A Comparative Study of ARIMA and SARIMA Models to Forecast Lockdowns due to SARS-CoV-2

Hardik Chhabra*, Anubhav Chauhan

Department of Electronics and Communication Engineering, Netaji Subhas University of Technology, Dwarka, New Delhi, India

ABSTRACT

By predicting the implementation of intelligent lockdown before the onset of waves, this research offers an efficient method to address the COVID-19 pandemic. This study introduces unified Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models that are capable of predicting lockdowns in more than 200+ nations. The proposed model was trained over 18,000 dataset of 237 nations and has a response time of 2.5 months. The auto-ARIMA model was used to pick the initial variations of the model parameters and then the optimal model parameters were found based on the best match between the forecasts and test data. The models reliability was evaluated using the analytical methods Auto Correlation Function (ACF), Partial Auto-Correlation Function (PACF), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These models are trained using data acquired from World Health Organizations data repository. The two ARIMA and SARIMA models are clearly gaining an advantage over other studies by having a rapid response time. Besides, a brief comparison of trained ARIMA and SARIMA models are presented and the ARIMA model gained an upper hand due to its accuracy. Additionally, the models are able to predict confirmed death and confirmed cases of COVID. This research shows to be highly beneficial for decision-making about the implementation of smart-lockdowns and could provide another dimension to time-series analysis, which is strongly dependent on models having better response time.

Keywords: ARIMA; SARIMA; Smart-lockdown; Autocorrelation function; Partial autocorrelation function; COVID-19

INTRODUCTION

The novel coronavirus was first identified in Wuhan, China in December 2019. The first case of the disease was reported to the World Health Organization (WHO) on December 31, 2019. COVID-19 is caused by the Severe Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV-2). It belongs to the coronavirus family, which comprises viruses that may infect both animals and humans. The virus is considered to have started in bats and spread to humans. This highly infectious strain of corona virus has already infected over 485,243,022 cases globally and killed 6,137,553 deaths. The previous corona viruses, SARS and MERS, were not as infectious and persistent as the COVID-19 as it has been called. The uncertainty and lack of clarity further exacerbated the issue and many nations were infected with the virus. The virus is extremely infectious and causes death due to

respiratory failure. The virus's distribution has differed between nations due to variances in epidemiological circumstances and diagnostic capabilities. Many nations throughout the world have been affected by Covid-19. The virus has infected people of all ages in practically every country. Germany, France, Italy, Spain and the United States have been hit the worst [1].

The proposed machine learning model in this study is a generalized model that means it is capable of making predictions for over 200+ countries having same parameters which is trained over World Health Organization's data consisting of data for 237 countries, the only requirement to develop this model is a dataset that contains attributes as stated in the meta-data [1]. This study is a novel application of the time series forecasting for COVID-19 data. Additionally, this model only focuses on forecasting lockdown periods, thus less data is required for training since the

---

model won't be forecasting on daily basis thereby increases the response time.

A lot of work has been accomplished in the related areas of research. In the study, there is significant utility in having a prognosticative model for COVID-19 mortality that's primarily based entirely on demographics and underlying conditions, and that doesn't need initial symptoms or biopsy findings [2]. There are many beneficiaries of such a model. Specifically, this can facilitate telemedicine for patients who test positive, warn patients unaffected by COVID-19 of potential risks, and undertake precautions. It can be utilized by the government to put intelligent lockdown techniques into place. Research is conducted to build a short-term predictive model tracking disease cases using the Poisson distribution for the number of daily incidences and the gamma distribution for the serial period to predict the future occurrence of COVID-19 [3]. Second, calculate the reproduction amount on the assumption that it would remain consistent for a brief period of time. Finally, create hypothetical incidence cases based on the posterior distributions of the current rate of transmission assuming this rate remains unchanged or changes by a small amount. The following is a study to predict the spread of COVID-19 in the United States under various resumption strategies. The model predicts that the number of COVID-19 cases will initially increase exponentially at the onset or shortly after the resumption of the infectious event, but then settle into a stable, slightly diminishing protracted stage of disease transmission. In contrast to standard Susceptible Infected Recovered (SIR) models, this pattern is consistent with the number of COVID-19 cases reported in the United States. A prediction for the incidence of new COVID-19 cases is based on various social distance criteria. It concludes that reducing social distance will significantly increase the number of COVID-19 cases [4]. And a comparative study in to analyze and forecast the COVID-19 outbreak in India [5]. In India, the adopted states of Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka and Uttar Pradesh. Three categories of forecasts were made using the ARIMA and prophet time series forecasting models: reported cases, fatalities, and recovered cases. Followed by much more such research on COVID-19 data.

The usage of ARIMA and SARIMA model in studies shown in Table 1 are similar to the study in this research paper, but the results are different and focus on forecasting lockdown periods i.e. start and end time of lockdowns during the crisis (Table 1) [6-9].

**Table 1:** Models reported with similar studies for COVID-19 forecasting.

S.No.	Reference Paper	Remarks
1	[6]	Study conducted on 16 countries to forecast the spread, recovery and confirmed death cases due to novel coronavirus. This paper suggests the use of new Autoregressive Integrated Moving Average (ARIMA)/ Seasonal Autoregressive Integrated Moving Average (SARIMA) models for each country individually. The discussed models required 7 months (January 22 to August 3, 2020) response time.
2	[7]	This study includes a comparison between Gated Recurrent Units (GRU), Long-Short Term Memory (LSTM), ARIMA and SARIMA to forecast the trend of COVID-19. Response time for the models reported in this research is 6 months.
3	[8]	This study reports the use of ARIMA model to forecast the COVID-19 confirm cases trend for 5 countries including India, Russia, Spain, US and Brazil. Total infected cases from the stated five nations are used for training in a daily time series approach from February 15 to June 30, 2020, i.e. response time of 4.5 months.
4	Proposed	Study conducted on 235 countries (as stated in WHO data repository) and developed a unified ARIMA/SARIMA model that works efficiently for 200+ countries. The said models were trained over 18,000 dataset of 237 nations, i.e. response time of 2.5 months approximately.

This paper is organized as follows. After a brief introduction to the research problem, a theoretical framework on ARIMA and SARIMA is presented in Section 2. Theoretical Framework discusses the mathematics and applications of ARIMA and SARIMA respectively. Section 3 introduces the Methodology for the research. The initial step to select the parameters of the proposed ARIMA and SARIMA model using ACF and PACF is discussed. Followed by model design and descriptions of ARIMA and SARIMA model respectively. Section 4 presents the observations for the ARIMA and SARIMA forecasts. Finally, section 5 concludes the paper by presenting results and future scope of this study. The APPENDIX presented the verified results for countries such as Brazil, South Africa, Chile, Ghana and Canada.

LITERATURE REVIEW

This study focuses on the Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) machine learning algorithms to create a machine learning model which is capable of forecasting lockdown periods for novel coronavirus (Figure 1).
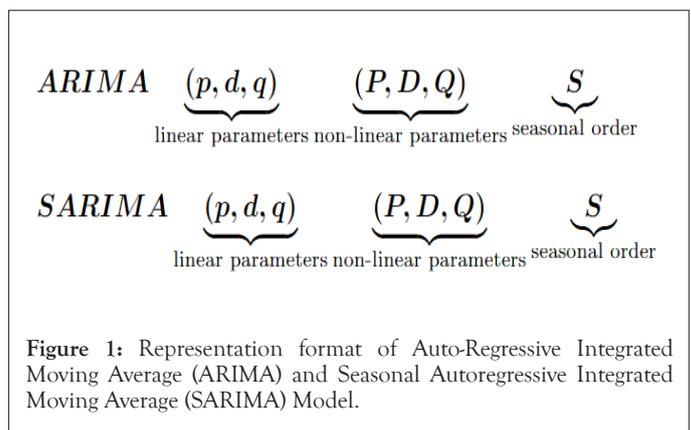
Figure 1: Representation format of Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) Model.

