International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 6 Issue 9, September - 2022, Pages: 27-38

Forecating Water Demand with ARFIMA models: A case study of Dhofar Governorate, Oman

Oladipupo Bello

Polytechnic Programmes Department, National Board for Technical Education, Kaduna, Nigeria
engroobello@gmail.com

Abstract— *Water demand models are indispensable tools in the planning, design and operations of water supply systems. These models are developed based on historical water consumption data which usually have trends and seasonality. The autoregressive fractionally integrated moving average (ARFIMA) process is used in this study to evaluate the performance of univariate water demand models. Several ARFIMA models were fitted to monthly water consumption data of Dhofar Governorate, Oman from January 2002 to October 2020. The best fitted ARFIMA model was selected based on Akaike Information Criterion, and compared with autoregressive integrated moving average (ARIMA) model. The simulation study showed that ARFIMA model performed better than the ARIMA model. The water demand forecast values for two (2) years was obtained for the Dhofar Governorate using the best fitted ARFIMA model. The decision makers in the water supply industry will find the results of the study useful to evaluate investment projects to expand their facilities and determine optimum water tariff for their customers.*

Keywords—ARFIMA, Akaike Information Criterion, Forecast, Stationarity, Water Demand

1. INTRODUCTION

Human lives whether in rural or urban areas depend on availability of clean water supplies. It is indispensable and essential for human survival, industrialization and rapid development. Therefore, provision of potable water for human consumption is one of the key services of Government needed to create healthy and safe society, and improve quality of human life [1]. In order to achieve orderly communal development, quantities of water required must be determined ahead and supply must be planned to meet the demand of end-users living and/or working in a particular geographical area or cluster of areas. Water demand forecasts are therefore essential aspect of cost optimization and sustainable management of water distribution systems [2].

Water demand forecast is applied to predict the future water consumption so as to achieve the objective of water distribution systems. It provides useful information to managers to make operational, tactical and strategic decisions regarding the operations and management of the water distribution systems. It drives sustainability and conservation of water resources. Water forecasting is very important in the development of strategies to save energy, cost and water resources [3]. Furthermore, accurate prediction of water demand can aid online detection of cyber-physical attacks, water theft, fraud and smart meter faults [4].

Water demand forecasting is a complex and dynamic process and thus requires mathematical modelling. Application of time series modelling techniques is a common approach to water demand forecast modelling because they are simple and direct methods that predict future water demand based on only past historical data without considering complex and uncertain variables [2,3].

The purpose of this study is to propose water demand forecast models based on time series data and identify the best fitted model among them. ARFIMA models are considered due to their abilities to model long run persistence or long memory processes. It has been used extensively to solve economic, climatic, network traffic, air quality, global carbon dioxide emission, and hydrological forecasting problems [5] –[10].

To the authors' knowledge, there are limited studies published in the literature that examined the application of ARFIMA process to water demand modeling. The significance of this study is the determination of the performance of water demand models based on ARFIMA modeling technique in comparison to ARIMA model. This study is made up of four sections. Section two presents a brief overview of Dhofar Governorate in Oman, ARFIMA process and model evaluation index. Section three discusses the simulation results. Finally, the last section presents the concluding remarks on the study.

2. MATERIALS AND METHODS

2.1 Research Design

The study is designed to examine and investigate the ability and performance of ARFIMA models to predict monthly water consumption for a geographical area. In the study, an in-depth descriptive analysis to detect the presence of stationarity in the data set was carried out before the data was employed for modeling and forecasting. After the data was transformed, it was divided into two: training and testing datasets. A set of ARFIMA models was trained with the training data set. After training, the best fitted model is then validated using the testing data set. The performance of the best fitted ARFIMA model is compared with ARIMA model.

2.2 Modelling Dataset

Dhofar is one the eleven governorates of Oman as shown in Figure 1. It lies in the southern part of Oman. It has a population estimated to be more than 420,000 persons. It is a mountainous area of 99,300 Km^2. It is known to be one of the key sources of frankincense in the world and exports most important and finest frankincense. It is the largest city in Oman in terms of land area. Dhofar governorate consists of ten provinces. Groundwater and desalination plants are key sources of water supply in Dhofar.

Fig. 1 Map of Sultanate of Oman (Source: www.mapsofworld.com)

Dhofar has a tropical climate. The temperature remains steady between upper 20ºC and mid 30ºC. The climate in Dhofar is different from the climate in the other parts of the Country due to annual monsoon winds from the Indian Ocean. During the Khareef season, between June and September, Dhofar is saturated with cool moisture and heavy fog. Winter temperatures are mild and pleasant ranging between 15ºC and 23ºC [11]. Dhofar has attractive beaches and natural reserves that attract large number of tourists to the Governorate for recreation, vacation and other related activities. The climatic condition in Dhofar is known for high temperature, low rainfall and precipitation.

Table 1 shows the descriptive statistics of the monthly water consumption dataset of Dhofar Governorate. The data used for the modeling are monthly water consumption measurements in cubic metres. It has 226 data points each, starting from January 2002 to October 2020. The data was obtained from World Bank Database [12]. The time series and box plots of the dataset are shown in Figure 2.

