Forecasting Groundwater Production and Rain Amounts Using ARIMA-Hybrid ARIMA:
Case Study of Deir El-Balah City in GAZA

Ihsan A. S. Abu Amra
Islamic University of Gaza,
Gaza, Palestine
ihsanamra2015@gmail.com

Ashraf Y. A. Maghari
Islamic University of Gaza
Gaza, Palestine
amaghari@iugaza.edu.ps

***Abstract*— Forecasting as a data mining technique offers the opportunity to leverage the numerous sources of time-series data, to support business decision-makers for effective planning and derive value from historical data. The lack and fluctuating rainfall and the Israeli occupation caused a crisis in Palestine and in Gaza in particular. Water resources are decreasing by the increase of population as the passage of time. Groundwater is the only water resource used in Gaza. Moreover, the increase in the demand for groundwater and the decrease in rainfall water, which is the main source of groundwater, will lead to depletion of groundwater wells, thus increase the salinity rate. In this paper, we conducted the forecasting technique on both the groundwater and the rain amounts in Dear El-Balah city of Gaza using the following forecasting algorithms: Auto-Regressive Integrated Moving Average (ARIMA), Hybrid ARIMA:(ARIMA+ NN), (ARIMA+ ets), (ARIMA+ tbats). The best performance of applied algorithms on rainfall data according to Mean Absolute Percentage Error (MAPE) measure was ARIMA combined with NN which gave the MAPE = 21%. On the other hand, ARIMA combined with tbats was the best algorithm applied on wells production data which achieved MAPE= 8.9%. The results showed that after 5 years the amounts of rainfall and groundwater production in comparison with the period from (2013 to 2017) will decrease by 8.4%, 0.03%, respectively.**

Keywords— Groundwater; Rain; Artificial Neural Networks; ARIMA; Hybrid ARIMA;

INTRODUCTION

Water is considered one of the scarce resources in Palestine. An important reason for this is the restriction imposed by the Israeli occupation on the Palestinian institutions in the field of the extraction and use of water. Furthermore, the economic situation in Gaza plays a vital role because of the shortage of the financial resources for investing in the water sector resulting in high rates of poverty [1].

Water resources in Palestine include runoffs, surface water; and groundwater. While the share of the Palestinians is around 11%, the remaining 89% is exploited by Israel [2]. According to the latest population statistics in Gaza, the number of citizens in Gaza until the beginning of April 2018 is 2090405, including the central governorate with a total of 294531 citizens [3]. By the 1970s Israel was physically controlling almost all available sources of fresh water from the Jordan River and its catchment (including the Golan Heights annexed from Syria) as well as the underground aquifer water [4].

The water situation in Gaza is facing critical condition, according to last reports and national plans of Palestinian Water Authority[5]. The only source of water is the ground aquifer, where the water level is decreasing, with increase in water demand for different uses. The Coastal Aquifer receives an annual average recharge of 55 -60 MCM/y mainly from rainfall, while the annual extraction rates from the aquifer is about 200 MCM. These unsustainable high rates of extraction have led to lowering the groundwater level, the gradual intrusion of seawater and up conning of the underneath saline groundwater. All these conditions made the aquifer unusable in 2016, and the damage to it will be irreversible by 2022.

The Gaza aquifer represents the sole water source of Gaza. It is threatened by seawater and salt groundwater intrusion due to over pumping[6]. Rainwater is the only renewable source that feeds the Coastal Aquifer in Gaza, especially because of the increased reliance on the amount of groundwater by municipalities. The maintenance of water levels in the wells is very important taking into account the amounts of rainfall falling to maintain Salinity and other minerals in groundwater [5].

The increasing consumption of groundwater and decreasing rainfall water, will lead to depletion of groundwater wells, thus increase the salinity rate.

Gaza is one of the scarcest water resources areas in the region. Groundwater is the only water resource of water for the Palestinians in Gaza and provides more than 90% of all water supplies[7]. It suffers from rising deficit in the water budget because of continuous increase in water demand for different uses which has been leading to fall in the quality and quantity of groundwater.

Data mining includes extracting information from large data sets. Data mining leads to either a descriptive model or a predictive model. The Descriptive model includes tasks such as Clustering, Association Rules, and Summarizations. The predictive model uses the strategy of predicting the data values taken from various datasets. This model includes classification, prediction, regression and analysis of time series[8].

