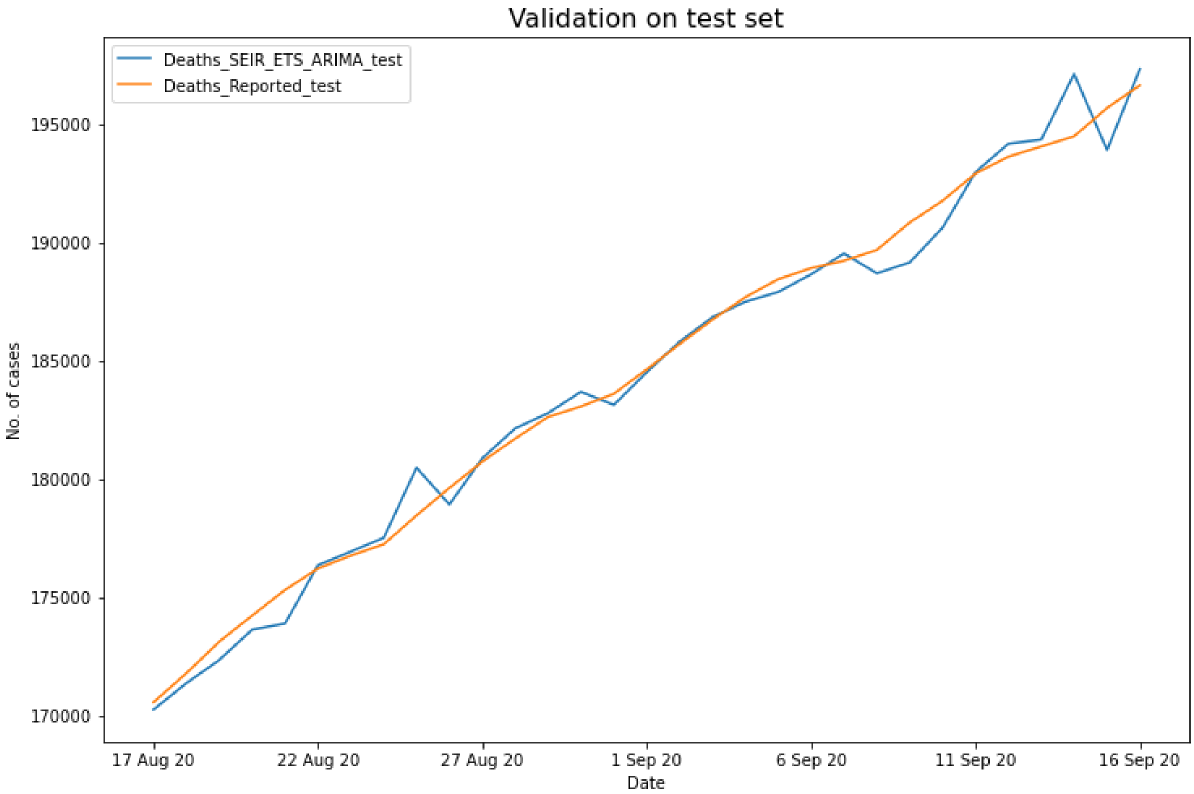


Figure 4.9(a) Deaths case prediction on test set

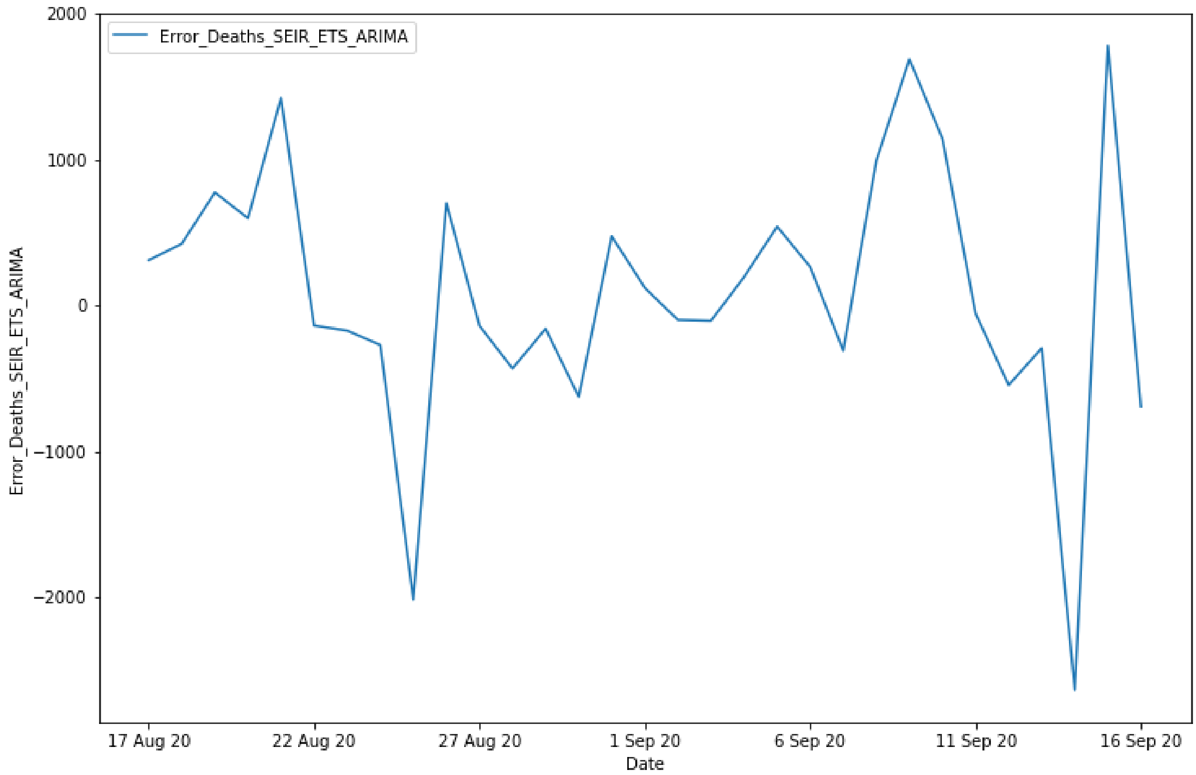


Figure 4.9(b) Error graph for deaths case (validation set)

4.4.2 MAE, MSE and MSLE compared to reference paper:

The MAE, MSE and MSLE has been calculated for the validation. The MAE, MSE and MSLE of SEIR-ETS-ARIMA model and SEIRD-ARIMA model for each case is shown in the table 4.8, 4.9 and 4.10. All the quality measures have been improved using SEIR-ETS-ARIMA model as compared to SEIRD-ARIMA model, except MSE of recovered case.