Table 1 Summary of the descriptive statistics of the monthly water consumption dataset

Description of Item	**Values**
Number of data points	226
Mean	2684477
Minimum	915285
Maximum	5615067
Standard Deviation	1336429
Skewness	0.6117845
Kurtosis	2.201432

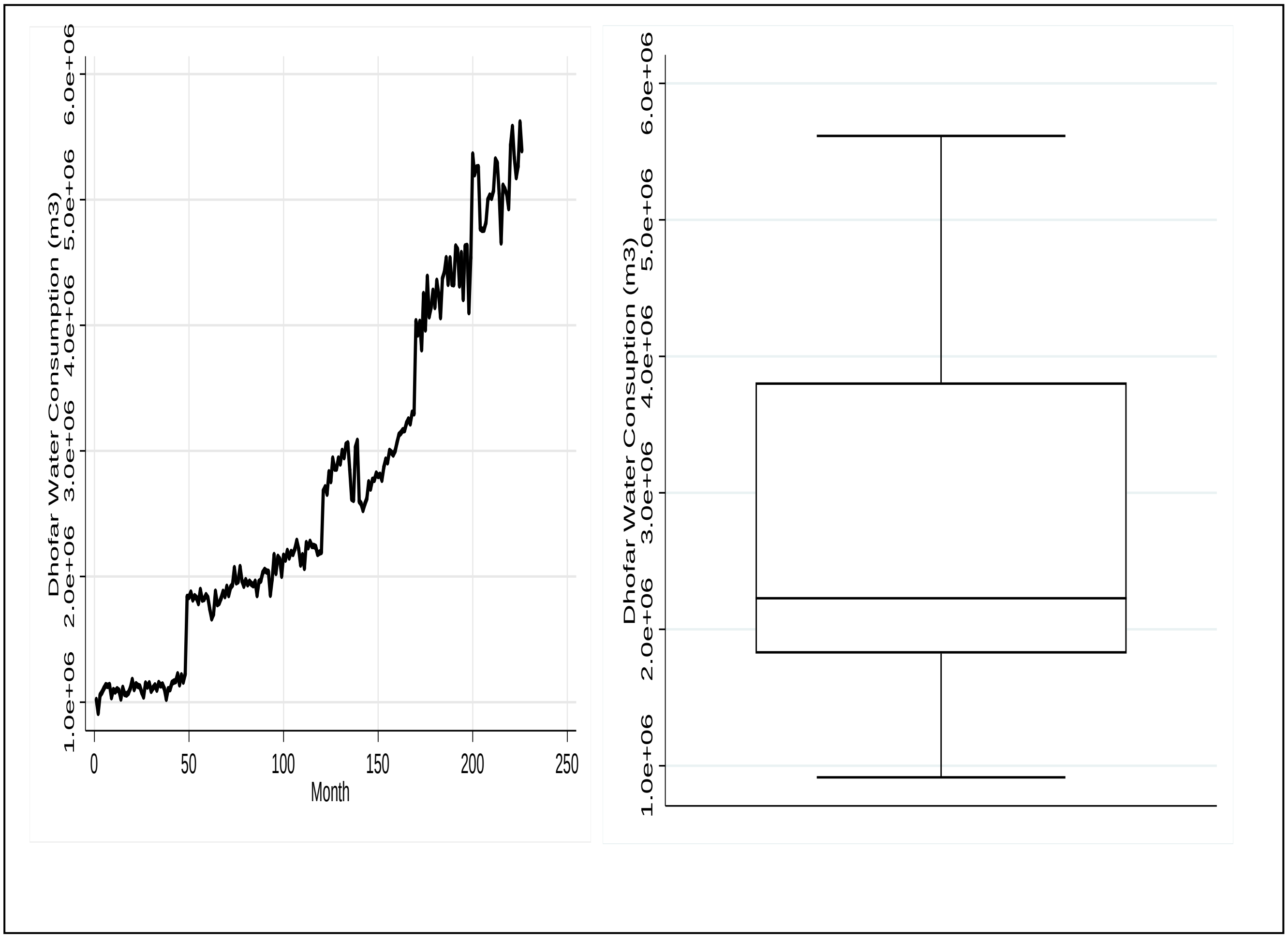


Fig. 2 Time series and box plot of water consumption in Dhofar Governorate

2.3 Autocorrelation and Partial correlation functions

The study examines the presence of trends and seasonality in the monthly water consumption data using the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. In Figures 3 and 4, the ACF and PACF plots showed that there is presence of trends and seasonality in the data which made it non-stationary.