Autoregressive integrated moving average (ARIMA): Developed in 1970, the Box-Jenkins forecasting system is sometimes called ARIMA. In essence, it is an extrapolation technique that forecasts the underlying variable using past time series data. The ARIMA model must be identified, estimated, and evaluated, which are the three processes in the model estimation process [9]. Any variable's future value is regarded in this as a linear function of its preceding values and errors, which are expressed as (Figure 2).
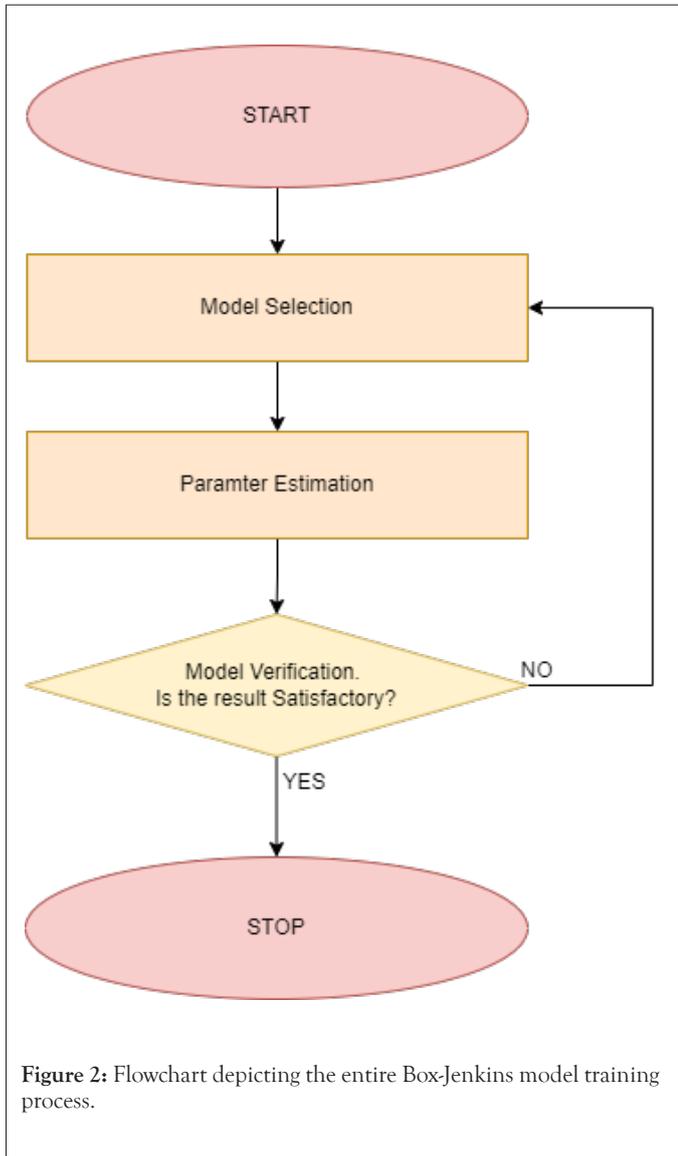

Figure 2: Flowchart depicting the entire Box-Jenkins model training process.

Autoregressive Model, a statistical model known as the AR model illustrates how one variable depends on a previous time period. In this model, the data X_t is dependent on its prior values. For instance, the model $MA(p)$, can be written as p of its previous values.

$$X_t = \alpha + \varepsilon_t + \sum_{i=1}^{p} \alpha_i . X_{t-i} \quad (1)$$

Where, ε_t is the error and α_1, α_2, α_3,..., α_p are coefficients. The order p for the $MA(p)$ model can be chosen based on the PACF plot's significant spikes. The slow decay of the ACF plot is another sign of the AR process.

Moving Average (MA), the weighted average of prior errors is used by the $MA(q)$ model to determine its forecast value. In time series data, it is mathematically derived that the MA model has the capacity to comprehend patterns and trends. For instance, $MA(q)$ for a data X_t can be described as

$$X_t = \mu + \sum_{i=1}^{q} \varepsilon_{t-i} . \theta_i \quad (2)$$

where, μ is the series' mean, θ_1, θ_2, θ_3,..., θ_q are coefficients, and ε_{t-1}, ε_{t-2}, ε_{t-3},..., ε_{t-q} are errors with a normal distribution having zero mean and standard deviation to be unity. (also known as white noise). In comparison to the AR model, we can determine the value of q if the ACF plot has a significant cut after lag q for model $MA(q)$. The slower decay of the PACF plot is another indication of the MA process. Now, an $ARMA(p,q)$ equation can be formulated to be

$$X_t = \alpha + \varepsilon_t + \sum_{i=1}^{p} \alpha_i . X_{t-i} + \mu + \sum_{i=1}^{q} \varepsilon_{t-i} . \theta_i \quad (3)$$

Now an integration operator d can be introduced, defined as follows, to help deal with non-stationary data.

$$X_t^{[d]} = \Delta^d X_t = X_t^{d-1} - X_{t-1}^{[d-1]} \quad (4)$$

where d is the order of differencing and $X_t^{[0]} = X_t$.

Now that we can fit an $ARMA(p,q)$ model to $X_t^{[d]}$ instead of X_t, we get the equation for $ARIMA(p,d,q)$

$$\Delta^d X_t = \alpha + \Delta^d \varepsilon_t + \sum_{i=1}^{p} \alpha_i . \Delta^d X_{t-i} + \mu + \sum_{i=1}^{q} \theta_i . \Delta^d \varepsilon_{t-i} \quad (5)$$

Here, X_t is the value predicted for the variable, which is calculated by adding the total of its arbitrary error term ε_t that occurred in the 't' time period and its lag variables in the previous 't' time period. The coefficients are α_i and θ_j. Here, p and q stand in for the moving average and autoregressive lags, respectively. The Moving Average component (MA) is represented by the second half of the equation, while the Autoregressive part (AR) is represented by the first half of the equation from α_o to $\alpha_p . X_{t-p}$. Consequently, the whole equation is known as the $ARMA(p,q)$ model. The level at which the dependent variable X_t stays static is indicated by the symbol $I(d)$, which stands for the order of integration. In order to account for the dependent variable's lag, the random error from estimation, and the sequence in which the variable becomes steady, ARIMA models are stated as $ARIMA(p,d,q)$. Plotting a correlogram with the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) will reveal the order of p and q [10].

An equation for $ARIMA(5,1,5)$ can mathematically be represented as,

$$\Delta^1 X_t = \alpha + (\varepsilon_t - \varepsilon_{t-1}) + \sum_{i=1}^{5} \alpha_i . (X_{t-i} - X_{(t-1)-i}) + \mu + \sum_{i=1}^{5} \theta_i . (\varepsilon_{t-i} - \varepsilon_{(t-1)-i}) \quad (6)$$

Applications of ARIMA: There have been earlier studies on COVID-19 data to forecast the spread using various supervised learning methods and other hybrid models [11,12]. In this paper, the ARIMA model has been used to forecast the time period at which a lockdown must be imposed over various regions/countries. One of the studies that have been completed on COVID-19 data using the ARIMA model is to forecast the spread of COVID across nine countries [13].

ARIMA model is popular for its usage in forecasting stocks. In, the automatic ARIMA algorithm is discussed to forecast stock

returns from 50 stocks of the Indian National Stock Exchange's minute-wise records [10]. And in the author recreates the ML model for stock market prediction to that of a weekly version of SAARC (India, Sri Lanka, Pakistan) countries using a hybrid model of ARIMA and RNN which decomposes the time-series into its linear and non-linear data components of stock prices [14]. Additionally, research of fifty-six equities from seven sectors is done in the report [15]. The stocks chosen are all listed on the National Stock Exchange (NSE). The set empirical study will use data from the last 23 months. The ARIMA model's accuracy in forecasting stock prices was tested. The best ARIMA model was chosen using AIC. This study also looked at how changing the time span of earlier or past data affected accuracy. The research provides a hybrid technique for predicting stock prices that makes use of the ARIMA and SVMs models' particular strengths. The organized selection of optimal parameters of the hybrid model is of significant relevance for future research [16].

There have been many other uses of the ARIMA Model in the latest AI technology. The ARIMA model is used to predict rainfall in certain regions using various parameters, there is a study of the ARIMA Model predicting the rainfall of the climatological station of Banjarbaru with the help of the Kalman Filter, even so, the ARIMA Model is used to an extent of making hybrid models with Long Short Term Memory (LSTM) to forecast drought using SPEI data [17,18].

Amongst the most important tools used by policymakers throughout the world is to make decisions concerning the forecasting of energy consumption in emerging nations. Most early research in Turkey used multiple statistical modelling techniques. But on the other side, time-series forecasting typically results in better outcomes since expected demographic and economic features typically deviate from the actual results. This study used ARIMA, and SARIMA methods to predict Turkey's primary energy consumption from 2005 to 2020 [19].

A study to develop a forecasting machine learning model to forecast inflation in Kenya is presented in [20]. And in the light of both monetary and revenue-based policymaking, inflation forecasting is critical. In Kenya, like in any other nation, inflation projections are envisioned to allow the Central Bank of Kenya (CBK) to maintain some control over future inflation. In order to model and forecast inflation in Kenya, the use of both ARIMA and GARCH approaches in this work. The results indicated that inflation will rise.

Forecasting demand is a crucial part of supply chain management. It is one of the most essential planning techniques a firm may apply in the future because of its link to other business activities. An ARIMA model for forecasting demand for completed goods in grocery stores is being developed using the time-series Box-Jenkins method [21]. Numerous models are built using historical demand data, and four performance metrics: standard error, maximum likelihood, Schwartz Bayes' criterion, and Akaike's criterion are used to determine which model has the best performance. ARIMA (1, 0, 1) is the chosen model to minimize the four previous criteria. The outcomes point to the possibility of using this model to simulate and estimate future needs in the food manufacturing industry. In this sector, these results will give

managers reliable criteria for making decisions.

Seasonal autoregressive integrated moving average: SARIMA is an extended algorithm that has a seasonal component along with the Auto-Regressive Integrated Moving Average (ARIMA) method. The model assumes that the WHO COVID-19 data comprises trends, seasonal components, and irregular terms.