The MAE for active, recovered and deaths case are improved by 68.96%, 12.78% and 79.597% respectively. The MSE for active and deaths case are improved by 96.79% and 99.82% respectively. However, the MSE for recovered case using SEIR-ETS-ARIMA model is greater than of SEIRD-ARIMA model. The MSLE for active, recovered and deaths case are improved by 99.40%, 52.51% and 99.94% respectively. This shows that the new model largely improve the quality measures for active and deaths cases than the SEIRD-ARIMA model.

Table 4.8: The MAE of SEIR-ETS-ARIMA model and SEIRD-ARIMA model

Cases	**Active**	**Improvement %**	**Recovered**	**Improvement %**	**Deaths**	**Improvement %**
MAE (SEIR-ETS-ARIMA)	$6.3*10^{-5}$	68.96%	$7.38*10^{-5}$	12.78%	$1.96*10^{-6}$	79.597 %
MAE (SEIRD-ARIMA)	$2.03*10^{-4}$		$8.46*10^{-5}$		$8.07*10^{-6}$	

Table 4.9: The MSE of SEIR-ETS-ARIMA model and SEIRD-ARIMA model

Cases	**Active**	**Improvement %**	**Recovered**	**Improvement %**	**Deaths**	**Improvement %**
MSE (SEIR-ETS-ARIMA)	$8.89*10^{-9}$	96.79%	$6.68*10^{-8}$	-157.05 %	$7.49*10^{-12}$	99.82%
MSE (SEIRD-ARIMA)	$2.77*10^{-7}$		$2.64*10^{-8}$		$2.76*10^{-9}$	

Table 4.10: The MSLE of SEIR-ETS-ARIMA model and SEIRD-ARIMA model

Cases	**Active**	**Improvement %**	**Recovered**	**Improvement %**	**Deaths**	**Improvement %**
MSLE (SEIR-ETS-ARIMA)	$1.64*10^{-9}$	99.40%	$1.24*10^{-8}$	52.51%	$1.41*10^{-12}$	99.94%
MSLE (SEIRD -ARIMA)	$2.75*10^{-7}$		$2.62*10^{-8}$		$2.76*10^{-9}$	

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1 Conclusion

- Seasonal period for active, recovered and deaths cases are found to be 30, 50 and 15 days respectively. Appropriate ETS model for active, recovered and deaths cases are found to be (M,M,M), (M,M,M) and (M,A,M) respectively, where M stands for multiplicative and A for Additive model. The appropriate ARIMA model for active and recovered case are found to be (2,0,0) and (3,1,2) respectively, while, from ACF and PACF plot, no appropriate model for death case is determined.
- On comparing SEIR-ETS-ARIMA model with SEIRD-ARIMA model by calculating quality measures such as MAE, MSE and MSLE (Table 4.8, 4.9 and 4.10), it is found that the new model largely improve the quality measures for active and deaths cases than the SEIRD-ARIMA model. For recovered case, all quality measures are improved except the MSE.
- The error graph of active and recovered cases in case of Nepal shows error generally resides near zero and fluctuate to few thousands, and there are irregular spikes in the graph, which shows the model is not biased for active and recovered cases.
- The error of the model is the difference between original time series (reported) data and the prediction made by the SEIR-ETS-ARIMA model. The error graph of deaths cases shows that the error generally resides near zero, but it shows decreasing trend, hence, we cannot claim the un-biasness of the model for death case.

5.2 Recommendation

- Other machine learning models like LSTM, Genetic Algorithm, Neural Network, etc. may also be used in combination with epidemiological model, or the combination of time series, epidemiological and other model can be used instead of two time series model for precise prediction than using any single model.
- The generalized SEIR model can be further modified to incorporate the new variant case, effect of vaccination over time, and other aspect which may affect the spread of the virus.
- The generalized SEIR model can be modified to include natural birth and natural deaths.

APPENDIX (1): RESULT OF ETS MODEL

Seasonal decomposition of Residual_1 to select appropriate ETS model:

The error after applying generalized SEIR model is the input for ETS model. To select appropriate ETS model, decomposition is done for each of the cases.

The error (Residual_1) is the input data, which is then decomposed into three components: Seasonal component, Trend component and the error (Remainder) component. The irregularity and seasonal nature in any of the above mention components represent the multiplicative mode (M) in ETS model. However, if there is upward or downward trend line in any components, additive mode (A) is chosen. For example, in the decomposition of Residual_1 of active case, we see that the seasonal component clearly has seasonal nature, the error component is irregular in nature, and the trend component also has repeating nature after a period of time. Hence, the E, T and S in ETS model is in multiplicative mode (M). Therefore, ETS model for active case is (M, M, M).

The decomposition of Residual_1 for active, recovered and cases are shown in figure 4, figure 2 and figure 3 respectively.

The seasonal period is chosen by observing the seasonal component in the decomposition. For active case, it is clear that the cycle is repeating on around 30 days, for recovered case, the cycle is repeating on around 50 days and for deaths case it is around 15 days. Hence, the seasonal period for active, recovered and deaths case is 30, 50 and 15 days respectively.