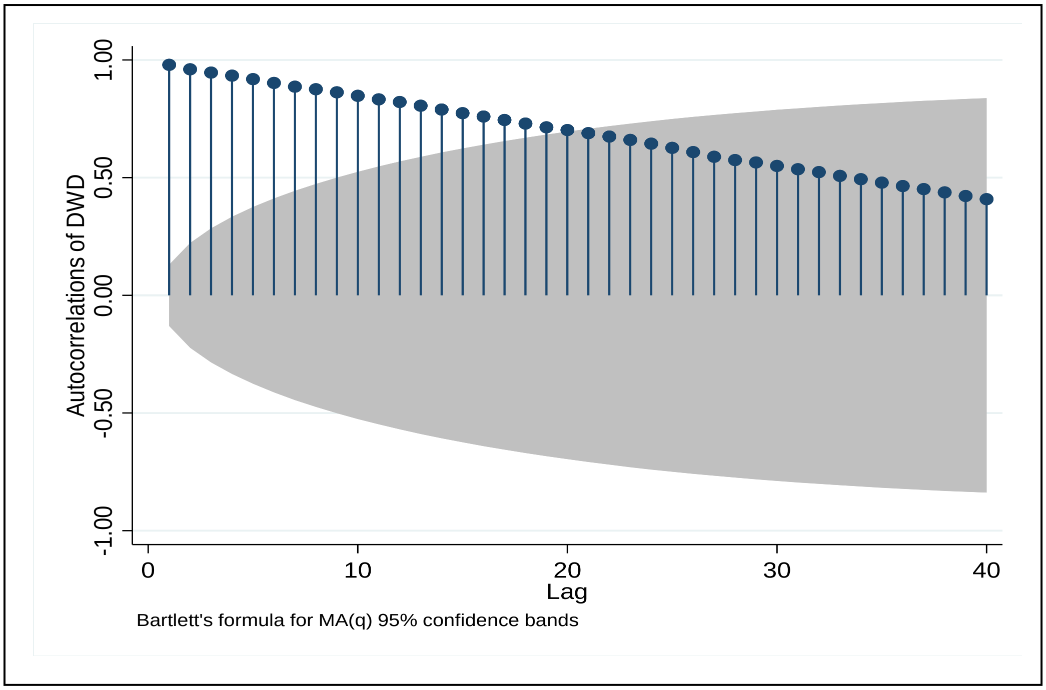


Fig 3 ACF plot of residential water consumption in Dhofar Governorate

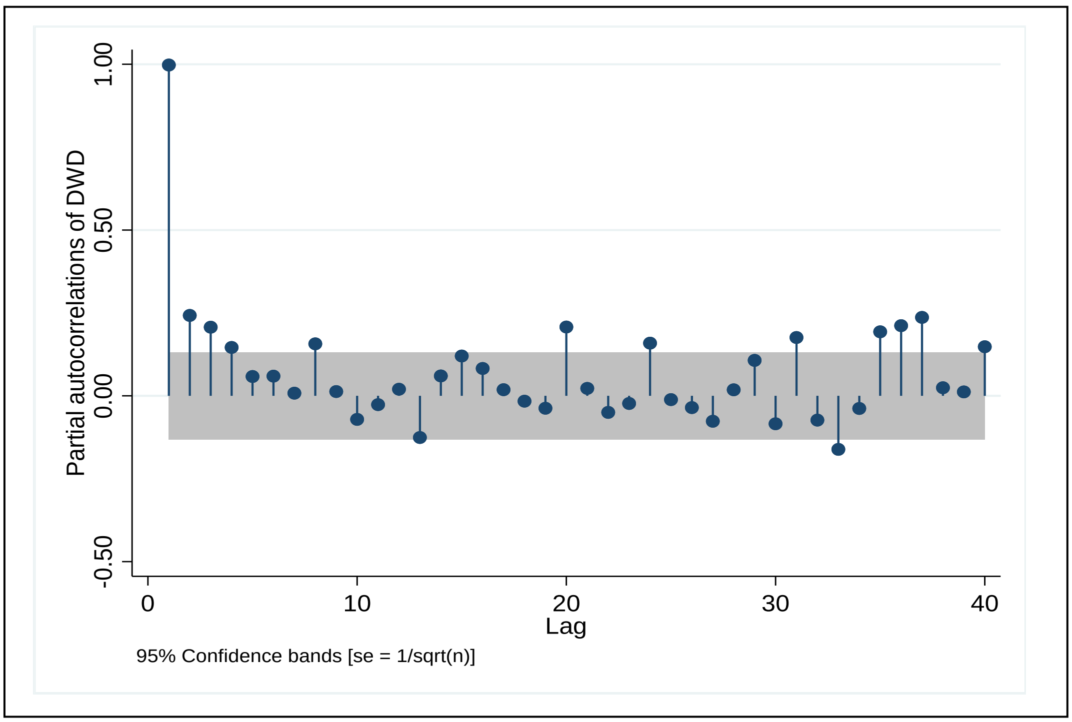


Fig 4 PACF plot of monthly water consumption in Dhofar Governorate

2.4 Distribution Pattern Analysis

Further analysis of the distribution pattern of the monthly water consumption of the Dhofar Governorate in Figure 5 showed that the variable is not normally distributed. There is a need to transform it in order to produce a new dataset that fit into the assumption of stationarity.

