International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced ARIMA Model for Water Demand Forecasting in Smart Water Distribution Network

Subha J[1], M. Kowsigan[*2]

Submitted: 23/04/2023 **Revised**: 27/06/2023 **Accepted**: 07/07/2023

Abstract*:* The fraction of the world's freshwater resources that are usable each year decreases. A poll conducted by the World Economic Forum predicts that during the next two decades, there will be severe water shortages all across the world due to rising demand. It is difficult to both stop the rising demand for water and cut down on the amount of water that is wasted in transit. Cities are increasingly adopting IoT-enabled water distribution systems that employ smart water meters to collect real-time data on water consumption and transfer it either to the cloud, fog, or edge. Then it can be stored, analysed for patterns, and used to plan for future water needs and create more effective infrastructure. It's crucial to anticipate and analyse client demand for water use. The enhanced auto-regressive integrated moving average (ARIMA) method is used to analyse the trend of water consumption data and forecast future water consumption demand based on previous historical information. When compared to other forecasting methods, they tend to provide better results. It is important to have an accurate forecast of water use. Planning and building water supply systems rely heavily on accurate and dependable forecasts. The ARIMA model was validated using the mean absolute scaled error (MASE) and root mean square error (RMSE).

Keywords*: accurate, validated, ARIMA, consumption, forecasting*

1. Introduction

1.1. Back ground

Many people's health problems now stem from their inability to reliably get clean water. It has been reported by the World Health Organization (WHO) that poor water quality and its unequal distribution to consumers are important causes of health problems across the world. On the demand side, issues with water quality and resource allocation are likely to arise as a result of population growth, fast urbanisation, and growing demands from agricultural and energy production [1]. There are two distinct kinds of water distribution networks, distinguished by the orientation of the distribution pipes. Both an aboveground and an underground system are part of the water distribution network [2]. The wireless sensor network is selected according to whether the water distribution system uses underground wireless sensor networks (UWSN) above or below ground. Underground pipe monitoring Until now, the water distribution network has not made use of the SCADA (Supervisory Control and Data Acquisition) system, which is used for supervisory control and automation in water distribution facilities [3].

The state of the art in water demand forecasting in water distribution involves the use of advanced modelling techniques and data analytics to improve the accuracy and reliability of predictions. Statistical models, such as autoregressive integrated moving average (ARIMA) [5], seasonal decomposition of time series (STL), and regression models, have been widely used for water demand forecasting. These models analyse historical water consumption data along with relevant explanatory variables like weather data and population data to identify patterns and trends. Machine learning algorithms, including artificial neural networks (ANNs) [10], support vector machines (SVMs), and random forests [13], have been applied to water consumption forecasting. These models can capture nonlinear relationships and complex interactions between various factors influencing water demand, leading to improved forecasting accuracy.

1.2. Problem statement

The traditional water distribution system typically involves a network of pipes, pump stations, storage tanks, and valves that are used to transport potable water from the treatment plant to consumers. Broken pipes, faulty demand prediction systems, inadequate monitoring, poor management, and insufficient distribution owing to leakage and theft are major problems in the water distribution system. A reliable and effective water distribution management system must be developed to prevent this uneven distribution. All of these problems are inherent in India's present water delivery infrastructure. The layout of a water distribution system can vary depending on factors such as geography, population density, and water demand. The conventional water

[1]*Research Scholar,Department of Computing Technologies,SRM Institute of Science and Technology,*
Kattankulathur, Chengalpattu, Tamilnadu, India-603203,sj2773@srmist.edu.in
[2]*Department of Computing Technologies,Faculty of Engineering and Technology,SRM Institute of Science and Technology*
Kattankulathur, Chengalpattu, Tamilnadu, India-603203,kowsigam@srmist.edu.in

distribution system is shown in Fig.1. Developing an accurate and reliable water demand forecasting system that can predict future water consumption patterns for a given region or water supply network The primary objective is to provide accurate estimates of water demand in order to optimise water resource management, improve infrastructure planning, and ensure the sustainable and efficient allocation of water resources.