The idea of forecasting is to use historical data to make forecasts and prediction about future data. Forecasting is one of the most common techniques of time series data analysis. It's used to predict future trends in water demands, retail sales, weather forecasting, and many other application scenarios[9]. Reliable forecast of groundwater consumption is necessary for proper planning of groundwater demand and water management resources. Forecasts can be classified, as suggested by [10], into: (1) Long-Term which forecasts decades, (2) Medium-Term which forecasts years to decade, (3) Short-Term which forecasts up to year.

From previous introduction, population in Gaza in specific is growing up by time, and water resources are limited due to low rainfall annual amounts or by Israeli politics against Palestinian Authority about water underground resources. So, handling these factors by using forecasting science in future water service projects is very important issue for mayors of municipalities and decision makers, by Routing the annual consumption of groundwater with planning between the salinity of groundwater and rainfall rates, which are the main source of groundwater recharge and thus reduce the salinity ratio, managing the consumption of groundwater through the coordination of the digging of wells by the citizens.

Coastal Municipalities Water Utility (CMWU) has time series data for groundwater wells in Dear El Balah and the amount of monthly production over years, as well as the Ministry of Agriculture[11] have a large time series data belongs to all its services on the city. Hundreds of records and data are collected. So, these data scientifically should be used in planning for determining the relation between groundwater and rainfall in the future.

In this paper, we collected two different historical series data, the first for groundwater wells from (CMWU) in Dear El-Balah (10 years: from 2008 to 2017) and the second: historical rainfall amounts since 1985 (33 years) from the Palestinian Ministry of agriculture. The data prepared for applying forecasting algorithms to choose the most accurate algorithm according to 'MAPE' for the prediction of the next five years. Hence, useful results will be presented to the managers and specialists to have a clear perspective view of the water consumption in the near future.

The rest of this paper is organized as follows: Section 2 reviews some related works. Section 3 shows experiments and results. Section 4 presents conclusion and future work.

RELATED WORK

Some researchers [12], proposed a neural network-based approach to automatically extract features from the time series measured at observation sites. They predicted short term rainfall amount based on multisite features.

Their proposed model outperformed a set of baseline models. The proposed approach demonstrated its effectiveness. While other researchers [13] focused on forecasting of long-term seasonal spring rainfall in Victoria. Artificial Neural Network (ANN) and Multiple Regression (MR) approach was used for this purpose.

Hassani et.al [14] presented the forecasting comparison among several non-parametric and parametric techniques e.g. the ARIMA, ETS,NN, TBATS, ARFIMA, MA, WMA, SSA-R and SSA-V. They used TBATS and SSA-R models for tourist arrival forecasting purposes. The results suggested that there is not a model that its forecasting accuracy consistently outperforms that of all other models for any of the countries under any of the forecasting horizons and investigation.

Mohanty et. al [16] applied artificial neural network (ANN) approach to the weekly forecasting of groundwater levels in multiple wells located over a river basin. They applied different algorithms to predict groundwater levels at 18 sites over the study area. The model performance was better while forecasting groundwater levels at shorter lead times than that for larger lead times. While Dhekale and Sahu [17] aimed at forecasting groundwater fluctuations using time series analysis groundwater data. The time series water data collected for nine years from the period 2005 to 2013. They foresee behavior of ground water in 2014. Data from 2005 to 2012 was used for analysis and training, while 2013 data used for validation and evaluation. The results achieved that there are differences of groundwater depth among the sites and seasons.

Mukhairez[18] applied four forecasting algorithms: ARIMA, hybrid ARIMA, singular spectrum analysis(SSA) and linear regression on a real dataset collected from the municipality in Khan Younis city. The best algorithm was Hybrid ARIMA which gave the least (MAPE). The results after applying Hybrid ARIMA compared to 2017 for the next five years was the minimum water revenue will decrease about 3.8%, but the minimum water consumption for the overall city will increase to about 8.4%.

Thiyagarajan et. al [19] proposed ARIMA model for forecasting the failure of a sensor that measures surface temperature from an urban sewer. The proposed approach was examined and compared with ETS and TBATS model. The prediction performance of TBATS model was better than the ETS model. Also, the prediction performance of ARIMA model was better than the ETS and TBATS model.