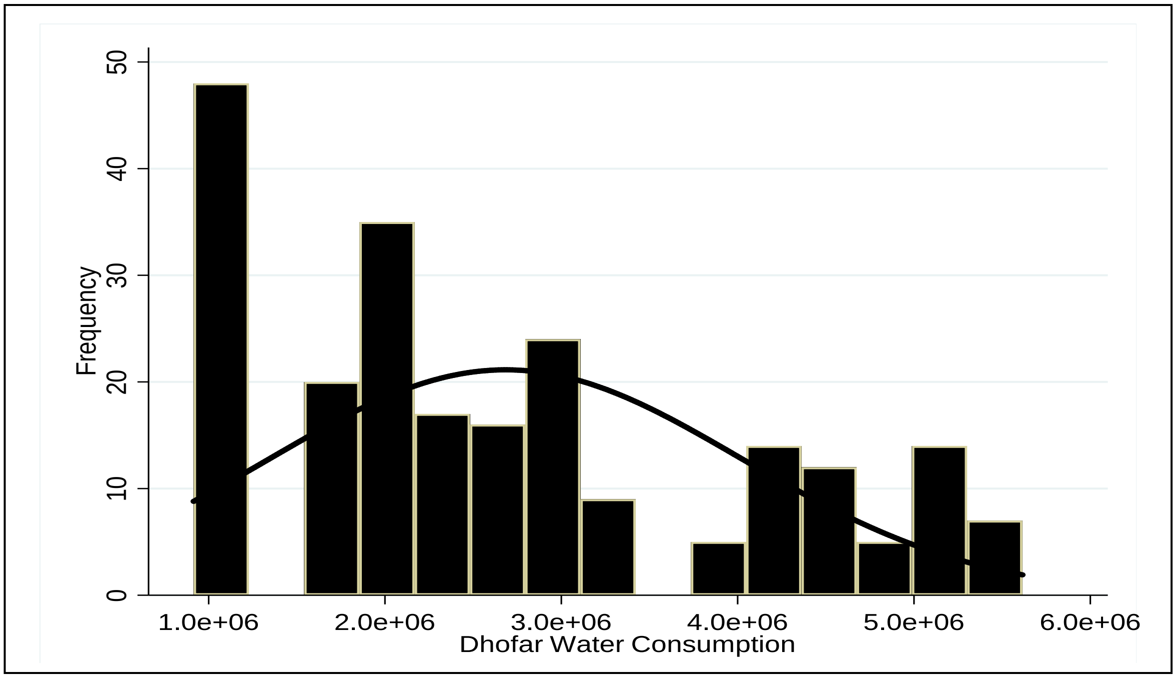


Fig. 5 Histogram Plot/Distribution of Water Consumption at Dhofar Governorate

The water consumption data was pre-processed to remove the trends and seasonality. The variable was transformed using the logarithm function and the seasonality was removed using time-series operator, seasonal differencing function. The transformation of the variable results to a normal distribution as shown in Figure 6. Consequently, the seasonal differenced of the logarithm of the variable were used as the variable for modeling and forecasting water demand for Dhofar Governorate.

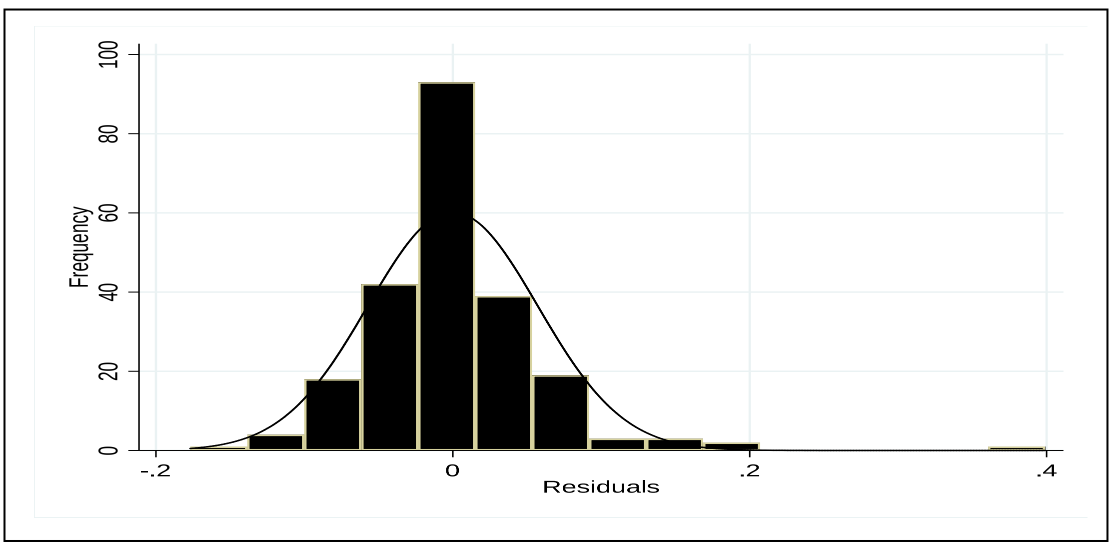


Fig 6 Histogram of Residuals (residuals are estimated by ordinary least square regression)

Figure 7 shows the quantile-quantile (Q-Q) plot of the residuals of the regression analysis of the transformed water consumption data of the Dhofar Governorate. The Q-Q plot is thin-tailed distribution with less deviation at the tail ends and thus it represents for normal distribution and fit for modelling.

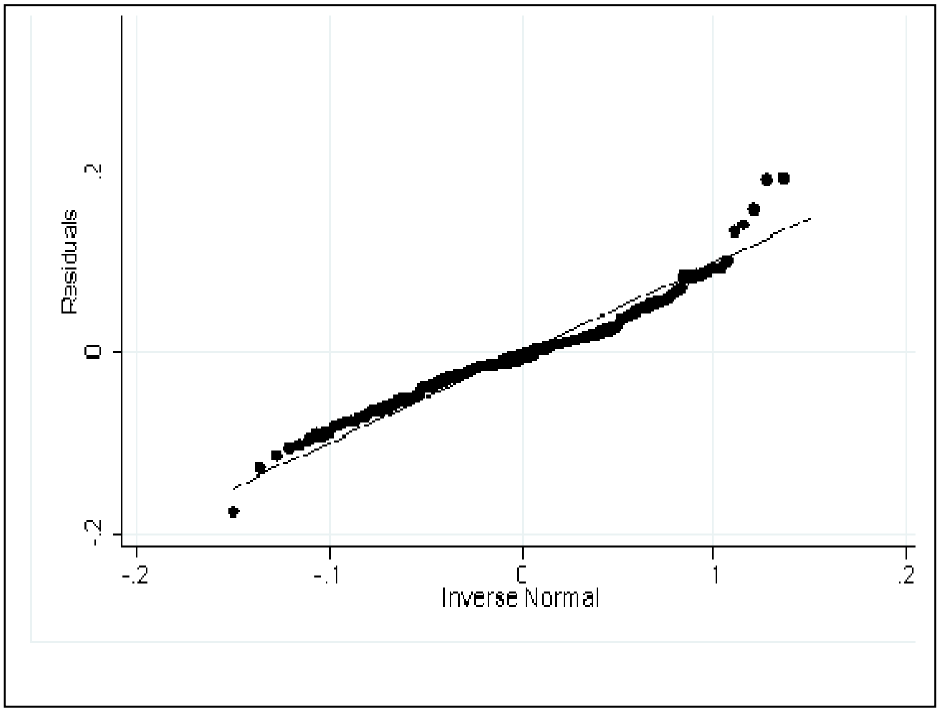


Fig 7 Q-Q plot of residuals of the water consumption data

2.5 ARFIMA model

An autoregressive model is a representation of a type of random or time varying process. The autoregressive model $AR(p)$ of order p can be expressed as:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t, \tag{1}$$