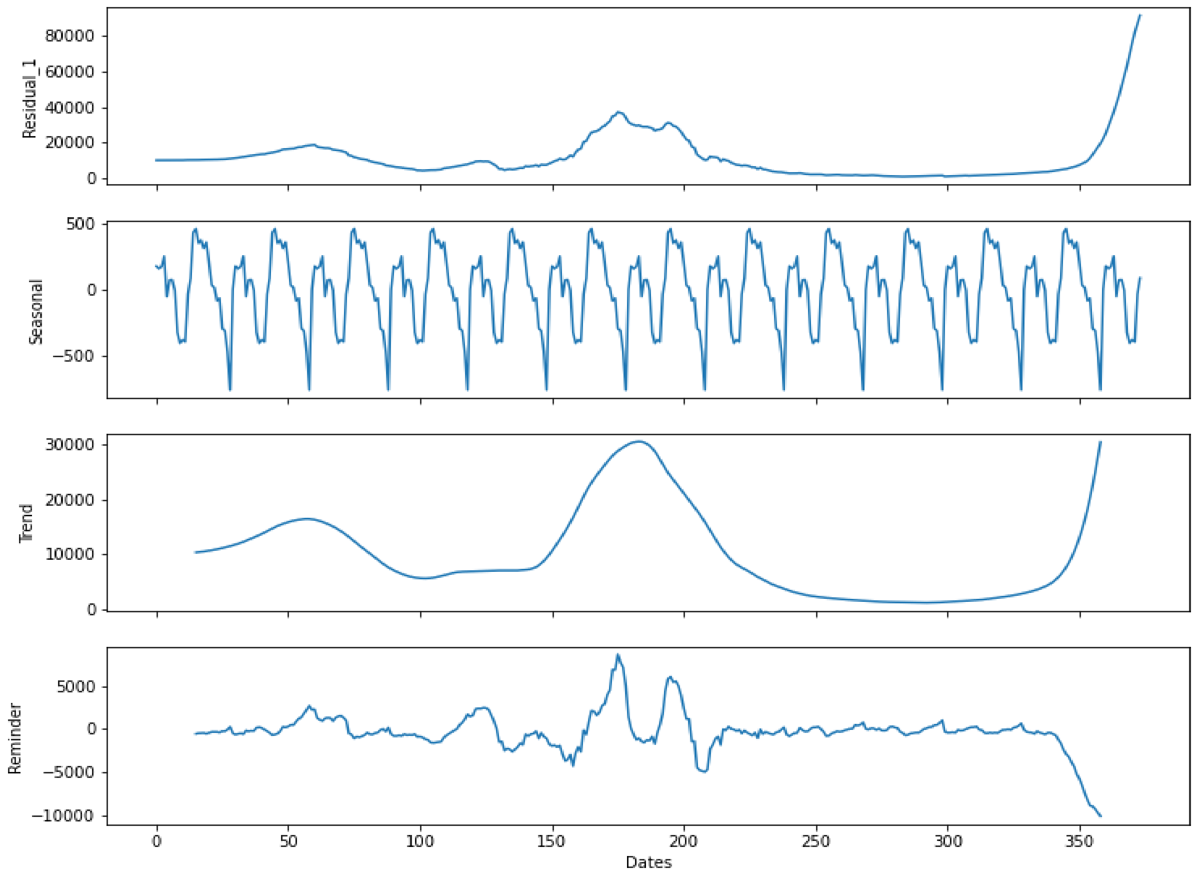


Figure 1: Seasonal decomposition for active case

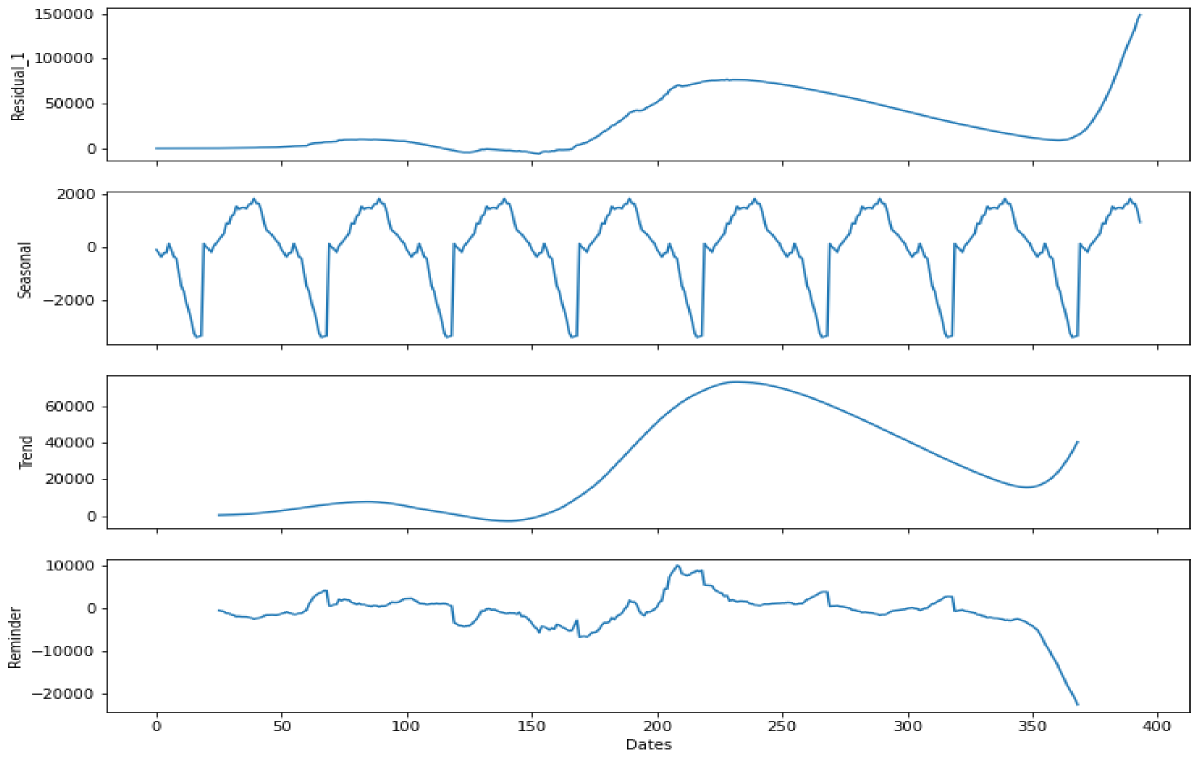


Figure 2: Seasonal decomposition for Recovered case

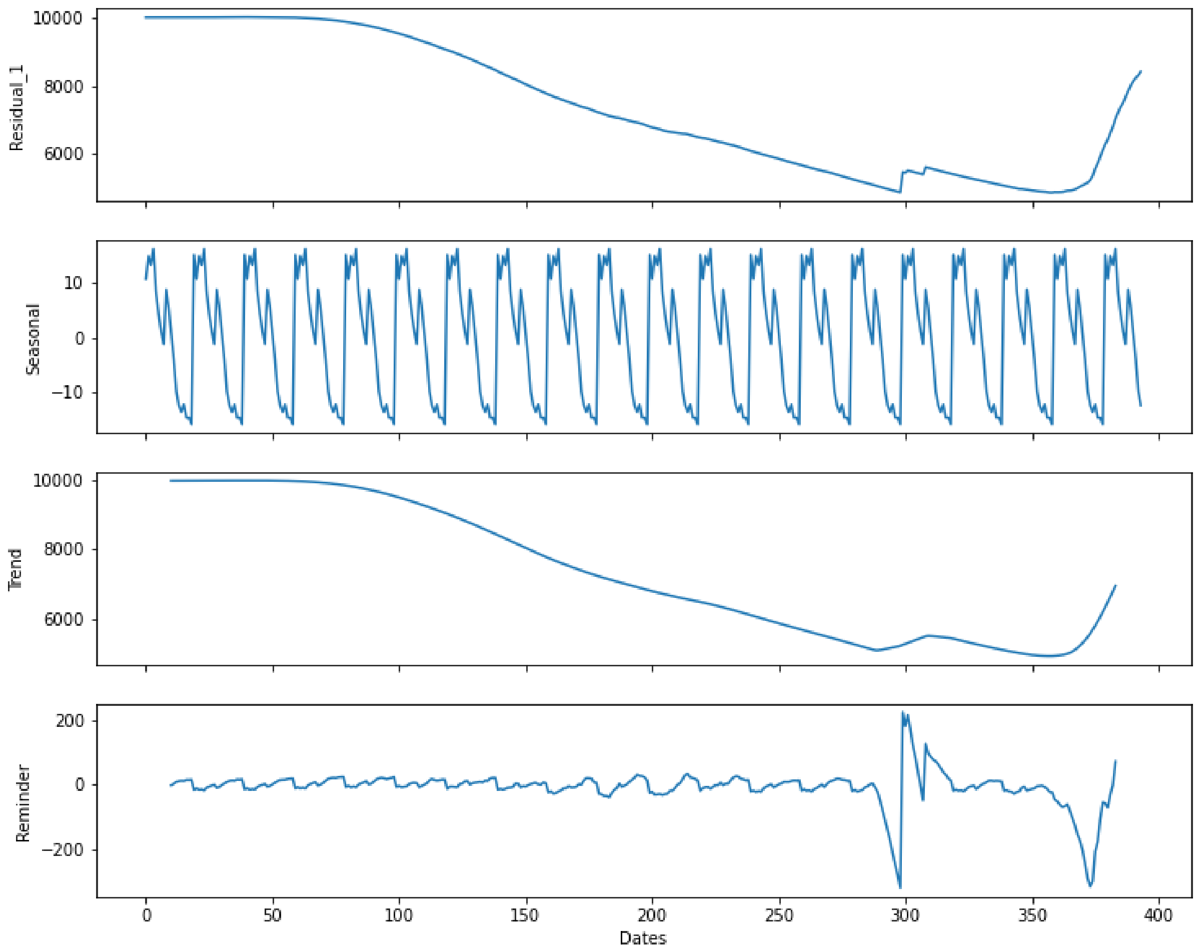


Figure 3: Seasonal decomposition for Deaths case

Prediction of SEIR-ETS model:

The prediction of each cases using generalized SEIR model and ETS model is shown as below:

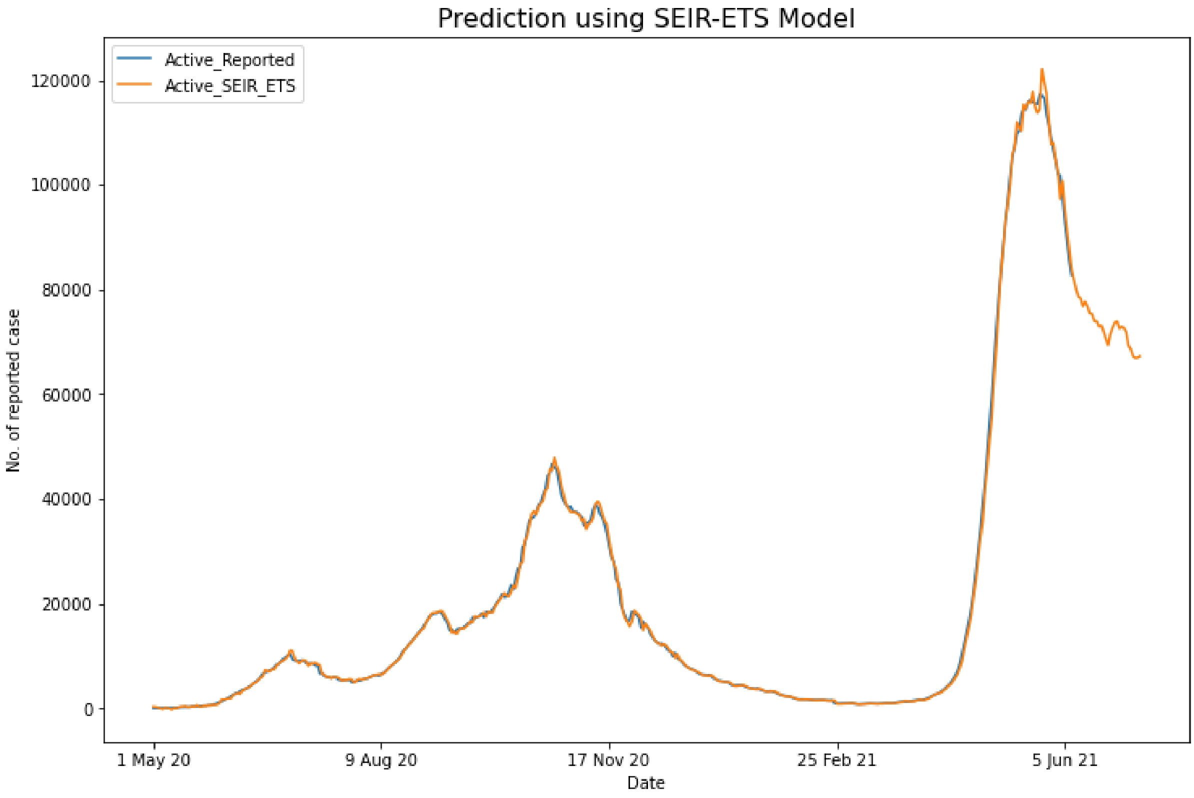


Figure 4: SEIR-ETS prediction for active cases

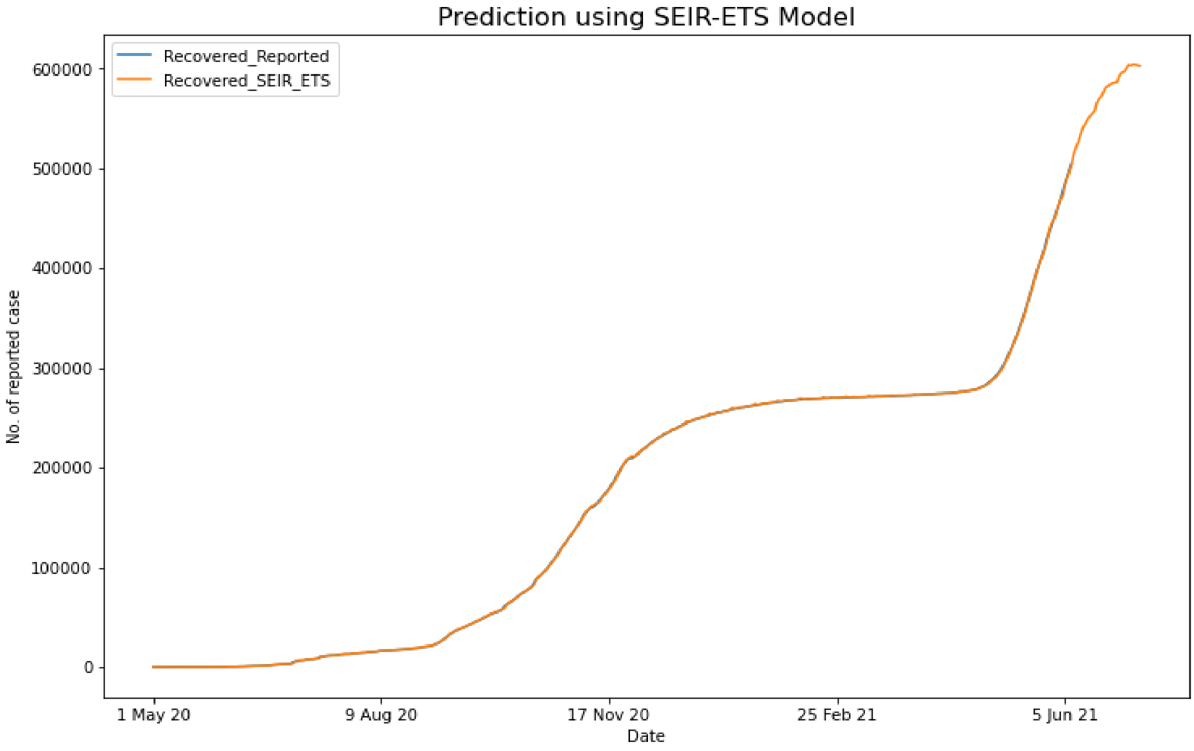


Figure 5: SEIR-ETS prediction for recovered cases

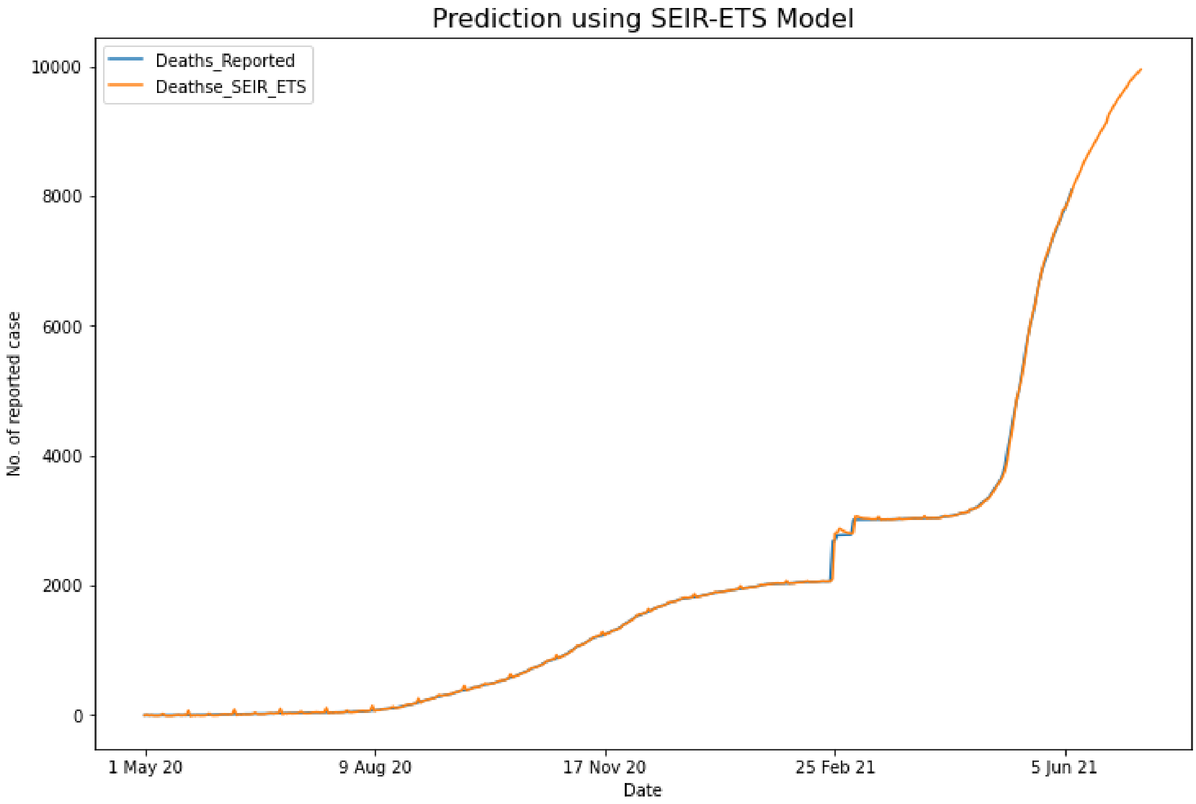


Figure 6: SEIR-ETS prediction for deaths cases

APPENDIX (2): RESULTS OF ARIMA MODEL

Selection of appropriate ARIMA (p,d,q) model using ACF and PACF plot:

The error of SEIR-ETS (Residual_2) is then fed to ARIMA model. The non-stationary data will be made stationary. The p – value of the data (Residual_2) is calculated using Akaike Information Criterion (AIC). If the p-value is less than 0.05 (the significance level), the data is said to be stationary. The p – value for Residual_2 of active and deaths cases are 0.03531017 and 2.172609*10^{-05} respectively. Hence, the order of integration (d) of the ARIMA model for active and deaths case is 0. The p-value for Residual_2 of recovered case is 0.3589365897 which is greater than 0.05. Hence, the differencing is required. After the first differencing, the p – value is 3.946795*10^{-11} which is less than 0.05. Therefore, the order of integration (d) of the ARIMA model for recovered case is 1.