Pati and Shukla [20] had experimentally verified the predictive performance of three models: ANN, ARIMA and the Hybrid Model (ARIMA + ANN).

Garima and Mallick[21] used ETS (Exponential Smoothing) and ARIMA (Autoregressive Integrated Moving Average) for analysis and predicting of weather parameters. The accuracy was estimated by different criteria such as: MAPE and RMSE by using different packages in R.

EXPERIMENTS AND RESULTS

The proposed model is about forecasting for both groundwater production and rain amounts in Dear El-Balah city of Gaza. The selected algorithms are the most recent algorithms and famous in forecasting technique. Our methodology in this research consists of the following five phases:

A. *Data Collection:*

Historical data was collected from different Institutions: (the Coastal Municipalities Water Utility (CMWU) And the Ministry of Agriculture) in Dear El-Balah city. The collected data from CMWU were for monthly groundwater production since 2007 and from the Ministry were rainfall amounts since

1985. The groundwater data contains information for 12 wells belongs to Deir El- Balah city in the middle area of Gaza. We selected the related tables and columns that belongs to our study and prepare it for the next phase.

B. Data Preprocessing:

At this essential step, required preprocessing tasks were applied to enhance data efficiency before applying the forecasting algorithms. Preprocessing includes several techniques e.g. cleaning, reduction, transformation and integration. We used Microsoft Excel 2016 version to perform the following data preprocessing:

1. *Data integration:*

Data integration means: combining data from multiple datastores into one consistent dataset. We gained our datasets from: the ministry of agriculture about yearly rain amounts of Dear El-Balah since 1985 and groundwater production amounts from (CMWU), so we merged them through the common ID and obtain two data sets for 12 groundwater wells belonging to the (CMWU), and rain amounts of Dear El Balah.

2. *Data reduction:*

In this step we reduced the representation of dataset in smaller volumes, we removed irrelevant attributes. The result of this step is two data sets on the form of time series for groundwater and rainfall. TABLE I represents a sample of groundwater dataset with two columns, the first column of time with the format (mm-yy), and the second column "amount(m3)".

TABLE I. SAMPLE OF GROUNDWATER TIME SERIES DATA SET

Month	Amount (m3)
Jan-08	339957
Feb-08	371711
Mar-08	396483
Apr-08	398139
May-08	398812
Dec-08	317989
Jan-09	299446
Feb-09	319041
Mar-09	372687
Jul-09	399822
Aug-09	423123

TABLE II represents rainfall data set of two columns: the year and the amount of rainfall.

TABLE II. SAMPLE OF RAIN TIME SERIES DATASET

Year	Amount (mm)
1985	108
1986	626.4
1987	130.2
1992	419.3
1993	209.6
1994	683
2002	435.7

3. *Missing values:*

We filled the missing value of (JUN 2013) in a groundwater dataset with the average values of the same month from the three previous years respectively: (JUN 2010, JUN 2011, and JUN 2012) as shown in TABLE III to ensure closing the missing monthly amount to the real value. Also, the monthly amount of rain in June is similar to the same month in previous years, because of the absence of rain in other months.

TABLE III. FILLING MISSING VALUE IN JUNE2013

Month	Amount (m3)
Feb-13	331622
May-13	417558
Jun-13	436912
Jul-13	474883
Aug-13	513654

C. Implementation

After preprocessing and preparing the real datasets of Deir El-Balah city on the form of time series data, we gained two datasets the first: monthly wells production from January-2008 to December 2017. It consists of two columns (Month, Amount (m^3)) as in Table1. The second data set of the annual rainfall amounts from 1985 to 2017. It consists of two columns (Year, Amount (mm)) as in TABLE II. Four types of forecasting algorithms were run on these datasets. These algorithms are: 1- ARIMA, 2-ARIMA +NN, 3- ARIMA+ETS, 4-ARIMA+ tbats. According to their results, the appropriate algorithm with the lowest computed MAPE value will be selected for real forecasting from 2018 to 2022.

- *Steps of applying models:*

1. We divided the original wells data as in TABLE IV into training set from January-2008 to December 2015 (80 % of the original data) and testing set starts from January-2016 to December- 2017 (20% of the original data). Also, we divided the rain data into two parts, training and testing sets: The training set starts from 1985 to 2010 (78.7 % of the original data) and testing set starts from 2011 to 2017 (21.2% of the original data). The periods of each splitted data was declared also in TABLE IV.