Where $\phi_1, \ldots, \phi_p$ are autoregressive parameters, c is a constant, and the random variable ε_t is the white noise. For the model to be stationary, some constraints are imposed on the values of the parameters [13].

Also, the moving average smoothens the time series that produces cyclic and trend-like plots even the original data are themselves independent random events with fixed mean. The moving average model $MA(q)$ refers to the moving average model of order q :

$$X_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{2}$$

Where $\theta_1, \ldots, \theta_q$ are the moving average parameters of the model, μ is the expectation of X_t (often assumed to equal 0), and $\varepsilon_t, \varepsilon_{t-1}, \ldots$ are again, white noise error terms.

The AR and MA models combined to form $ARIMA(p,d,q)$ expressed above and can be generalized as follows:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1-B)^d (X_t - \mu) = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right)\varepsilon_t \tag{3}$$

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 6 Issue 9, September - 2022, Pages: 27-38

The $(1-B)^d$ is referred to as difference operator, ∇^d . The ARMA and ARIMA models have the ability to capture the short range dependence (SRD) property, since d is confined in the range of integer order.

ARFIMA model is introduced to capture the long memory or long-range dependence (LRD) of the fractional systems. It is one of the best known examples of long memory models. LRD is an indication that the decay of the autocorrelation function (ACF) is algebraic and slower than exponential decay so that the area under the function curve is infinite. The operator can be expressed using binomial expansion for any real number d with Gamma function:

$$(1-B)^d = \sum_{k=0}^{\infty}\binom{d}{k}(-B)^k = \sum_{k=0}^{\infty}\frac{\Gamma(d+1)}{\Gamma(d+1)\Gamma(d+1-k)}(-B)^k \tag{4}$$

The general form of $ARFIMA(p,d,q)$ process X_t is expressed as:

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \tag{5}$$

Where $d \in (-0.5, 0.5)$ and $(1-B)^d$ is defined as the fractional difference operator. $ARFIMA(p,d,q)$ processes are widely used in modeling long memory time series, where p is the autoregressive order, q is the moving average order and d is the level of differencing. The $ARFIMA(p,d,q)$ process is the natural generalization of the standard ARMA or ARIMA processes. It is widely applied in modeling long memory time series which take into cognizance the slowly decaying autocorrelations.

The ARFIMA model is stationary when $-0.5 < d < 0.5$.The ARFIMA model turns to non-stationary when $d \geq 0.5$ and stationary but non-invertible when $d \leq -0.5$. The ARFIMA model is termed a short memory if $d = 0.5$ and a unit root process if $d = 1$ Furthermore, it has a positive dependence among distant observations, if $0 < d < 0.5$; and it also has an anti-persistent property if $-0.5 < d < 0$ [13,14].

Non-stationary time series data is transformed into stationary data before it is suitable for further analysis and modeling using the approximated binomial expression of the long memory filters to estimate the memory parameters in the $ARFIMA(p,d,q)$. One of the methods to remove trends from time series data is to use models that involve fractional differencing filters [15].

When a non-stationary time series pass through fractional order difference filter, the series will yield residuals that are uncorrelated and normally distributed with constant variance as depicted in Figure 8.

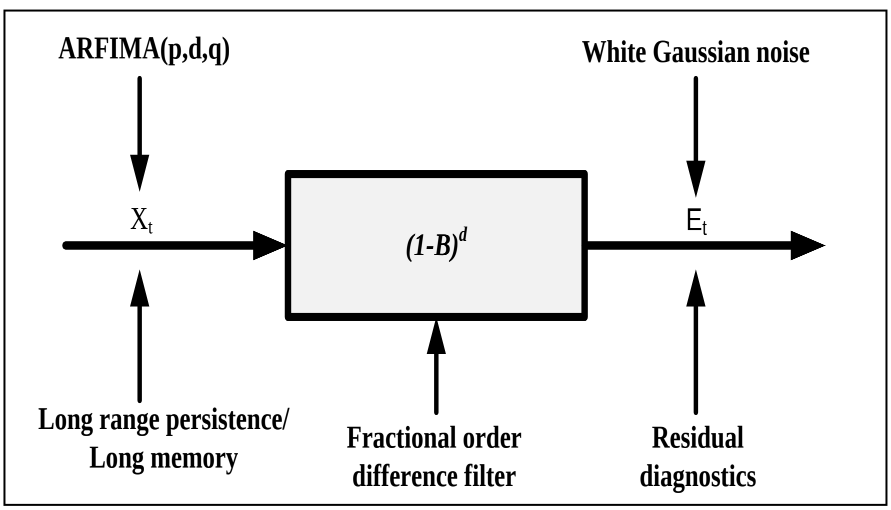


Fig. 8 Fractional order difference Filter in ARFIMA process

Finally, d is the parameter to be estimated first and foremost. Thus, the fractional order filter is needed to eliminate long memory property to obtain the stationary series [15].