For $ARMA(p,q)$ equation we will use L operator, which denotes the lag operator, where $L^n x_t = x_{t-n}$

$$X_t = \alpha + \sum_{i=1}^{p} \alpha_i . L^i X_t + \mu + \sum_{i=1}^{q} \theta_i . L^i \varepsilon_t + \varepsilon_t \qquad (7)$$

The equation can be further represented by a polynomial operator, here polynomial operator results in

$\sum_{i=1}^{q} \theta_i . L^i \varepsilon_t = \theta(L)^q \varepsilon_t$, thus we obtain

$$X_t = \alpha(L)^p X_t + \theta(L)^q \varepsilon_t + \varepsilon_t \quad (8)$$

It can be assumed that $ARIMA(p,d,q)$ equation will turn out to be

$$\Delta^d X_t = \alpha(L)^p \Delta^d X_t + \theta(L)^q \Delta^d \varepsilon_t + \Delta^d \varepsilon_t \quad (9)$$

Now, using seasonal lags and an $ARMA(P,Q)$ model on the differenced values, we can extract any remaining structure. In other words, we use L^s rather than the standard lag operator L. Once more, P and Q are seasonal time lags.

$$\Delta_s^D X_t = A(L^s)^p \Delta_s^D X_t + \Theta(L^s)^Q \Delta_s^D \varepsilon_t + \Delta_s^D \varepsilon_t \quad (10)$$

We can now apply another $ARIMA(p,d,q)$ model to $\Delta D_s X_t$ by multiplying the seasonal model by the new ARIMA model in order to remove any remaining seasonality and obtain a mathematical representation of $SARIMA(p,d,q)(P,D,Q,S)$. Thus, the equation for $SARIMA(5,1,5)(1,0,2,9)$ can be formulated by replacing the values in the equation (11).

$$\Delta^d \Delta_s^D X_t = \alpha(L)^p A(L^s)^p \Delta^d \Delta_s^D X_t + \theta(L)^q \Theta(L^s)^Q \Delta^d \Delta_s^D \varepsilon_t + \Delta^d \Delta_s^D \varepsilon_t \quad (11)$$

The SARIMA models are created using the approaches described in this article. The order of seasonal differencing is determined by the OCSB test [22]. Then the order of differencing is established by the KPSS unit-root test [23]. The order of the autoregressive and moving average components, p, q, P, and Q, is then measured by traversing the model space using stepwise processes. Finally, the best model is selected using tests based on AIC or BIC and estimated residuals from Oliveira et al. [24-27].

Applications of SARIMA: The study on the SARIMA model has shown that this model has various uses in forecasting data related to-Weather and Temperature forecasting, fire frequency or accidents forecasting, healthcare industry, Tourism, and the financial sector [28-32]. Talking about the healthcare industry, the current paper deals with the analysis and forecasting of region-wise smart lockdowns. Paper describes a SARIMA Model capable of forecasting monthly temperature values in Ghana's northern area [28]. The SARIMA (1,0,0)(1,0,0)(12) is the best-fitting model, according to the results. The model produced will aid in the development of future agriculture and tourist initiatives in the region.

A study to look into the implementation of the SARIMA model is conducted to predict the fire accident frequency as studied in reference Ma et al., [29]. Here the authors have worked on the China Fire Service Data from 2004 to 2018 and developed

an optimized predictive model SARIMA (1,1,1)(1,1,1)(12). It is claimed to have increased the accuracy of the model by 11.5% with a root mean square of the predicted data from 2018 to 2019 to be 2826.93. This result may provide a better understanding of fire aggregation laws and provide theoretical support for fire protection and fire protection activities by fire departments.

Qi et al., [30] is a comparative study of SARIMA and Seasonal Auto-Regressive Fractionally Integrated Moving Average (SARFIMA) to predict the incidence of Haemorrhagic Fever with Renal Syndrome (HFRS). Its conclusion states the better predictive performance of SARFIMA over SARIMA with 6.7% Root Mean Square Error which is 5% less than the SARIMA model. Moreover, there are studies that are able to predict the dengue cases in Campinas, State of São Paulo, Brazil [33]. The model is based on the monthly incidence of dengue from 1998 to 2008. And is tested using data gathered between January and December 2009. SARIMA (2,1,2) (1,1,1) (12) was found to be the model that best suited the data. The number of dengue cases in a given month may be anticipated using the number of dengue cases one, two, and twelve months before, according to this model. The anticipated 2009 numbers are quite similar to the actual ones. And a study is conducted in KwaZulu-Natal, South Africa to predict monthly malaria cases [34]. The research was conducted utilizing a two-part clinically confirmed monthly malaria case dataset. The training dataset was used to create a SARIMA model using the Box-Jenkins technique from January 2005 to December 2013, while the testing dataset was used to evaluate the forecast given by the best-fit model from January to December 2014. As a consequence, the SARIMA (0,1,1) (0,1,1) (12) model may be used to model and predict monthly malaria incidences in KZN.

Cross-correlation functions are used to create comparative structures using different models in monthly natural gas production and consumption forecasts to evaluate relationships between extrinsic variables. In addition, for the first time in a literature review of study [35], SARIMA model for monthly predictions of US natural gas production and consumption will be tested using the proposed method. The performance of the model is compared with SARIMA (p,d,q)*(P,D,Q)S. The Root-Mean-Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) results show that the best-fit model is superior. This strategy forecasts monthly natural gas production and consumption in the United States by 2025.

A technique to predict the detection duration of different objects used in the reported smart home context. Prognosis will help quantify an elderly person's health and alert them to routine or abnormal behavior in daily activities. Seasonal Automatic Regression Integrated Average (SARIMA) procedures are used in health modelling to predict the functioning of common objects used by an elderly person living alone [36].

The purpose of the paper's research is to determine the utility of analytics models in developing a system capable of providing an approximate estimate of future pollution levels within a reasonable confidence interval [37]. For time-dependent data, conventional linear regression approaches are shown to be inadequate. In this context, a time series forecasting technique is employed to estimate future levels of several contaminants with a high degree of confidence. The experimental investigation of forecasting for Bhubaneswar City's air pollution levels shows that the suggested strategy employing SARIMA and the Prophet model is successful.

Some Machine Learning Models have also been trained in hybrid mode with the SARIMA Algorithm. For forecasting the daily tourism arrivals at Macau SAR, China, the SARIMA-LSTM Hybrid model (using 2-year tourism data) is developed by the authors in reference Wu et al., [31] which claims to outperform the previously used SARIMA model for the same purpose. In addition, the hybrid SARIMA-FFNN model was tested to predict Kediri money inflows and outflows using hidden layers selected by the minima of Mean Absolute Percentage Error (MAPE) and Root-Mean-Squared Error (RMSE). The hybrid SARIMA-FFNN shows excellent predictive performance [32].

A hybrid model study that combines the benefits of both strategies is provided [38]. This model, executed in four steps, (SARIMA-ANFIS) incorporates a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) and a new hybrid model, the Adaptive Neurophagy Inference System (ANFIS). In the first step, the data is pre-processed for linear component modelling. In the second stage, the SARIMA model estimates linear and nonlinear factors. In the third phase, the linear and non-linear components of the decomposed inflow and the validity of the parameters of the SARIMA model are examined in a specific fitting test. After selecting the component's sovereign, ANFIS models the non-linear component in the fourth step. This technique uses ANN modelling (SARIMA-ANN model) instead of ANFIS. The results show that the accuracy of ANFIS is more than ANN and SARIMA coupled-ANN models, making SARIMA-ANFIS the best of all. A similar study in is employed where a new hybrid approach SARIMA-GEP combines a Seasonal Autoregressive Integrated Moving Average (SARIMA) model with a Gene Expression Programming (GEP) model to predict monthly reservoir inflows [39].

```
Index: 194103 entries, 2020-01-03 to 2022-03-31
Data columns (total 7 columns):
 #   Column             Non-Null Count    Dtype
---  ------             --------------    -----
 0   Country_code       193284 non-null   object
 1   Country            194103 non-null   object
 2   WHO_region         194103 non-null   object
 3   New_cases          194103 non-null   int64
 4   Cumulative_cases   194103 non-null   int64
 5   New_deaths         194103 non-null   int64
 6   Cumulative_deaths  194103 non-null   int64
dtypes: int64(4), object(3)
memory usage: 11.8+ MB
```
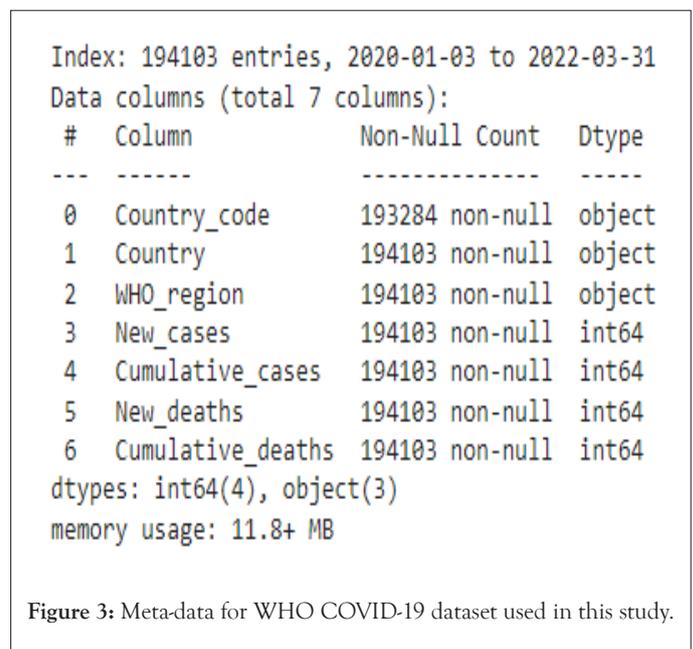
Figure 3: Meta-data for WHO COVID-19 dataset used in this study.