TABLE IV. SPLITTING THE ORIGINAL DATA INTO TRAINING AND TESTING

Data sets	Training set	Testing set
Well data	Jan-2008 to Dec-2015 (96 months)	Jan -2016 to Dec-2017 (24 months)
Rain data	1985 to 2010 (26 years)	2011to 2017 (7 years)

2. We ran the four algorithms on the training set of wells with horizon value (24) to reach December 2017(the last record in testing set).
3. The same algorithms have been run on the training set of rain data with horizon value (7) to reach 2017(the last record in rain testing set). The algorithms have been evaluated by comparing the results of each algorithm with the original data of rain and groundwater separately.

D. Evaluation the algorithms

First, Evaluation process of the performance for the four selected algorithms done over the amount attribute for groundwater data set, by MAPE measure. We computed the MAPE value between the predicted results and actual values in testing set according to the following equation:

$$\mathbf{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Actual_{i-Predicted_i}}{Actual_i}\right| 100\%$$

Then, we compared the MAPE values between forecasting algorithms. The algorithm with the lowest MAPE value will be the most accurate algorithm and will be a candidate for the actual forecast task.

The computed MAPE value for each algorithm is shown in Table V. It is observed that the hybrid ARIMA (ARIMA + tbats) was the best forecasting algorithm to be used over groundwater data set.

TABLE V. COMPUTED MAPE FOR THE ALGORITHMS OVER GROUNDWATER DATA

Algorithm	MAPE%
ARIMA	**25.9**
Hybrid ARIMA	**11.4**
ARIMA+ ets	**9.1**
ARIMA+ tbats	**8.9**

Second, we evaluated in the same way the performance for the selected five algorithms done on the rain amounts attribute over the rain dataset. By computing the MAPE for each algorithm. It is observed as shown in TABLE VI that the hybrid ARIMA (ARIMA +NN) was the best forecasting algorithm for real forecast over rain data set

TABLE VI. COMPUTED MAPE FOR THE ALGORITHMS OVER RAINFALL DATA

Algorithm	MAPE%
ARIMA	22.7
ARIMA+NN	21.0
ARIMA+ ets	23.8
ARIMA+ tbats	21.2

E. Forecasting

After determining the most accurate algorithm for rain and groundwater prediction, we forecasted five years in advance for rain amounts and groundwater production separately. Then we computed the deviation percentage for every forecasted year and compared it with year 2017.

1. Forecasting rain amounts

Figure 1 shows the results of forecasting the rain amounts from year 2018 to 2022.

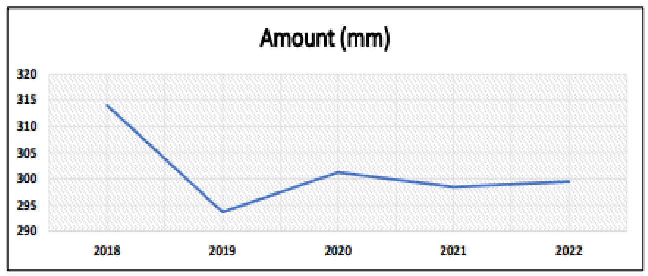

Fig. 1. Forecasting rain amounts from 2018 to 2022

2. Forecasting wells production amounts of groundwater

The results of forecasting the groundwater production amounts is plotted in Figure2. The rain amounts from 1985 to 2022 and wells production amounts from 2008 to 2022 is represented in Figure3 and Figure4, respectively.