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 6 Issue 9, September - 2022, Pages: 27-38

2.6 Model Evaluation

In order to evaluate the models, Akaike Information Criterion (AIC) was used. AIC is a model selection metric for determining how well a model fits into a dataset that was used to build it. AIC compares different possible models and suggests which one is the best fit for the data. It evaluates the relative information value of the model using the maximum likelihood estimate and the number of parameters in the models [16].

The mathematical expression of AIC is:

$$AIC = 2K - 2\ln(L) \tag{6}$$

Where K is the number of independent variable(s) used and L is the maximum likelihood estimate. The default K is always and increases with the number of independent variable. When comparing models, the model with lowest AIC is considered to be the best-fit model.

2.7 Model Estimation

The ARFIMA model parameters are estimated in following three steps: The first step is the estimation of the differencing parameter (d). This is followed by fitting the time series to $ARFIMA(p,d,q)$ models using several combinations of autoregressive and moving average functions with different orders. The last step is the use of the Akaike Information Criterion (AIC) to compare the different models and select the best fit.

3 RESULTS AND DISCUSSIONS

The study examined twenty five (25) different ARFIMA models. All ARFIMA models are trained with 70% of the data and the remaining 30% of the data was used to test models' performance. All models were simulated using the same training data. Ten (10) out of the ARFIMA models did not converged when they were trained. Parameter estimation in the ARFIMA models was performed. The water consumption was forecasted using the best fitted ARFIMA model. All models are implemented using the Stata 15 software.

For the water demand model in Dhofar Governorate, the ARFIMA model with the lowest AIC was found to be ARFIMA (0,-0.22,0). The AIC value is obtained as -456.9095. The summary of the AIC values of the fifteen (15) fitted models is shown in Table 2.

Table 2 AIC values of ARFIMA models

Models	**Auto-regressive (p)**	**Moving average (q)**	Differencing parameter **(d)**	**Akaike information criterion (AIC)**
ARFIMA 1	0	0	-0.22	-456.9095
ARFIMA 2	0	1	-0.25	-454.994
ARFIMA 3	0	2	-0.22	-454.9417
ARFIMA 4	0	3	-0.22	-454.9417
ARFIMA 5	1	0	-0.25	-454.9999
ARFIMA 6	1	2	-0.28	-453.0422
ARFIMA 7	2	0	-0.21	-454.9107
ARFIMA 8	2	1	-0.27	-453.0316
ARFIMA 9	3	0	-0.22	-454.9343
ARFIMA 10	3	1	-0.27	-453.1129
ARFIMA 11	1/1	0	-0.25	-454.9999
ARFIMA 12	1/1	2	-0.28	-453.0422
ARFIMA 13	1/2	0	-0.28	-453.0479
ARFIMA 14	2/1	0	-0.28	-453.0479
ARFIMA 15	2/1	2	-0.39	-453.9638

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 6 Issue 9, September - 2022, Pages: 27-38

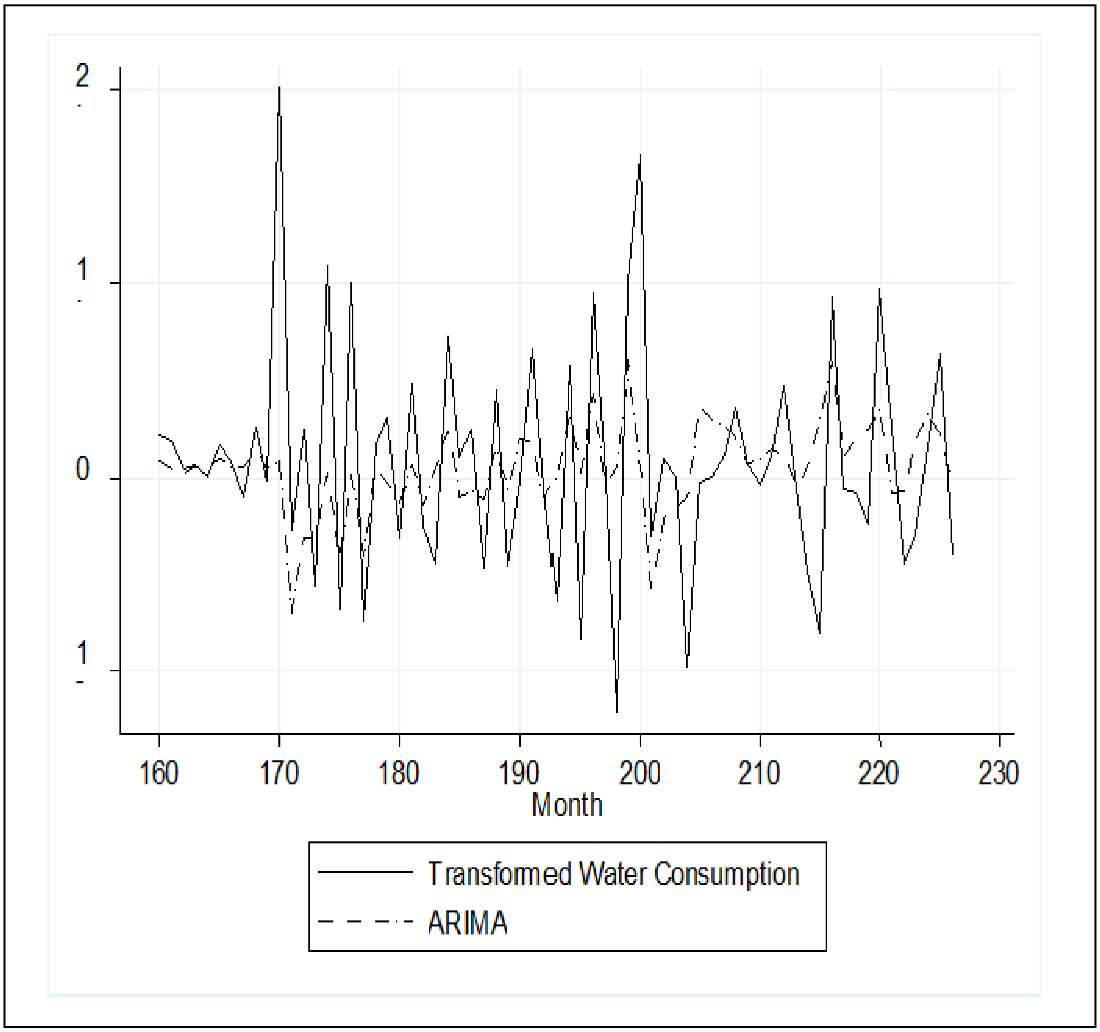
The best fitted ARFIMA model for the water consumption in Dhofar is