The order of auto-regression (p) and moving average (q) are determined using ACF and PACF plot as shown in figure 7, 8 and 9. The plots is observed up to 50 lags.

Table 2: Nature of ACF and PACF plot for choosing appropriate model

Model	ACF $\rho(k)$ plot	PACF ϕ_{kk} plot
AR (p)	Damped exponential and/or sine functions	$\phi_{kk} = 0$ for $k > p$, where, ϕ_{kk} is value of partial auto correlation function at lag k
MA (q)	$\rho(k) = 0$ for $k > q$, where, $\rho(k)$ is value of autocorrelation function at lag k	Dominated by damped exponential and/or sine functions
ARMA (p, q)	Damped exponential and/or sine functions after lag max (0, q-p)	Dominated by damped exponential and/or sine functions after lag max (0, p-q)

The ACF and PACF of Residual_2 of active case is shown in figure 7. The value at lag 0 is the correlation of the present data with itself, so, it is not significant for the analysis, hence, is ignored. The ACF plot is exponentially decaying. The PACF plot has sharp cut-off, at lag = 3, the value of partial auto correlation is 0 and at lag > 3, the value of partial auto correlation is below the significant value. Hence, according to the table above, the AR model with p-value 2 is chosen. Therefore, ARIMA model for active case is (2,0,0).