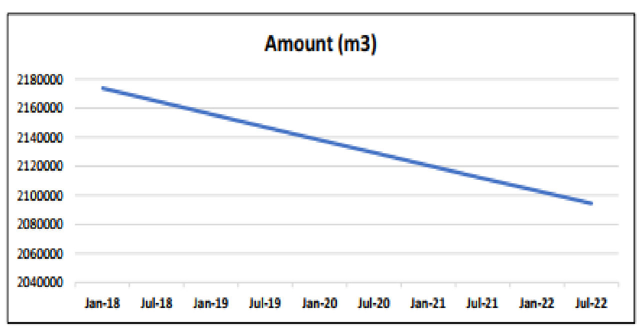

Fig. 2. Forecasting Groundwater production from 2018 to 2022

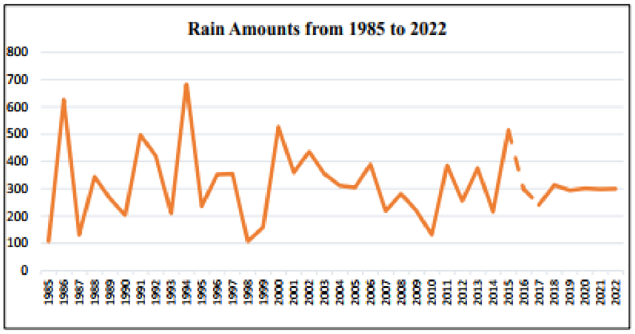

Fig. 3. Rain amounts from 1985 to 2022

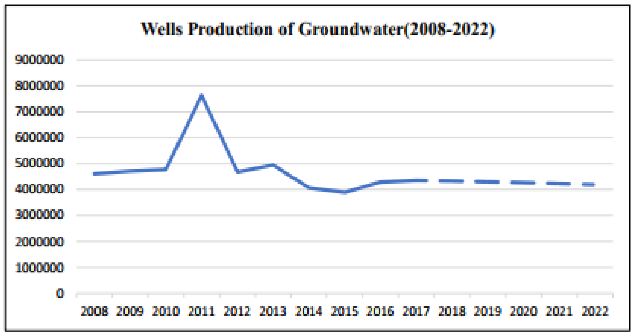

Fig. 4. Ground water production from 2008 to 2022

3. *Results Deviation*

To give strong future vision for decision makers about the amounts of rainfall and groundwater production for Deir El-Balah city, we compared the last year of the original dataset (2017) with each forecasted year from (2018-2022) for both 'rain amounts' and 'wells production'.

TABLE VII. DEVIATION IN COMPARISON TO 2017.

Year	Rain amounts (mm)	Rain amounts Deviation %	wells production (m^3)	Wells production Deviation%
2017	314	0	4358586	0
2018	294	+30	4283491	-2
2019	301	+22	4303291	-1
2020	298	+25	4314447	-1
2021	299	+24	4324261	-1
2022	314	+24	4332709	-1

In TABLE VII the forecasted rain amounts comparing to 2017 will increase with rate range between 22% and 30% from 2018 to 2022. On the other hand, groundwater production will maximally decrease by 2% in comparison to 2017.

TABLE VIII. DEVIATION OF RAIN FORECASTED YEARS COMPARED TO LAST 5 YEARS

Year	Rain amounts(mm)	Groundwater amount (m3)
2013-2017	1645	21563737
2018-2022	1507	21558198
Deviation%	-8.4	-0.03

Generally, we can say that after 5 years the rainfall amounts will decrease with 8.4% in comparison with rain amounts in the period from (2013 to 2017) but the wells production of groundwater after 5 years will be decreased with (0.03%) in comparison with the period from (2013 to 2017) of groundwater production as shown in TABLE VIII.

CONCLUSION AND FUTURE WORK

In this paper we presented medium term forecasting for both groundwater production and rainfall amounts of Dear El-Balah city in Gaza. This is very useful to leverage the numerous sources of time-series data, to support business decision-makers, for effective planning. We focused on the relationship between rainfall - which feeds the groundwater resovir and reduces its salinity - and the percentages of the demand for the groundwater. Lack and fluctuating rainfall and the Israeli occupation caused a crisis in Palestine and in Gaza in particular. Moreover, the increase in the demand for groundwater (the only water resource in Gaza) and the decrease in rainfall water, which is a main source of groundwater, will lead to depletion of groundwater wells, thus increase the salinity rate. We conducted four forecasting algorithms: Auto-Regressive Integrated Moving Average (ARIMA), Hybrid ARIMA (ARIMA+ NN), (ARIMA+ ets), and (ARIMA+ tbats) on two real data sets for groundwater production amounts and the rain amounts. We found that the more accurate algorithm was applied for rain amounts forecasting is (ARIMA combined with NN) which gave (MAPE) of 21.0%, while, (ARIMA combined with tbats) was the best algorithm for groundwater forecasting which achieved (MAPE) 8.9%.

Finally, the results showed that after 5 years the rainfall amounts will decrease with 8.4% in comparison with rain amounts in the period from (2013 to 2017) but the wells production of groundwater will decrease with (0.03%) in comparison with the same period.