MATERIALS AND METHODS

The data for this study was acquired from World Health Organization (WHO). Figure 3 is the metadata for the COVID data being used to analyze and develop forecasting machine learning models.

ACF and PACF

The data is tested for ACF and PACF plots. This is a very crucial step in getting the values of AR(p), I(d), and MA(q) to build the optimized model. Although the Automatic ARIMA function is used that finds out the best-fit model by minimizing the Akaike Information Criterion (AIC), Figure 4 shows the full implementation of the automatic ARIMA function which shows the testing of various ARIMA models. Calculations of the graph for the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) are made [40].

```
Performing stepwise search to minimize aic
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=129589.685, Time=40.97 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=136467.249, Time=0.86 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=135286.437, Time=1.76 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=133472.207, Time=18.38 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=136465.252, Time=9.41 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=132427.499, Time=25.22 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=132282.045, Time=22.80 sec
 ARIMA(3,1,2)(0,0,0)[0] intercept   : AIC=129045.163, Time=40.84 sec
 ARIMA(3,1,1)(0,0,0)[0] intercept   : AIC=132128.042, Time=22.50 sec
 ARIMA(4,1,2)(0,0,0)[0] intercept   : AIC=126518.405, Time=66.66 sec
 ARIMA(4,1,1)(0,0,0)[0] intercept   : AIC=131667.661, Time=41.57 sec
 ARIMA(5,1,2)(0,0,0)[0] intercept   : AIC=125081.883, Time=66.11 sec
 ARIMA(5,1,1)(0,0,0)[0] intercept   : AIC=129725.032, Time=25.28 sec
 ARIMA(5,1,3)(0,0,0)[0] intercept   : AIC=124986.023, Time=85.17 sec
 ARIMA(4,1,3)(0,0,0)[0] intercept   : AIC=129625.657, Time=73.17 sec
 ARIMA(5,1,4)(0,0,0)[0] intercept   : AIC=122408.768, Time=94.69 sec
 ARIMA(4,1,4)(0,0,0)[0] intercept   : AIC=125983.957, Time=92.35 sec
 ARIMA(5,1,5)(0,0,0)[0] intercept   : AIC=122153.733, Time=108.90 sec
 ARIMA(4,1,5)(0,0,0)[0] intercept   : AIC=122949.180, Time=97.43 sec
 ARIMA(5,1,5)(0,0,0)[0]             : AIC=122096.887, Time=97.99 sec
 ARIMA(4,1,5)(0,0,0)[0]             : AIC=122944.088, Time=88.36 sec
 ARIMA(5,1,4)(0,0,0)[0]             : AIC=122404.895, Time=83.79 sec
 ARIMA(4,1,4)(0,0,0)[0]             : AIC=125982.138, Time=80.60 sec

Best model:  ARIMA(5,1,5)(0,0,0)[0]
```

Figure 4: Testing of various Autoregressive Integrated Moving Average (ARIMA) models by Automatic ARIMA function.

ARIMA using ACF and PACF for new cases and new deaths: The ACF and PACF plots are depicted in Figures 5-8, for ARIMA model that helps in determining the initial parameters for the ARIMA model (Figures 5-8).
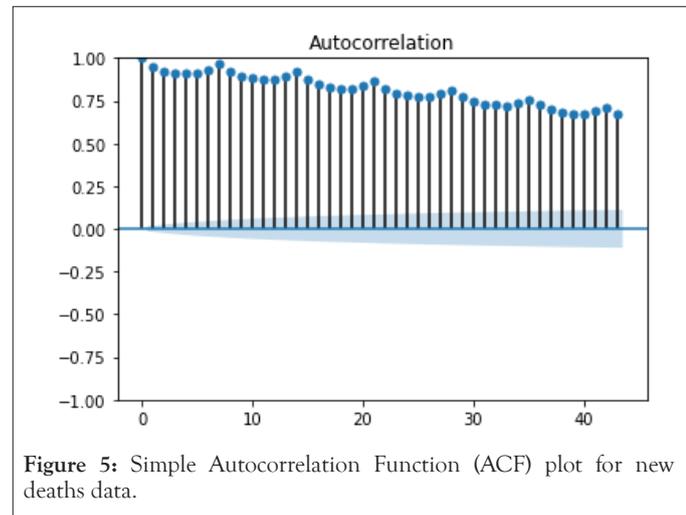
Figure 5: Simple Autocorrelation Function (ACF) plot for new deaths data.
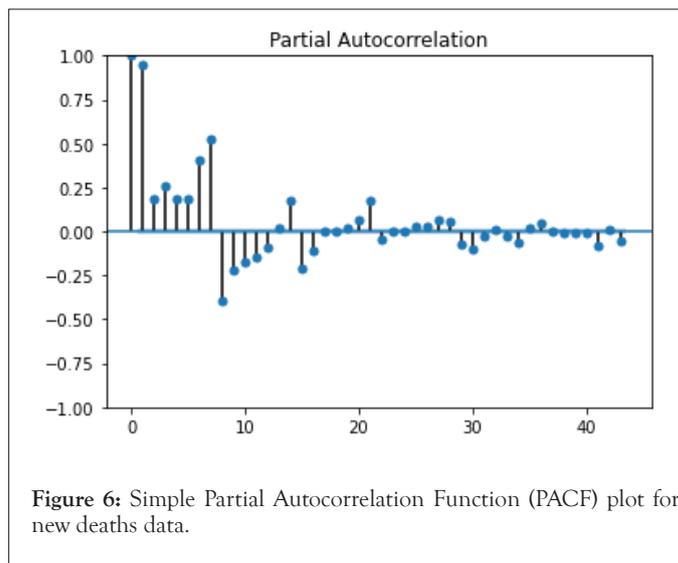
Figure 6: Simple Partial Autocorrelation Function (PACF) plot for new deaths data.
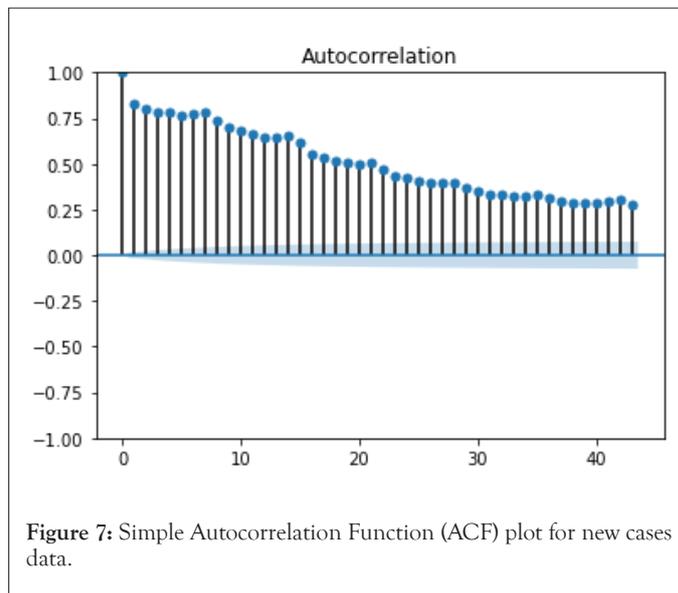
Figure 7: Simple Autocorrelation Function (ACF) plot for new cases data.
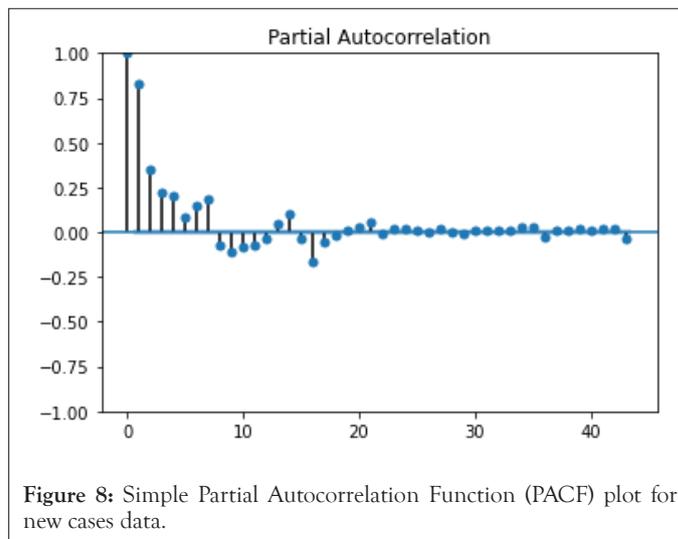
Figure 8: Simple Partial Autocorrelation Function (PACF) plot for new cases data.