The ACF and PACF of Residual_2 of recovered case after first differencing is shown in figure 8. The ACF plot is oscillating while the PACF plot is exponential. Hence, according to the table 2, the ARMA model is chosen. There are 2 and 3 spikes in ACF and PACF plot respectively. Hence, ARMA model of order (3,2) is chosen. Therefore, ARIMA model for recovered case is (3,1,2).

The ACF and PACF of Residual_2 of deaths case is shown in figure 9. Both of the ACF and PACF plot shows no significant correlation between lags. Also, the value of d is 0. Therefore, ARIMA model is not applied in the Residual_2 of deaths case.

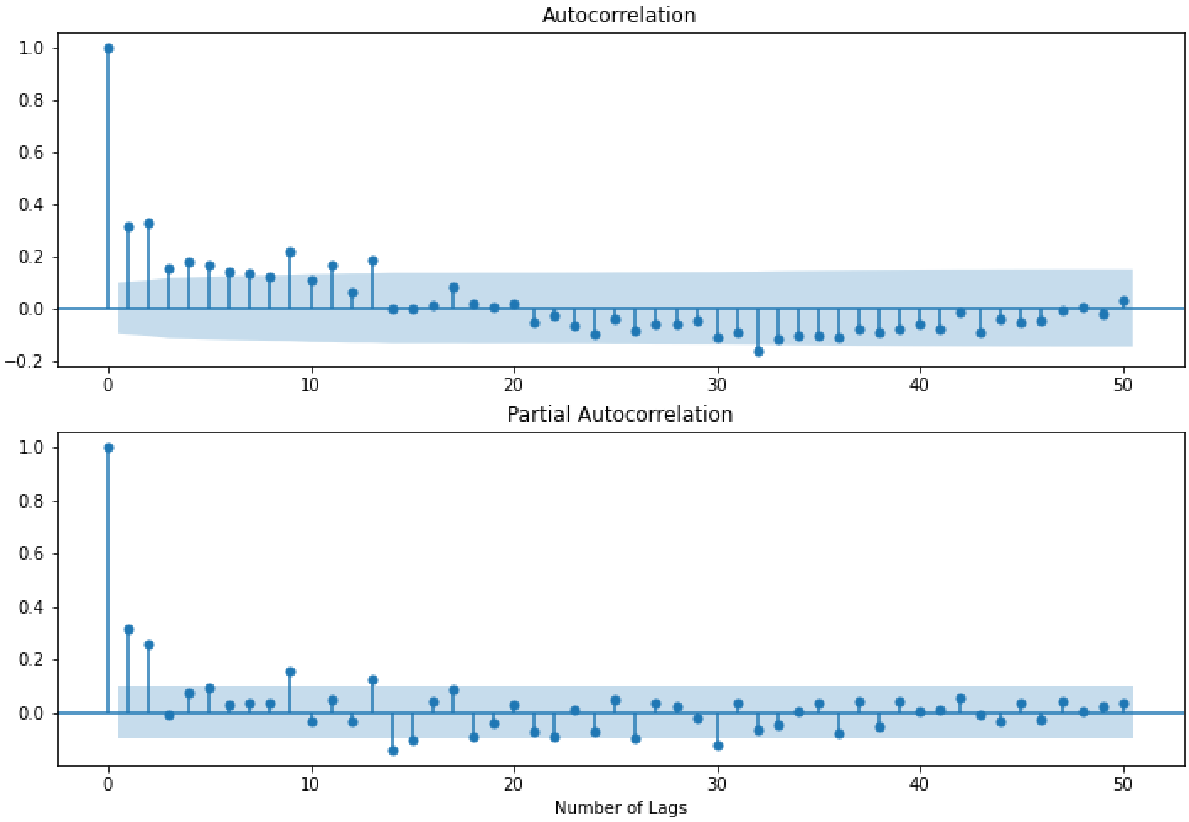


Figure 7: The ACF and PACF plot for active case