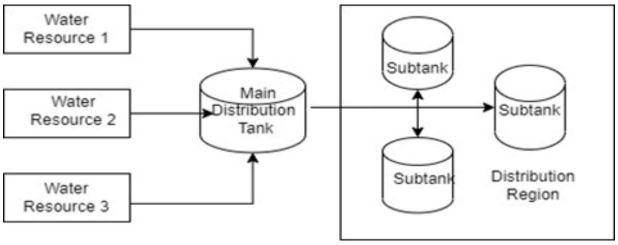

Fig. 1. Treditional Water Distribution System.

1.3. The Proposed work

To overcome the issue of traditional water distribution The Internet of Things (IoT)-integrated Water Distribution System (WDS) will be developed to meet customer demand. A smart water distribution system uses advanced technologies such as sensors, automation, and data analysis to monitor and control the distribution of water in a city or town. It allows for real-time monitoring of water flow, pressure, and leaks, as well as the ability to remotely control valves and pumps to optimise the distribution of water. The system also includes smart meters that can be used to track water consumption using information and communication technologies (ICT). Water conservation may be encouraged via the adoption of digital water metres, which provide for individualised and comprehensive user feedback on the consumption of water. Researchers have studied the impact of various water-use feedback mechanisms since the emergence of the digital meter [4]. And this water consumption data can be collected remotely without detecting potential leaks in homes and businesses, as depicted in fig. 2.

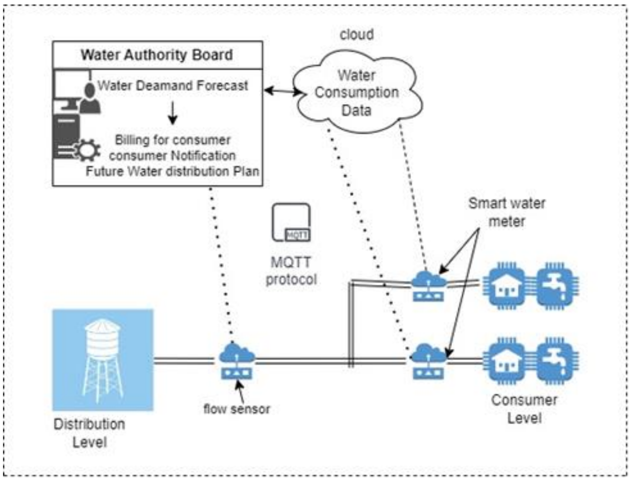

Fig 2. Smart Water distribution System with Water demand forcast model.

1.4. ARIMA Model

Demand and analysis of daily water consumption in Austin, Texas, USA, using a deep learning algorithm known as ARIMA (Auto Regression Integrated Moving Average) have been made, including an assessment of the margin of error. The working model of ARIMA is shown in Fig 3. Comparing the ARIMA model to other algorithms used in water forecasting, the ARIMA model provides more precise forecasts with a smaller error. This information can be used to identify and address problems more quickly and efficiently, conserve water, and reduce costs. The ARIMA model is comprised of the autoregressive (AR), differencing (I), and moving average (MA) components. Algorithm 1 shows the general framework for building an ARIMA model for water demand forecasting.

Algorithm 1: ARIMA Model

Input: Time series water consumption data

Output: Water demand forecast values

Begin:

Step 1: Data Preparation

Step 2: Stationarity Check on data

Step 3: Differencing between data patterns

Step 4: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Step 5: Model Selection and Training