SARIMA using ACF and PACF for new cases and new deaths: The ACF and PACF for SARIMA model are presented for new deaths and new cases data to calculate the initial parameters for the model, as shown in Figures 9-12.
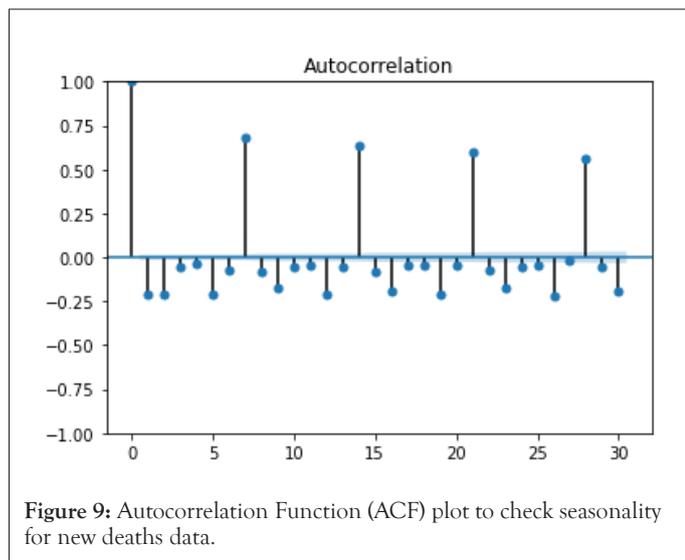
Figure 9: Autocorrelation Function (ACF) plot to check seasonality for new deaths data.

Figure 10: Partial Autocorrelation Function (PACF) plot to check seasonality for new deaths data.
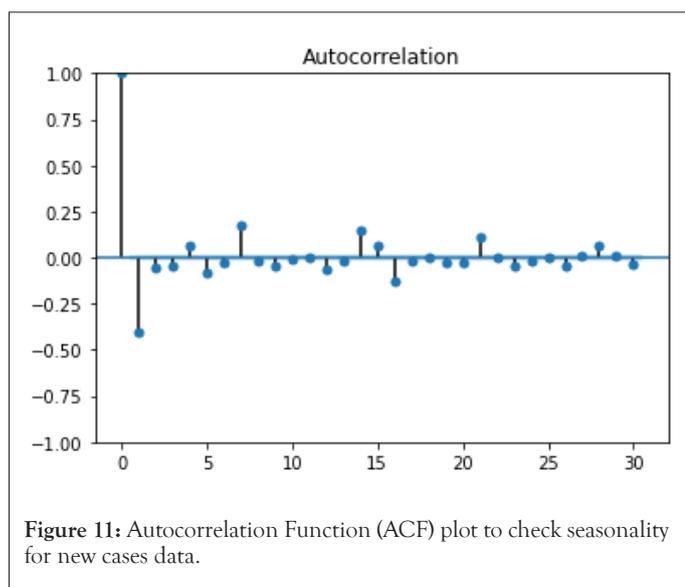
Figure 11: Autocorrelation Function (ACF) plot to check seasonality for new cases data.

Figure 12: Partial Autocorrelation Function (PACF) plot to check seasonality for new cases data.

Model design for ARIMA and SARIMA

The statistics of the trained ARIMA and SARIMA model as depicted in Figures 13-16, where ar.L1, ar.L2, ar.L3, ... corresponds to $\alpha_1, \alpha_2, \alpha_3, ...$ from the equations (6) and (11) i.e. they belong to linear parameters of Autoregressive (AR) component. Similar to AR component, ma.L1, ma.L2, ma.L3,... belongs to Moving Average (MA) component from linear parameters and sigma2 represents error. However in Figures 15 and 16, ar.S.L9 belongs to non-linear parameters of AR components representing seasonality to be 9 and ma.S.L9 and ma.S.L18 represents MA of non-linear components. Moreover, it is clearly mentioned that the ARIMA and SARIMA models took 18,000 data for confirmed cases and confirmed deaths for 237 countries and the model is judged on the basis of AIC, BIC and Hannan-Quinn Information Criterion (HQIC) (Figures 13-16).
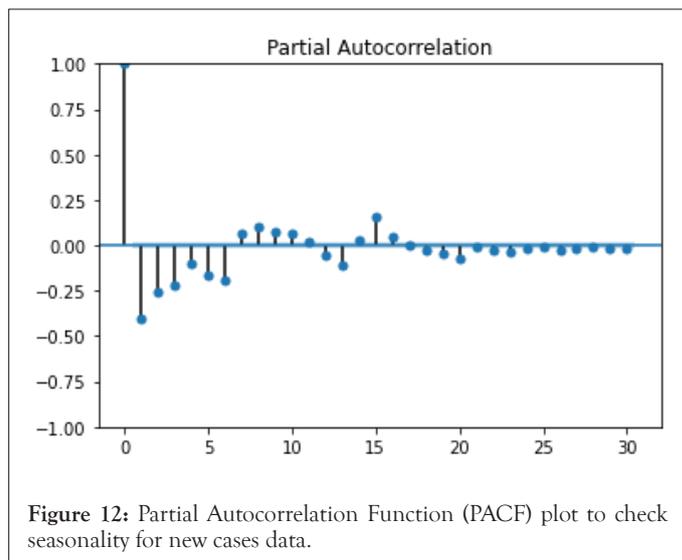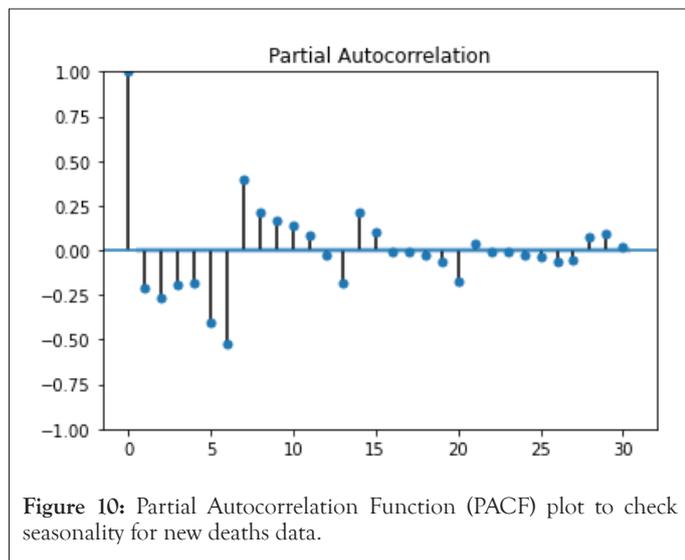
```
                              SARIMAX Results
================================================================================
Dep. Variable:              New_deaths   No. Observations:             18000
Model:               SARIMAX(5, 1, 5)    Log Likelihood           -67608.889
Date:                Mon, 08 Aug 2022    AIC                      135243.779
Time:                        08:33:20    BIC                      135345.154
Sample:                             0    HQIC                     135277.116
                              - 18000
Covariance Type:                  opg
================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
x1            -3.1690      3.584     -0.884      0.377     -10.194       3.856
x2             0.7334      1.489      0.492      0.622      -2.186       3.652
ar.L1          0.4090      0.002    167.964      0.000       0.404       0.414
ar.L2         -1.1125      0.002   -534.224      0.000      -1.117      -1.108
ar.L3          0.2249      0.004     63.622      0.000       0.218       0.232
ar.L4         -0.6708      0.002   -334.414      0.000      -0.675      -0.667
ar.L5         -0.3853      0.002   -160.016      0.000      -0.390      -0.381
ma.L1         -1.0526      0.002   -427.469      0.000      -1.057      -1.048
ma.L2          1.4259      0.003    486.888      0.000       1.420       1.432
ma.L3         -1.0350      0.004   -284.671      0.000      -1.042      -1.028
ma.L4          0.9147      0.002    387.569      0.000       0.910       0.919
ma.L5         -0.2383      0.002   -120.188      0.000      -0.242      -0.234
sigma2       107.5040      0.139    770.700      0.000     107.231     107.777
================================================================================
Ljung-Box (L1) (Q):                0.13   Jarque-Bera (JB):       30642510.56
Prob(Q):                           0.71   Prob(JB):                      0.00
Heteroskedasticity (H):            2.24   Skew:                          3.41
Prob(H) (two-sided):               0.00   Kurtosis:                    205.02
================================================================================
```

Figure 13: ARIMA model summary for death as the dependent variable.

```
                         SARIMAX Results
==============================================================================
Dep. Variable:              New_cases   No. Observations:           18000
Model:              SARIMAX(5, 1, 5)   Log Likelihood        -133380.485
Date:             Mon, 08 Aug 2022    AIC                    266786.970
Time:                     08:31:40    BIC                    266888.345
Sample:                          0    HQIC                   266820.307
                           - 18000
Covariance Type:               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
x1          -1446.3423     22.218    -65.098      0.000   -1489.888   -1402.796
x2            273.4060      7.549     36.218      0.000     258.611     288.201
ar.L1           0.3627      0.008     47.579      0.000       0.348       0.378
ar.L2          -1.0875      0.006   -171.301      0.000      -1.100      -1.075
ar.L3           0.1861      0.011     16.890      0.000       0.164       0.208
ar.L4          -0.6377      0.006   -102.236      0.000      -0.650      -0.626
ar.L5          -0.4082      0.007    -54.480      0.000      -0.423      -0.394
ma.L1          -0.7619      0.008   -100.323      0.000      -0.777      -0.747
ma.L2           1.2280      0.008    152.513      0.000       1.212       1.244
ma.L3          -0.6947      0.012    -60.018      0.000      -0.717      -0.672
ma.L4           0.7503      0.007    103.195      0.000       0.736       0.765
ma.L5          -0.0290      0.007     -4.152      0.000      -0.043      -0.015
sigma2       2.106e+05    592.302    355.530      0.000    2.09e+05    2.12e+05
===================================================================================
Ljung-Box (L1) (Q):                0.06   Jarque-Bera (JB):       14650678.83
Prob(Q):                           0.81   Prob(JB):                      0.00
Heteroskedasticity (H):            6.65   Skew:                         -2.81
Prob(H) (two-sided):               0.00   Kurtosis:                    142.66
===================================================================================
```