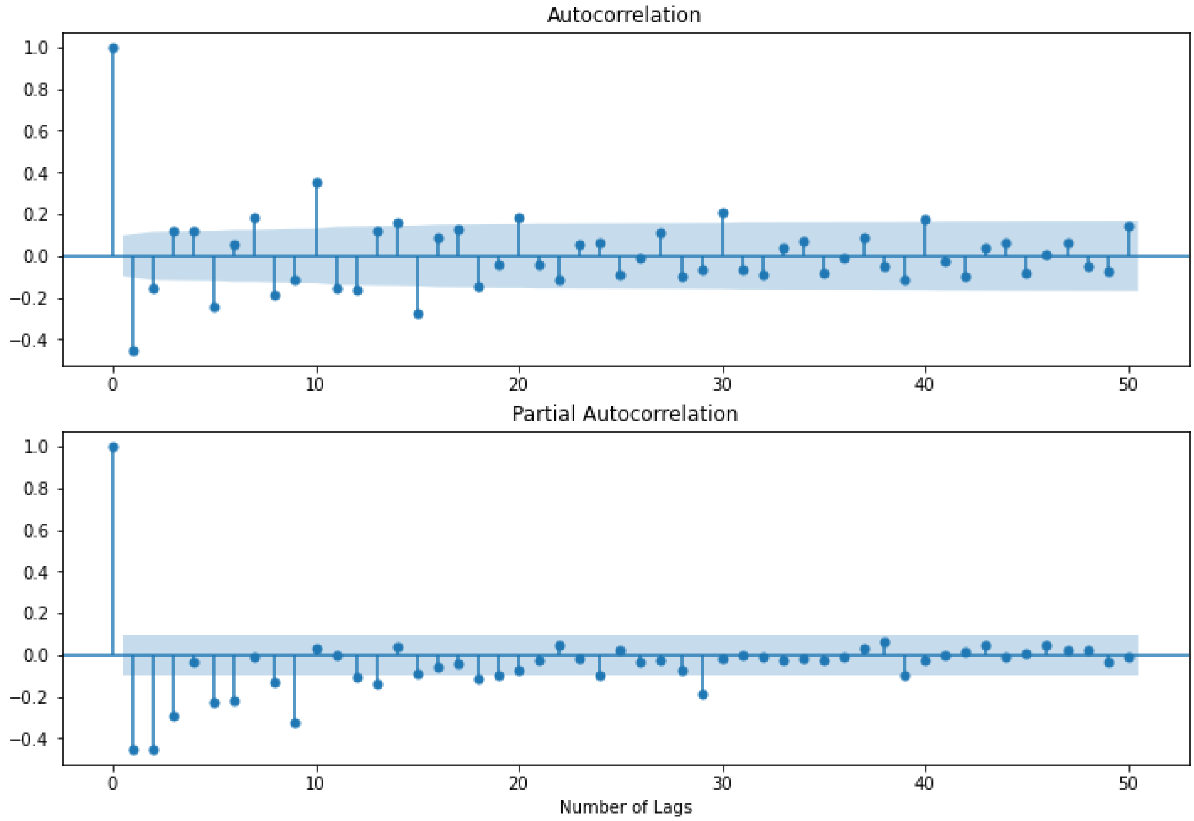


Figure 8: The ACF and PACF plot for recovered case

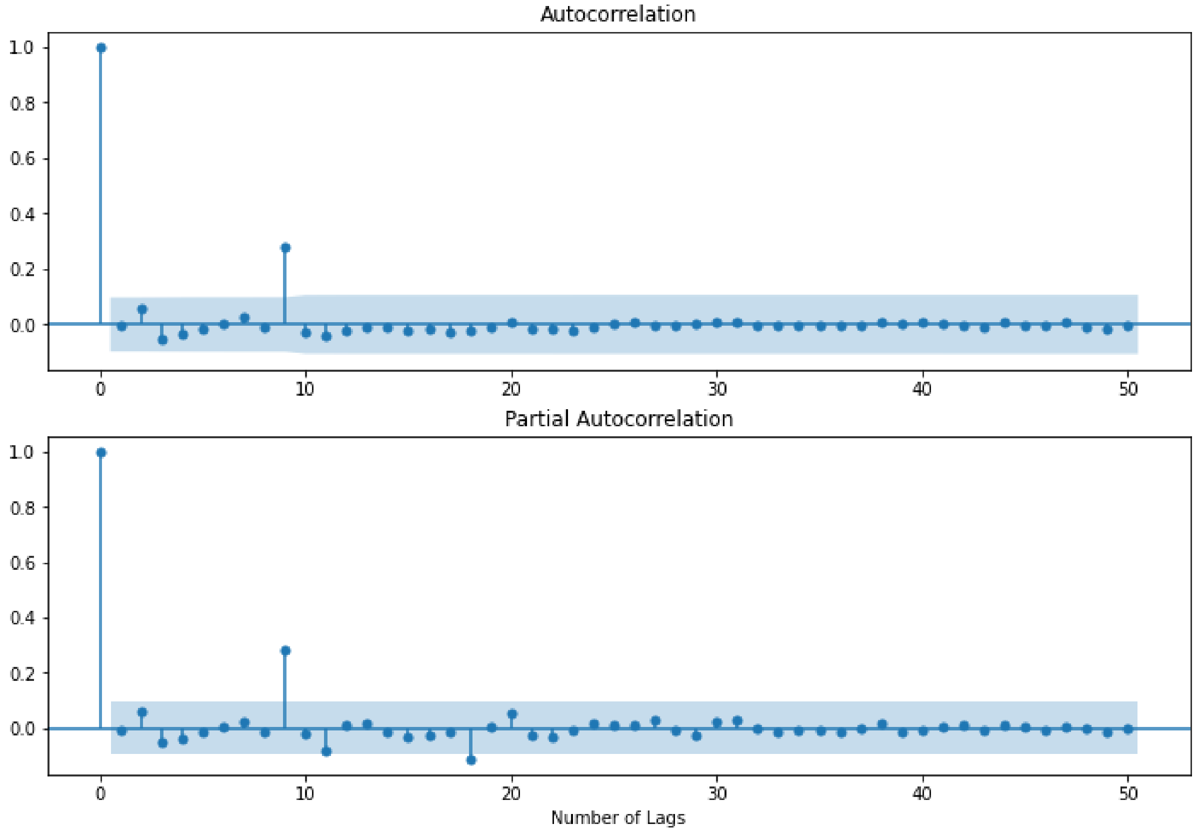
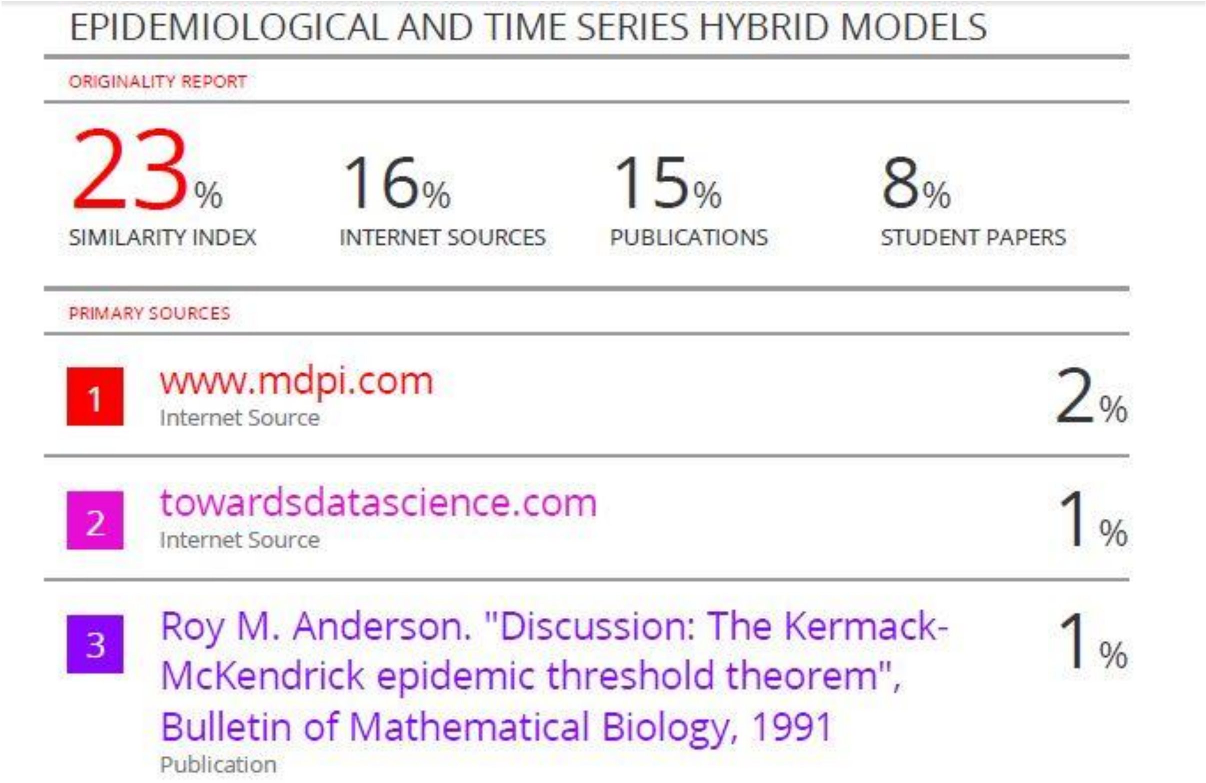


Figure 9: The ACF and PACF plot for deaths case

APPENDIX (3): PLAGIARISM REPORT



EPIDEMIOLOGICAL AND TIME SERIES HYBRID MODELS

ORIGINALITY REPORT

23%	16%	15%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	towardsdatascience.com Internet Source	1%
3	Roy M. Anderson. "Discussion: The Kermack-McKendrick epidemic threshold theorem", Bulletin of Mathematical Biology, 1991 Publication	1%

REFERENCES

[1] "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCov)," World Health Organization, 2020.