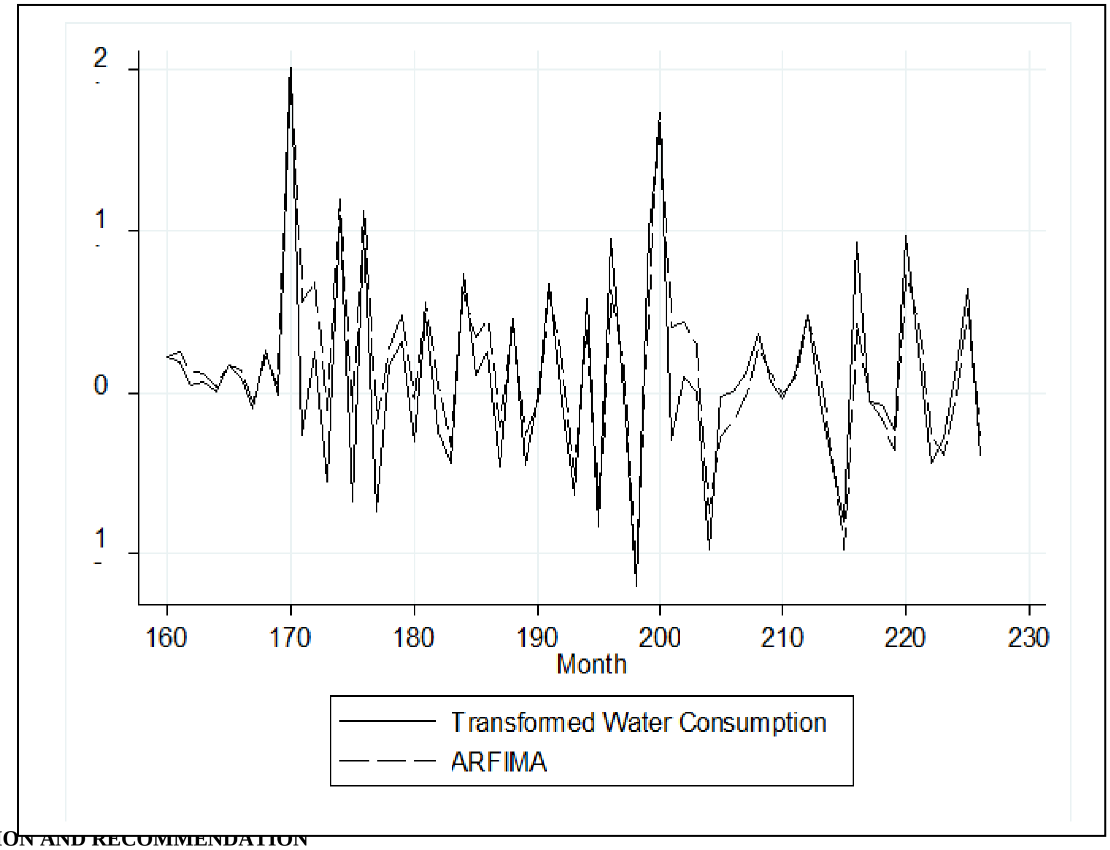
$$(1-B)^{-0.22} X_t = Z_t + .0084 \tag{7}$$

where $\{Z_t\} \sim WN(0,\ 0.0027471)$

In order to assess the performance of the best fitted model, the ARFIMA model was compared with ARIMA model. Figures 9 and 10 show the simulation plots of ARIMA and ARFIMA models, respectively. It is shown that the ARFIMA model closely reproduces the actual water consumption data set better than ARIMA model. AIC values for ARFIMA and ARIMA model was found to be -198.08 and -187.47, respectively. ARFIMA model had lower AIC value. This is an indication that the ARFIMA model yield more accurate results and suitable to predict future water consumption in Dhofar better than ARIMA model. This finding agrees with the results in [5].

Furthermore, the ARFIMA model was applied to forecast the 2-year water demand for Dhofar Governorate. The predicted water demand is presented in Table 3 at 95% confidence interval.





4 CONCLUSION AND RECOMMENDATION

Effective planning and management of water supply systems involves development of reliable water demand models. ARFIMA models are statistical tools used for modeling time series data. In this study, the monthly water consumption data of Dhofar Governorate in Oman was modeled and predicted for two years ahead using ARFIMA model. Fifteen (15) ARFIMA models were involved in the simulation study. The best fitting model was found to be ARFIMA (0,-0.22,0). The importance of the study is that the ARFIMA model has been demonstrated to be a useful modelling tool to forecast water demand for municipalities and large cities. The designers, engineers and managers in the water supply industry could easily adopt this model to design systems that will supply water to meet the varying demands of the public.