Figure 14: ARIMA model summary for cases as the dependent variable.

```
                         SARIMAX Results
==========================================================================================
Dep. Variable:                      New_deaths   No. Observations:           18000
Model:           SARIMAX(5, 1, 5)x(1, 0, [1, 2], 9)  Log Likelihood      -66919.147
Date:                           Mon, 08 Aug 2022    AIC                   133866.294
Time:                                   13:58:11    BIC                   133975.467
Sample:                                        0    HQIC                  133902.195
                                         - 18000
Covariance Type:                             opg
==========================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
ar.L1           0.2569      0.002    109.742      0.000       0.252       0.262
ar.L2          -0.9971      0.002   -508.550      0.000      -1.001      -0.993
ar.L3           0.0305      0.003      9.125      0.000       0.024       0.037
ar.L4          -0.5432      0.002   -276.487      0.000      -0.547      -0.539
ar.L5          -0.5125      0.002   -225.557      0.000      -0.517      -0.508
ma.L1          -0.9084      0.002   -379.202      0.000      -0.913      -0.904
ma.L2           1.1471      0.003    387.445      0.000       1.141       1.153
ma.L3          -0.7569      0.004   -215.350      0.000      -0.764      -0.750
ma.L4           0.6916      0.002    284.426      0.000       0.687       0.696
ma.L5          -0.0876      0.002    -41.091      0.000      -0.092      -0.083
ar.S.L9        -0.1644      1.997     -0.082      0.934      -4.078       3.749
ma.S.L9         0.4948      1.997      0.248      0.804      -3.418       4.408
ma.S.L18        0.0536      0.660      0.081      0.935      -1.241       1.348
sigma2         96.0538      0.132    728.047      0.000      95.795      96.312
==============================================================================================
Ljung-Box (L1) (Q):                0.35   Jarque-Bera (JB):       17890365.50
Prob(Q):                           0.55   Prob(JB):                      0.00
Heteroskedasticity (H):            2.16   Skew:                          2.05
Prob(H) (two-sided):               0.00   Kurtosis:                    157.40
==============================================================================================
```

Figure 15: SARIMA model summary for deaths as the dependent variable.

```
                         SARIMAX Results
==========================================================================================
Dep. Variable:                       New_cases   No. Observations:           18000
Model:           SARIMAX(5, 1, 5)x(1, 0, [1, 2], 9)  Log Likelihood     -133422.095
Date:                           Mon, 08 Aug 2022    AIC                   266872.191
Time:                                   14:00:13    BIC                   266981.364
Sample:                                        0    HQIC                  266908.092
                                         - 18000
Covariance Type:                             opg
==========================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
ar.L1           0.3115      0.008     40.105      0.000       0.296       0.327
ar.L2          -1.0430      0.006   -161.040      0.000      -1.056      -1.030
ar.L3           0.1039      0.011      9.205      0.000       0.082       0.126
ar.L4          -0.5949      0.006    -92.141      0.000      -0.608      -0.582
ar.L5          -0.4583      0.008    -60.404      0.000      -0.473      -0.443
ma.L1          -0.6986      0.008    -89.607      0.000      -0.714      -0.683
ma.L2           1.1032      0.009    126.705      0.000       1.086       1.120
ma.L3          -0.5587      0.012    -45.776      0.000      -0.583      -0.535
ma.L4           0.6629      0.008     85.071      0.000       0.648       0.678
ma.L5           0.0332      0.007      4.599      0.000       0.019       0.047
ar.S.L9        -0.2230      2.871     -0.078      0.938      -5.850       5.404
ma.S.L9         0.4102      2.871      0.143      0.886      -5.216       6.037
ma.S.L18        0.0401      0.539      0.074      0.941      -1.016       1.096
sigma2       2.148e+05    355.770    603.800      0.000    2.14e+05    2.16e+05
==============================================================================================
Ljung-Box (L1) (Q):                0.62   Jarque-Bera (JB):       35137597.14
Prob(Q):                           0.43   Prob(JB):                      0.00
Heteroskedasticity (H):            8.18   Skew:                         -4.78
Prob(H) (two-sided):               0.00   Kurtosis:                    219.24
==============================================================================================
```

Figure 16: SARIMA model summary for cases as the dependent variable.

After finding the appropriate parameters to work with the ARIMA and SARIMA models, these models were then trained using those parameters. Otherwise the automatic ARIMA function gave us the ARIMA (5,1,5)(0,0,0,0) as the best-fit model. Elsewhere, Figures 13 and 14 have the complete ARIMA model summary of deaths and cases due to COVID as the dependent variables respectively.

The SARIMA model summary with Deaths as the dependent variable on which the model has been trained for SARIMA (5,1,5) (1,0,2,9) as shown in Figure 15. And a SARIMA model with the same parameters is developed over new cases as the dependent variable.

ARIMA model description: There is a need to calculate the order (p, d, q) of an autoregressive moving average sequence for building a dependable forecasting ARIMA model, and studies [41-45] tell us different algorithms to determine the order of ARIMA model. These suggest an iterative/recursive approach to determine the order of the ARIMA model [46,47] or the use of a pattern identification approach using the autocorrelation function else by minimizing the AIC [48,49].

Following the selection of a specific model from the general class, effective statistical techniques are used to ascertain the model's parameters. To estimate parameters, the most widely used computer programs use a non-linear least-squares technique, maximum likelihood, or harmony search algorithm as used in the study [24,41].

$$AIC = (-2)\log(\text{maximum likelihood}) + 2$$

$$(\text{number of independently adjusted parameters within the model}) \quad (7)$$

$$BIC = (-2)\log(\text{maximum likelihood}) + 2(\text{number of samples used for fitting}) \quad (8)$$