This is a dangerous indicator and needs decision makers to take in control these results in holding water projects or while demanding water from groundwater wells. As future work, other forecasting algorithms can be applied for medium term forecasting and may include more factors like weather conditions or temperature degrees.

REFERENCES

[1] P. W. Authority, “Water Authority Strategic Plan 2016-2018 1,” pp. 1–48, 2018.

[2] R. Barakat and R. Heackock, *Water in Palestine ISBN 978-9950-316-47-8*. 2013.

[3] Ministry of Interior and National Security, “No Title,” 2018.

[4] Z. Al-Shalalfeh, F. Napier, and E. Scandrett, “Water Nakba in Palestine: Sustainable Development Goal 6 versus Israeli hydro-hegemony,” *Local Environ.*, vol. 23, no. 1, pp. 117–124, 2018.

[5] M. Aiash and Y. Mogheir, “Comprehensive Solutions for the Water Crisis in Gaza Strip,” vol. 25, no. 3, pp. 63–75, 2017.

[6] F. Daibes-Murad1, “Water Resources in Palestine A Fact Sheet and Basic Analysis of the Legal Status,” 2004.

[7] S. Report and O. F. Water, “in the Occupied State of Palestine-2012,” no. October, p. 22, 2013.

[8] Deepashri and A. Kamath, “Survey on Techniques of Data Mining and its Applications,” *Int. J. Emerg. Res. Manag. &Technology*, vol. ISSN, no. 62, pp. 2278–9359, 2017.

[9] C. C. Aggarwal, *Data Mining The Textbook*. 2015.

[10] R. B. B. V. Jones, *Forecasting Urban Water Demand*, Second edi. America: american water works association, 2011.

[11] M. of A. S. of Palestine, “No Title,” 2017.

[12] M. Qiu *et al.*, “A short-term rainfall prediction model using multi-task convolutional neural networks,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2017–Novem, pp. 395–404, 2017.

[13] F. Mekanik, M. A. Imteaz, S. Gato-Trinidad, and A. Elmahdi, “Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes,” *J. Hydrol.*, vol. 503, pp. 11–21, 2013.

[14] H. Hassani, E. S. Silva, N. Antonakakis, G. Filis, and R. Gupta, “Forecasting accuracy evaluation of tourist arrivals,” *Ann. Tour. Res.*, vol. 63, pp. 112–127, 2017.

[15] M. Bakker, H. Van Duist, K. Van Schagen, J. Vreeburg, and L. Rietveld, “Improving the performance of water demand forecasting models by using weather input,” *Procedia Eng.*, vol. 70, pp. 93–102, 2014.

[16] S. Mohanty, M. K. Jha, S. K. Raul, R. K. Panda, and K. P. Sudheer, “Using Artificial Neural Network Approach for Simultaneous Forecasting of Weekly Groundwater Levels at Multiple Sites,” *Water Resour. Manag.*, vol. 29, no. 15, pp. 5521–5532, 2015.

[17] B. S. Dhekale, P. K. Sahu, K. P. Vishwajith, and L. Narsimhaiah, “Structural Time Series Analysis towards Modeling and Forecasting of Groundwater Fluctuations in Murshidabad District of West Bengal,” vol. 5, pp. 117–126, 2015.

[18] D. A. E.-H. Hosam Mukhairez, “Medium-Term Forecasting for Municipal Water Demand and Revenue (KhanYounis City as A Case Study) Medium-Term Forecasting for Municipal Water Demand and Revenue (KhanYoun.”

[19] K. Thiyagarajan, S. Kodagoda, and L. V. Nguyen, “Predictive Analytics for Detecting Sensor Failure Using Autoregressive Integrated Moving Average Model,” *12th IEEE Conf. Ind. Electron. Appl.*, pp. 1923–1928, 2017.

[20] J. Pati and K. K. Shukla, “A comparison of ARIMA, neural network and a hybrid technique for Debian bug number prediction,” *Proc. - 5th IEEE Int. Conf. Comput. Commun. Technol. ICCCT 2014*, pp. 47–53, 2015.

[21] E. Garima Jain and B. Mallick, “A Study of Time Series Models ARIMA and ETS,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 4, pp. 57–63, 2017.