Step 6: Forecasting water demand

Step 7: Model Evaluation and Iteration by measuring Errors

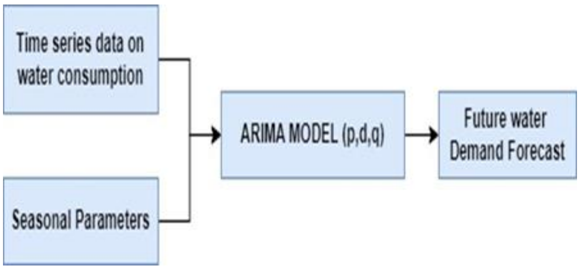

Fig. 3. Working of ARIMA model**.**

2. Related Works

The accuracy of the water distribution system is enhanced by the demand forecasting model. Based on data collected from January 1, 2010, through March 31, 2018, this Water Demand Study (WDS) forecasts future water demand and analyses daily water consumption in Austin, Texas, USA, using a deep learning algorithm known as LSTM (long short-term memory) and a time series algorithm known as ARIMA (Auto Regression Integrated Moving Average). A forecast of water use from January to March 2019 has been made, including an assessment of the margin of error. Comparing LSTM to ARIMA, LSTM is shown to provide more precise forecasts with smaller standard deviations. The 24-hour water demand was also included in the design of the water distribution system [5].

The Internet of Things (IoT) and fog computing have been suggested as a unified architecture for an underground WDS. For water distribution planning, work has been done on both predicting water demand and proposing an Internet of Things-based architecture. Three months of daily forecasts for water consumption have been made using ARIMA and regression analysis. Water supply design for an Internet of Things-based framework has been carried out, making use of hydraulic systems engineering for efficient water transfer with minimal losses to set up an intelligent system for distributing water. This has been accomplished with the help of the EPANET simulation tool. The analysis of regularity in water demand data led to the selection of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model as the basis for the forecast procedure. [6].

The Future Industrial Kitchen (FIK) project's FIKWater dataset is the result of two to four weeks' worth of collecting time series data for the two types of water needs from three Portuguese restaurants using the FIKWater framework. The samples were collected at a rate of 0.2 Hz using ultrasonic flow meters. Specifics about the areas under surveillance are also disclosed. On March 1, 2021, an ultrasonic flow meter, model number TUF2000M, was put on the primary water entry pipe to track the daily water use. Instantaneous flow rate, fluid velocity, active vibration speed, and overall consumption are only some of the characteristics measured by the TUF2000M [7].

A long-term water consumption forecast model has been created for SIBU City, Malaysia. To obtain prediction consumption with minimal error, we apply a model predicated on deep learning in neural networks (DLNN) using the LSTM method, which first employs a multi-perception layer to learn patterns from historical data. With the use of the RMSE (root mean squared error) statistic, this model is accurate [8].

This research recommended that Nanchang, China's capital city, use a new dynamic Firefly algorithm (NDFA) to estimate future water resource needs. There are three different models for making predictions: linear, exponential, and hybrid. In this scenario, we look at water use statistics from 2003 to 2015. The ideal model weights are determined for the years 2003–2012. The years 2013–2015 are used in the model to test. The water use from 2017 to 2020 has been forecast with an accuracy of up to 97.91%. [9]

The short-term prediction model uses a variety of machine learning and statistical methods to predict future water consumption, including artificial neural networks, deep neural networks, extreme learning machines, least square support vector machines (LSSVM), Gaussian process regression, random forests, and multiple regression. R-squared, root-mean-square error, mean-square error, and mean-absolute error have all been used to evaluate the relative efficacy of the different methods. The ANN model had the greatest performance while making predictions for 1, 12, and 24-hour periods. Regarding processing and precision, the LSSVM is the superior option [10].

To produce accurate short-term water demand estimations, the authors of this work suggest a parallel global optimization approach to fine-tune the hyperparameters of a trained support vector machine for 24 hours. four Each SVM takes the first six hours of water use as input characteristics and uses the next eighteen hours of use to forecast a single hourly demand. Black-box objective functions are optimised based on the mean average percentage error (MAPE), which is calculated via leave-one-out validation [11].

The water consumption data of South East water retail company from 2006 to 2015 is utilised to predict monthly demands in Melbourne, Australia, taking into account climatic factors via the use of discrete wavelet transform, principal component analysis, and particle-swarm optimisation. Efficiency, accuracy, and precision have all been verified by rigorous testing of this model. These measurements demonstrate the high efficiency and low error rate of the suggested approach [12].