One can calculate the maximum likelihood for the ARIMA model to perform the statistical task for model verification [41]. Furthermore, we require some objective function that will evaluate model performance on a validation set in order to determine the ideal pairing of p and q. Normally, this task in accomplished by Akaike Information Criterion (AIC) and Bayesian Information Criterion. Calculations of both techniques are based on the maximum likelihood for the autoregressive model [26]. The latest automatic ARIMA model uses AIC to find the best-fit model.
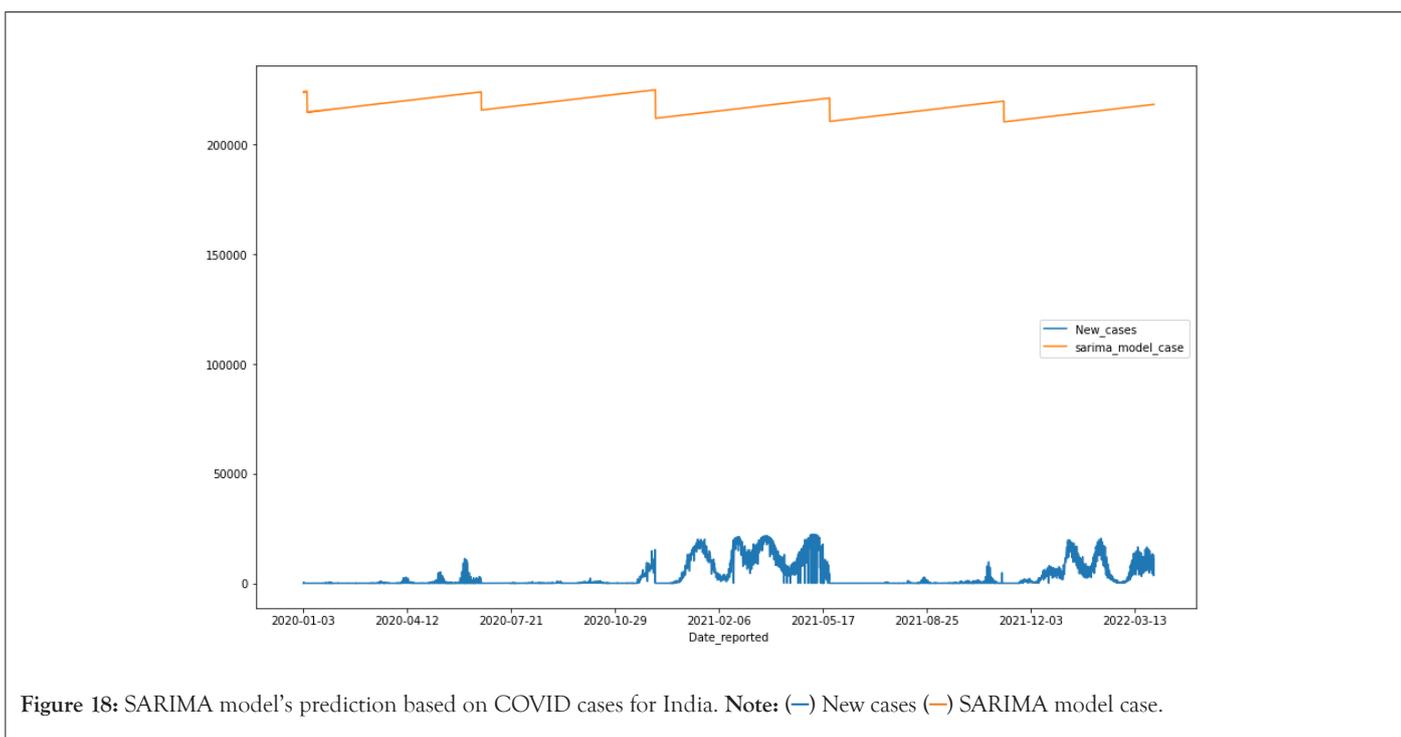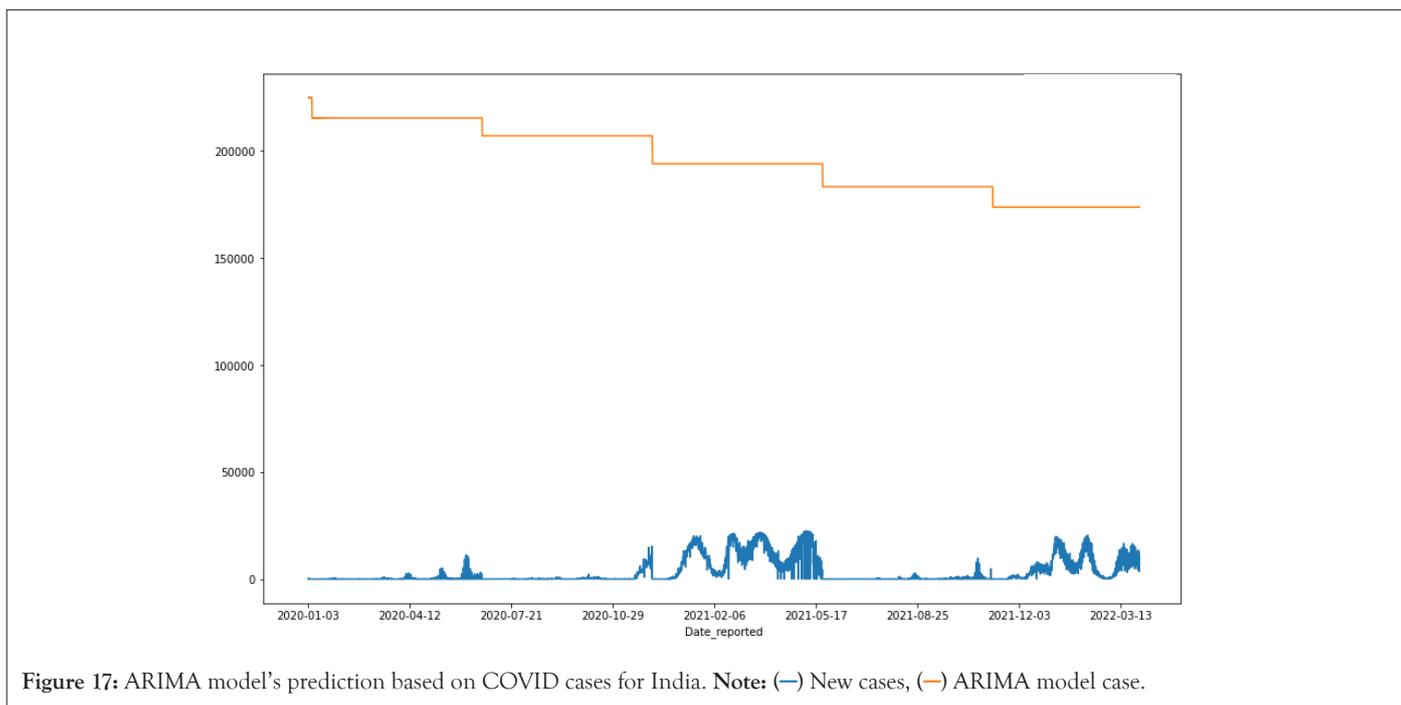
The next step is to fit/train the best-fit autoregressive model as studied in [27]. Once a suitable model has been fitted to the data, it may be extrapolated to get predictions of future time series data. It's also feasible to construct confidence intervals around these point projections if one considers normalcy.

SARIMA model description: Before a SARIMA function is implemented; various tests were conducted to analyse the data set. These tests are separated into stages. Firstly, the data is pre-processed to make sure that it is in the appropriate form so that a time-series-based analysis can be conducted over it. Next, the data is processed *via* ACF and PACF functions to check for reliability of the data in the case of AR(p), I(d), and MA(q) factors. Later comes the stage where OCSB Test and KPSS unit-root tests occur. Now, let the various algorithms perform training over various

parameters of the SARIMA model and find the best-fit model based on parameters. The last stage is to train the best fit model for SARIMA and forecast the values against the test data. This stage will determine the results that are obtained, whether it is up to the expectations or failed to meet the expectations. One can verify this by Root Mean Square Error, R-squared function, Mean Squared Error, etc. As discussed for the ARIMA, the same steps are employed to build a SARIMA model capable of forecasting. One can also use Akaike's information criterion, maximum likelihood, or harmony search algorithm [24, 41] to obtain the parameters that will determine the best-fit model.

Observation

Observations from the resulting graphs show the forecasted result along with actual values. The forecasted result of the ARIMA and SARIMA model are shown in Figures 17 and 18 respectively. One can observe that the predicted result suggests smart lockdown at various time intervals that will help control the spread of the novel coronavirus. The observations also show that the same ARIMA and SARIMA model parameters can be used to predict the results of models with deaths occurring due to COVID Virus as the dependent variable. Predictions made by the ARIMA and SARIMA models are depicted in Figures 19 and 20.
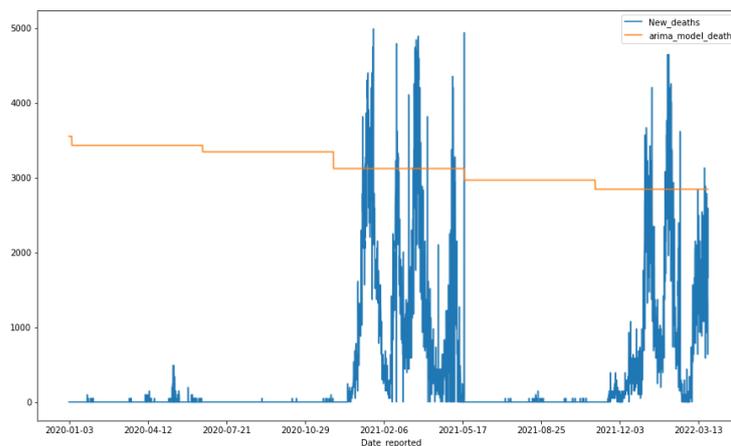

Figure 17: ARIMA model's prediction based on COVID cases for India. **Note:** (—) New cases, (—) ARIMA model case.

Figure 18: SARIMA model's prediction based on COVID cases for India. **Note:** (—) New cases (—) SARIMA model case.

Figure 19: ARIMA model's prediction based on death for India. **Note:** (—) New deaths, (—) ARIMA model deaths.
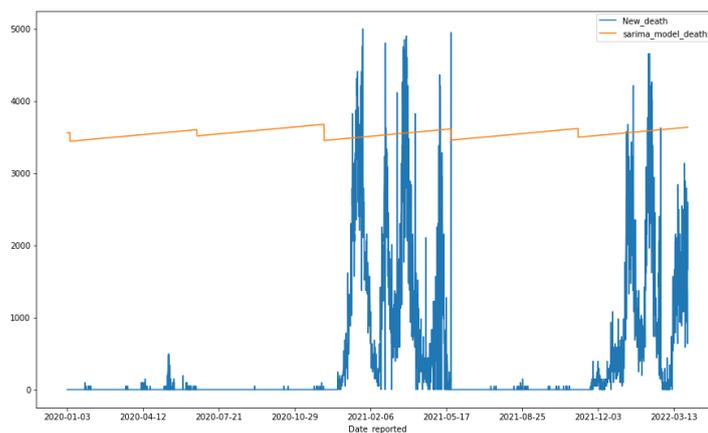

Figure 20: SARIMA model's prediction based on death for India. **Note:** (—) New deaths, (—) SARIMA model deaths.

RESULTS AND DISCUSSION

Time-series analysis and machine learning are used to make predictions about the spread and fatalities of the novel coronavirus. This information can be used to alert stakeholders and prepare for potential spikes in cases. It can also help with planning lockdowns and informing citizens, which is especially important in countries with high population of migrant workers, such as India and China. The goal is to use this information to assist people and government agencies in managing the pandemic effectively.

The ARIMA and SARIMA machine learning models produced resulting graphs which can predict the spike of new deaths and new cases due to the COVID pandemic. Based on this prediction, the models have forecasted the time periods to execute smart-lockdowns. These are step-line graphs and the severity to implement a lockdown depends on the size of each step. The feature that can inform us about the urgency of lockdown implementation

can either be numerically computed or computationally acquired from the given forecasted result. The bigger the size of the step means our machine learning model forecasted that there will be a severe spread of cases and consecutive deaths thus a need to impose lockdown in the upcoming time period is required. Refer to APPENDIX and confirm the results of nations such as Brazil, Ghana, Canada, Chile and South Africa.