[2] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, A. Zumla and E. Petersen, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China," *International journal of infectious diseases: IJID,* vol. 91, p. 264–266, 2020.

[3] "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)," World Health Organization, 2020.

[4] W. O. Kermack and A. McKendrick, "A Contribution to the Mathematical Theory of epidemics," *Society for Mathematical Biology,* vol. 115, pp. 700-721, 1927.

[5] I. Syahrini, Sriwahyuni, V. Halfiani, S. Yuni, T. Iskandar, Rasudin and M. Ramli, "The epidemic of Tuberculosis on vaccinated population," *Journal of Physics: Conference Series,* no. 012017, 2017.

[6] P. E. Lekone and B. F. Finkenstadt, "Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study," *Biometrics,* vol. 62(4), p. 1170–1177, 2006.

[7] T. Yoneyama, S. Das and M. Krishnamoorthy, "A Hybrid Model for Disease Spread and an Application to the SARS Pandemic," *Journal of Artificial Societies and Social Simulation,* vol. 15, 2012.

[8] J. M. Carcione, J. E. Santos, C. Bagaini and J. Ba, "A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model," *Frontiers in public health,* vol. 8, p. 230, 2020.

[9] Z. Tang, X. Li and H. Li, "Prediction of New Coronavirus Infection Based on a Modified SEIR Model," *Preprint,* 2020.

[10] K. Pandey, A. Subedee, B. Khanal and B. Koirala, "COVID-19 Control Strategies and Intervention Effects in Resource Limited Settings: A Modeling Study," *Preprint,* 2020.

[11] S. He, Y. Peng and K. Sun, "SEIR modeling of the COVID-19 and its dynamics," *Springer Nature B.V.,* 2020.

[12] A. Godio, F. Pace and A. Vergnano, "SEIR modeling of the Italian epidemic of SARS-CoV-2," *International Journal of Environmental Research and Public Health,* 2020.

[13] O. Ostashchuk, "Time Series Data Prediction and Analysis," *Czech Technical University in Prague,* 2017.

[14] C. Jofipasi, M. Miftahuddin and H. Sofyan, "Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District," *IOP Conference Series: Materials Science and Engineering 352(1):012055,* 2018.

[15] Y. Peng, B. Yu, P. Wang, D.-G. Kong, B.-H. Chen and X.-B. Yang, "Application of seasonal auto-regressive integrated moving average model in forecasting the incidence of hand-foot-mouth disease in Wuhan, China," *Journal of Huazhong University of Science and Technology,* vol. 37, pp. 842-848, 2017.

[16] O. Olayemi, O. Oluwatosin and O. Segun, "Time Series Analysis on Reported Cases of Tuberculosis in Minna Niger State Nigeria," *Open Journal of Statistics,* vol. 10, pp. 412-430, 2020.

[17] M. Nayak and N. K. A., "Forecasting Dengue Fever Incidence Using ARIMA Analysis," *International Journal of Collaborative Research on Internal Medicine and Public Health,* vol. 11, pp. 924-932, 2019.

[18] L. Ismail, H. Materwala, T. Znati, S. Turaev and A. M. Khan, "Tailoring time series models for forecasting coronavirus spread: Case studies of 187 countries," *Computational and Structural Biotechnology Journal,* vol. 18, no. 2001-0370, pp. 2972-3206, 2020.

[19] O. Ilie, R.-O. Cojocariu, A. Ciobica, S. Timofte, I. Mavroudis and B. Doroftei, "Forecasting the Spreading of COVID-19 across Nine Countries from Europe, Asia, and the American Continents Using the ARIMA Models," *Microorganisms,* vol. 8, no. 1158, 2020.

[20] R. K. Singh, M. Rani, A. S. Bhagavathula, R. Sah, A. J. Rodriguez-Morales, H. Kalita, C. Nanda, S. Sharma, Y. D. Sharma, A. A. Rabaan, J. Rahmani and P. Kumar, "Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model," *JMIR public health and surveillance,* vol. 6(2), no. e19115, 2020.

[21] X. Duan and X. Zhang, "ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data," *Xingde Duan, Xiaolei Zhang, ARIMA modelling and forecasting of irregularly patterned COVID-Data in Brief,* vol. 31, no. 105779, pp. 2352-3409, 2020.

[22] ArunKumar K. E., D. V. Kalaga, C. M. S. Kumar, G. Chilkoor, M. Kawaji and T. M. Brenza, "Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA)," *Applied Soft Computing,* vol. 103, no. 107161, pp. 1568-4946, 2021.

[23] M. Spyros, S. Evangelos and A. Vassilis, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS ONE,* 2018.

[24] P. Zhang and G. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing,* vol. 50, pp. 159-175, 2003.

[25] A. Aslanargun, M. Mammadov, B. Yazici and S. Asma, "Comparison of ARIMA, neural networks and hybrid models in time series: Tourist arrival forecasting," *Journal of Statistical Computation and Simulation,* vol. 77, pp. 29-53, 2007.

[26] L. Yu, L. Zhou, L. Tan, H. Jiang, Y. Wang, S. Wei and S. Nie, "Application of a New Hybrid Model with Seasonal Auto-Regressive Integrated Moving Average (ARIMA) and Nonlinear Auto-Regressive Neural Network (NARNN) in Forecasting Incidence Cases of HFMD in Shenzhen, China," *PloS one,* vol. 9, no. e98241, 2014.

[27] A. Swaraj, A. Kaur, K. Verma, G. Singh, A. Kumar and L. Sales, "Implementation of Stacking Based ARIMA Model for Prediction of Covid-19 Cases in India," *Preprint,* 2020.

[28] M. Ala'raj, M. Majdalawieh and N. Nizamuddin, "Ala'raj, Maher & Majdalawieh,Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections," *Infectious Disease Modelling,* 2020.