Use meter readings from 90 people to make hourly estimates of water usage. In this research, we explore the use of machine learning methods including random forests (RFs), artificial neural networks (ANNs), and support vector regression (SVR) with delayed supply, periodicity, climate, and demographic details. RMSE is used to evaluate the model's efficacy. The K-means clustering algorithm was used to classify typical daily water consumption habits. The accuracy rates of RF and ANN are higher than those of SVR [13].

A method has been created to enable short-term projections of daily farming water consumption using a dynamic artificial neural network (ANN) architecture and the Bayesian framework, even when such data is scarce. To achieve this goal, the ANN design is optimised with the use of a genetic algorithm (GA). Specifically, the Bembezar MD Irrigation District used this strategy. The created model outperformed its predecessors by an increase of 3 to 11 percent in terms of accuracy of prediction. Standard Error Prediction (SEP) was 8.7%, and R2 was 96% for the ANN model [14].

In this study, we combine the linear model with the exponential model and the logarithmic model. An enhanced whale optimization algorithm using a social-learning-based wavelet mutation technique is offered as a means of more precisely predicting future water-resource demands. The new method creates a new linear incremental probability, which boosts the algorithm's capacity for global search. As compared to WOA, the proposed algorithm achieves a more optimal balance between exploiting and exploring the space. Using the most recent CEC 2017 benchmark. Shaanxi Province, China's water use data from 2004–2016, is utilised for the study.

Compared to existing algorithms, the suggested one performs better while solving the three water resource forecasting models. Up to 99.6 percent of the forecast's accuracy may be guaranteed [15].

This report recommends doing in-depth research into the tuning of AI neural networks to foretell the demand for drinkable water. Feed-forward neural networks, long short-term memory, simple recurrent neural networks, and gated recurrent units are some of the designs used, with prediction intervals ranging from an hour to a week. For a certain neural network architecture, prediction horizon, and dataset, the optimal number of layers and nodes may vary greatly. The models in the proposed research will be statistically evaluated using the following Values of MAPE that are higher than those of R2 signal that the model is more effective [16].

This research employed several statistical approaches to analyse and pick the model's best input variables to determine how effectively meteorological factors can predict future urban water demand. The GSA-ANN and BSA-ANN algorithms were employed for these predictions. This study showed that employing statistical criteria to choose model inputs works and that the GSA-ANN hybrid model is more accurate than other hybrid models [17].This research predicted near-term water usage using ensemble learning. Future water demands are predicted using the ensemble learning model and the equal-dimension, new-information model. The suggested method was compared to the single back-propagation neural network (BPNN) and seasonal autoregressive integrated moving average (SARIMA) models with the help of a real-world water distribution system.Water demand projections are more accurate and reliable using the suggested technique. RMSE, APE, and MAPE are used to assess forecast model accuracy [18].

The multi-stage, windowed SSA-AR model uses single-spectrum estimates and autoregressive models with numerous windows to reduce noise. The SSA stages evaluate the original water data series. An autoregressive (AR) model predicts freshwater demand. The model is validated using 2006–2015 data from Baghdad's Al-Wehda treatment center. SSA-AR can predict future water usage from incorrect historical data [19].Long-term water demand projections have been made for the Blue Mountains Water Supply System in New South Wales, Australia, and the results showed that different variable selection processes gave different collections of predictor characteristics. In addition, some of the selection strategies resulted in a number of irrational independent variables and regression equations. In contrast, when PCA was used to prepare the datasets of the variables that were predicted, the resulting water requirement model provided more precise simulation results of the water needs. [20].