Future scope

This study to develop Machine Learning models capable of forecasting the lockdown periods for the spread of new cases as well as death cases of diseases such as novel coronavirus which spreads *via* air medium that makes this disease capable of widespread. This study proves to be very useful in predicting the lockdown periods which alerts the people about the potential threat from the disease. It is believed that such methodology must be adopted if such pandemic occurs in future.

It has been emphasized that gathering sufficient training data

is essential for the machine learning model to be accurate and generalizable. However, precedence to the quality, not quantity, of the data is given. The forecasted data from the model has the feature of determining severity of situation based on the size of each step. Determining the step size requires more computation power.

CONCLUSION

Time-series analysis and machine learning can be used to forecast the SARS-CoV-2 spread and deaths. The predictions can alert stakeholders and enable them to prepare for potential increases in cases. In addition, this information can help plan lockdowns and inform citizens, which is particularly significant in countries with a large population of migrant workers like India and China. The aim is to utilize this information to aid individuals and government organizations in effectively managing the pandemic.

This research proposes an effective method to address the COVID-19 pandemic by predicting the implementation of intelligent lockdowns in over 200 countries using unified ARIMA and SARIMA models. The models were trained on data from 237 countries and had a response time of 2.5 months. The models' reliability was evaluated using analytical methods. They have a rapid response time and predict both confirmed cases and deaths due to COVID-19.This study add a new perspective to time-series analysis and aid decision-making regarding implementation of intelligent lockdowns.

ACKNOWLEDGEMENT

FUNDING

COMPETING INTEREST

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

DATA AVAILABILITY

The COVID-19 data that support the findings of this study are available in/from WHO COVID Repository, https://covid19.who.int/data.

REFERENCES

1. "WHO Coronavirus (COVID-19) Dashboard." 2022.

2. Schwab P, Mehrjou A, Parbhoo S, Celi LA, Hetzel J, Hofer M, et al. Real-time prediction of COVID-19 related mortality using electronic health records. Nat Commun. 2021;12(1):1058.

3. Zhao H, Merchant NN, McNulty A, Radcliff TA, Cote MJ, Fischer RS, et al. COVID-19: Short term prediction model using daily incidence data. PloS one. 2021;16(4):e0250110.

4. Liu M, Thomadsen R, Yao S. Forecasting the spread of COVID-19 under different reopening strategies. Sci. Rep.2020;10(1):1-8.

5. Chordia S, Pawar Y. Analyzing and forecasting COVID-19 outbreak in India. 11th International Conference on Cloud Computing, Data Science & Engineering. 2021.

6. ArunKumar KE, Kalaga DV, Kumar CM, Chilkoor G, Kawaji M, Brenza TM. 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 Auto-Regressive Integrated Moving Average (SARIMA). Appl. Soft Comput. 2021;103:107161.

7. ArunKumar KE, Kalaga DV, Kumar CM, Kawaji M, Brenza TM. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. Alex. Eng. J. 2022;61(10):7585-7603.

8. Sahai AK, Rath N, Sood V, Singh MP. ARIMA modelling & forecasting of COVID-19 in top five affected countries. Diabetes & Metabolic Syndrome: Clin. Res. Rev. 2020;14(5):1419-1427.

9. Young WL. The Box-Jenkins approach to time series analysis and forecasting: principles and applications. RAIRO-Operations Research-Recherche Opérationnelle. 1977;11(2):129-143.

10. Yermal L, Balasubramanian P. Application of auto ARIMA model for forecasting returns on minute wise amalgamated data in NSE. IEEE Int Conf Comput Intell Comput Res. 2017.

11. Newbold P. ARIMA model building and the time series analysis approach to forecasting. J. Forecast. 1983;2(1):23-35.

12. Ong CS, Huang JJ, Tzeng GH. Model identification of ARIMA family using genetic algorithms. Appl Math Comput. 2005;164(3):885-912.

13. Ilie OD, Cojocariu RO, Ciobica A, Timofte SI, Mavroudis I, Doroftei B. Forecasting the spreading of COVID-19 across nine countries from Europe, Asia, and the American continents using the ARIMA models. Microorganisms. 2020;8(8):1158.

14. Peng Z, Khan FU, Khan F, Shaikh PA, Yonghong D, Ullah I, et al. An application of hybrid models for weekly stock market index prediction: Empirical evidence from SAARC countries. Complexity. 2021;2021:1-10.

15. Mondal P, Shit L, Goswami S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. Int J Comput Ssci Eng Appl. 2014;4(2):13.

16. Pai PF, Lin CS. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega. 2005;33(6):497-505.

17. Mehta WA, Sukmawaty Y. Rainfall prediction climatological station of Banjarbaru using arima kalman filter. J Phys: Conference Series. 2021.

18. Xu D, Zhang Q, Ding Y, Zhang D. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ Sci Pollut Res. 2022;29(3):4128-4144.

19. Ediger VS, Akar S. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy policy. 2007;35(3):1701-1708.

20. Nyoni T. Modeling and forecasting inflation in Kenya: Recent insights from ARIMA and GARCH analysis. D ReJ. 2018;5(6):16-40.

21. Fattah J, Ezzine L, Aman Z, El Moussami H, Lachhab A. Forecasting of demand using ARIMA model. Int J Eng Bus Manag. 2018;10:1847979018808673.

22. Osborn DR. Seasonality and the order of integration for consumption. Oxf Bull Econ Stat. 1988;50(4):361-377.

23. Otero J, Smith J. The KPSS test with outliers. Comput Econ. 2005;26:59-67.

24. Oliveira PJ, Steffen JL, Cheung P. Parameter estimation of seasonal ARIMA models for water demand forecasting using the Harmony Search Algorithm. Procedia Eng. 2017;186:177-185.

25. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19(6):716-723.

26. Akaike H. Canonical correlation analysis of time series and the use of an information criterion. Math Sci Eng. 1976.

27. Akaike H. Fitting autoregressive models for prediction. Ann Inst Stat Math. 1969;21(1):243-247.

28. Ayitey E, Kangah J, Twenefour FB. Sarima Modeling of Monthly Temperature in the Northern part of Ghana. Asian J Probab Stat. 2021;12(3):37-45.

29. Ma S, Liu Q, Zhang Y. A prediction method of fire frequency: Based on the optimization of SARIMA model. PLoS one. 2021;16(8):e0255857.

30. Qi C, Zhang D, Zhu Y, Liu L, Li C, Wang Z, et al. SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with SARIMA. BMC Med Res Methodol. 2020;20(1):1-7.

31. Wu DC, Ji L, He K, Tso KF. Forecasting tourist daily arrivals with a hybrid Sarima–Lstm approach. J Hosp Tour Res. 2021;45(1):52-67.

32. Monica M, Suharsono A, Otok BW, Wibisono A. Hybrid SARIMA-FFNN model in forecasting cash outflow and inflow. J Phys: Conference Series. 2021.

33. Martinez EZ, Silva EA, Fabbro AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Rev Soc Bras Med Trop. 2011;44:436-440.

34. Ebhuoma O, Gebreslasie M, Magubane L. A seasonal autoregressive integrated moving average (SARIMA) forecasting model to predict monthly malaria cases in KwaZulu-Natal, South Africa. S Afr Med J. 2018;108(7)..

35. Manigandan P, Alam MS, Alharthi M, Khan U, Alagirisamy K, Pachiyappan D, et al. Forecasting natural gas production and consumption in United States-Evidence from SARIMA and SARIMAX models. Energies. 2021;14(19):6021.

36. Manigandan P, Alam MS, Alharthi M, Khan U, Alagirisamy K, Pachiyappan D, et al. Forecasting natural gas production and consumption in United States-Evidence from SARIMA and SARIMAX models. Energies. 2021;14(19):6021..

37. Samal KK, Babu KS, Das SK, Acharaya A. Time series based air pollution forecasting using SARIMA and prophet model. International Conference on Information Technology and Computer Communications. 2019.

38. Moeeni H, Bonakdari H, Ebtehaj I. Integrated SARIMA with neuro-fuzzy systems and neural networks for monthly inflow prediction. Water Resour Manag. 2017;31:2141-2156.

39. Moeeni H, Bonakdari H, Ebtehaj I. Monthly reservoir inflow forecasting using a new hybrid SARIMA genetic programming approach. J Earth Syst Sci. 2017;126:1-3.

40. Flores JH, Engel PM, Pinto RC. Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting. Int J Conf Neural Netw. 2012.

41. Akaike H. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika. 1973;60(2):255-265.

42. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978:461-464.

43. Woodward WA, Gray HL. On the relationship between the S array and the Box-Jenkins method of ARMA model identification. J Am Stat Assoc. 1981;76(375):579-587.

44. Tsay RS, Tiao GC. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. J Am Stat Assoc. 1984;79(385):84-96.

45. Tsay RS, Tiao GC. Use of canonical analysis in time series model identification. Biometrika. 1985;72(2):299-315.

46. Hannan EJ, Quinn BG. The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological). 1979;41(2):190-195.

47. Hannan EJ, Rissanen J. Recursive estimation of mixed autoregressive-moving average order. Biometrika. 1982;69(1):81-94.

48. Chan WS. A comparison of some of pattern identification methods for order determination of mixed ARMA models. Stat Probab Lett. 1999;42(1):69-79.

49. Shibata R. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika. 1976;63(1):117-126.