The limitation of the study is that only one Governorate in Oman was examined. Future studies can be extended to more Governorates to determine and compare the performance of ARFIMA models. Also, this study could be extended by developing multivariate models that take into consideration economic and climatic variables that affect water demand.

Table 3 Forecast values for water consumption in Dhofar Governorate

Year	**Month**	**Water Demand (m^3)**
2020	November	5387712
	December	5379912.5
2021	January	5372260.5
	February	5364951
	March	5357986.5
	April	5351351.5
	May	5345031.5
	June	5339011
	July	5333276
	August	5327813
	September	5322609
	October	5317652
	November	5312930

	December	5308431.5
2022	January	5304147
	February	5300065
	March	5296177
	April	5292473
	May	5288945
	June	5285584
	July	5282382.5
	August	5279333
	September	5276428
	October	5273660.5

5 REFERENCES

[1] Yılmaz, V. (2022), "The use of band similarity in urban water demand forecasting as a new method", *Water Supply*, vol. 22, 1, pp. 1004-1019.

[2] Adamowski, J., & Karapataki, C. (2010). "Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms", *J. Hydrol. Eng*., vol. 15, pp. 729-743.https://www.medrxiv.org/content/10.1101/2020.05.03.20089615v1.full-text.

[3] Ristow, D., Henning, E., Kalbusch, A., & Petersen, C. E. (2021), "Models for forecasting water demand using time series analysis: a case study in Southern Brazil", *Journal of Water, Sanitation and Hygiene for Development*, vol. 11, no. 2, pp. 231-240.

[4] Candelieri, A. (2017), "Clustering and support vector regression for water demand forecasting and anomaly Detection", *Water*, vol. 9, no. 3, 224 .https://www.mdpi.com/2073-4441/9/3/224/htm.

[5] Pan, J. N., & Chen, S. T. (2008), "Monitoring long-memory air quality data using ARFIMA model", *Environmetrics: The official journal of the International Environmetrics Society*, vol. 19, no.2, 209-219.

[6] Zhou, D., Chen, S., & Dong, S. (2013), "Network traffic prediction based on ARFIMA model" arXivpreprint. arXiv:1302.6324.https://arxiv.org/abs/1302.6324

[7] Belbute, J. M., & Pereira, A. M. (2015), "An alternative reference scenario for global CO2 emissions from fuel consumption: An ARFIMA approach",. *Economics Letters*, vol. 136, pp. 108-111.

[8] Boateng, A., Gil-Alana, L. A., 'Maseka, L., Siweya, H., & Belete, A. (2016), "*Long memory and ARFIMA modelling: The case of CPI inflation rate in Ghana*", *The Journal of Developing Areas*, pp. 287-304.

[9] Balah, B., & Djeddou, M. (2020), "Forecasting COVID-19 new cases in Algeria using Autoregressive fractionally integrated moving average Models (ARFIMA)", medRxiv.

[10] Polotzek, K., & Kantz, H. (2020), "An ARFIMA-based model for daily precipitation amounts with direct access to Fluctuations". *Stochastic environmental research and risk assessment*, vol. 34, no.10, pp. 1487-1505.

[11] Al Amri, I. T. (2019), "Tourism constituents and tourism marketing in Oman (Dhofar Governorate as a Model" https://mrk.journals.ekb.eg/article_100405_165853468a3d5f06b9156f88895c8596.pdf

[12] World Bank Group (2022). Oman Water Statistics. https://www.data.worldbank.org

[13] Shitan, M., Wee, P. M. J., Chin, L. Y., &Siew, L. Y. (2008). "Arima and integrated arfima models for forecasting annualdemersal and pelagic marine fish production in Malaysia", *Malaysian Journal of mathematical sciences*, vol. 2, no. 2, 41-54.

[14] Dhliwayo, L., Matarise, F., &Chimedza, C. (2020), "Modeling Seasonal Fractionally Integrated Autoregressive Moving Average-Generalized Autoregressive Conditional Heteroscedasticity Model with Seasonal Level Shift Intervention", *Open Journal of Statistics*, vol. 10, no. 05, pp. 810-831.https://doi.org/10.4236/ojs.2020.105047.

[15] Liu, K., Chen, Y., & Zhang, X. (2017), "An evaluation of ARFIMA (autoregressive fractional integral moving average) programs", *Axioms*, vol. 6, no. 2, 16. https://www.mdpi.com/203608.

[16] Wang, Y., & Liu, Q. (2006), "Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships", *Fisheries Research*, vol. 7, no. 2, 220-225.

Authors

Oladipupo Bello, Author's profile.

Oladipupo Bello is a graduate of the Federal University of Technology, Akure, Nigeria with a Bachelor of Engineering (Electrical/Electronic). He obtained Master of Engineering (Electrical) from the Bayero University, Kano, Nigeria in 2010. He is a registered member of the Council for the Regulation of Engineering in Nigeria (COREN). He has coordinated quality assurance scheme, development and review of Engineering curricula for the Nigerian Polytechnic System. He holds doctoral degree in Electrical Engineering from the Tshwane University of Technology, Pretoria, South Africa. He has co-authored several publications in peer-review journals and conference proceedings on coagulation process modeling, model predictive control schemes, and applications of computational intelligence to water supply systems.