To predict future mistakes, we first use a model based on the Least Squares Support Vector Machine (LSSVM), then convert the error time series into a chaotic time series, and lastly, we apply the LSSVM technique. The hybrid approach is tested in three district metering areas in Beijing with various demand trends. As a result of the error-rectification section, the hybrid model was able to reduce the MAPE of the expected demand [21].The Backtracking Search Algorithm (BSA-ANN) optimises an ANN to predict monthly water requirements. South Africa's Gauteng province's monthly water use data was used to develop and validate the method between 2007 and 2016. The BSA-ANN model has the lowest RMSE (0.0099 megaliters) and highest efficiency (0.979). With a lower error rate than the Crow Search Algorithm (CSA-ANN), it was also more reliable [22].

Over the following six months, a slime mould technique (SMA-ANN) refined an updated artificial neural network (ANN) model for urban water demand forecasting. Over 16 years, ten weather elements modelled water demand stochastically. SMA-ANN hybrid models outperform traditional neural networks in statistical testing, and this approach gives accurate findings with a mean absolute relative deviation of 0.001 and a coefficient of 0.9. This research may help municipal water managers better administer the existing water system and prepare for expansions to meet future needs [23]. In the first step of this model, the price series is decomposed using the SSA. Next, a nonlinear autoregressive neural network (NARNN) is trained using each component to predict prices in the future [27]. Predicting future hardware sales using ARIMA and a recurrent neural network-long short-term memory (RNN-LSTM) [28] .

3. Research Method

3.1. Dataset

In order to evaluate the accuracy and precision of our simulation model, let's first talk about how we'd go about projecting future water demands in the proposed system.The dataset used in the experiment was compiled from information gathered by an IoT system in Austin TX ,that monitors water use and weather data across many regions to foresee the region's future water demand. The system is put to work analysing the pattern of water use. Among the 4701 data points included in the dataset are the following features: date and monthly water usage (in gallons) at the distributor level. Consumption data from 2012–2020 is utilized, with a monthly time step used for analysis.

3.2. Pre-processing

The unprocessed data undergoes a metamorphosis at the pre-processing stage. We get the month values from the Date variable and also retrieve and group them as numbers from 1 to 12. Use class() to determine whether the data is a time series. ts() will transform non-time series data into time series data by finding the minimum and maximum from the Date variable. Then choose monthly reports that are generated from the gathered information. To ensure the data is stable, use the ACF () and PACF () methods on it. If the data is steady, a P-value of 0.05 or less will be obtained in the Dicky-Fuller test. Data that has a P-value higher than 0.05 suggests instability. To use it in a forecasting model, it must first be transformed into a stationary form.

3.3. Methodology Used

The forecasting methods used in this study are briefly discussed here. Because of its reliability and low margin for error, the ARIMA model was chosen for this analysis. When 'p' is the relationship between data points and the number of observations, "d" is its dependency on the variances between subsequent observations, and "q" is an observation's reliance on residual errors, the ARIMA (p, d, q) model with the best fit is selected using auto Arima is the function that automatically selects the best ARIMA model for a given time series dataset based on its Akaike information criterion (AIC) value. It has been determined that the best-fitting model can be used to predict water use up to the year 2025. ARIMA (0,0,1) is selected as the best model, shown below in Table 1, for the lagging data with a 95% confidence interval. The Ljunk-Box test is then used to determine whether the model has an autocorrelation problem and whether or not the model has been verified. Every lag value's p-value must be greater than 0.05.

4. Results and Discussion

To evaluate the efficacy of the model so far, it was built in RStudio, a statistical program, using the ARIMA predictive model. Water consumption forecasts have been made using the various evaluation criteria that have been incorporated into the model and are discussed in Section 5. Water consumption patterns throughout several city areas in the United States are shown in Fig. 4. from January 2012 to September 2020. Data on water usage was collected from January 2015 through September 2020 to predict water demand for the subsequent five years. i.e., until December 2025. Fig. 5. depicts the demand pattern from January 2015 to September 2020. The variations in data points from 2015 to 2020 consumption are shown in Fig. 6. should be aware of the changes in water use so that we can analyse the data and make accurate projections on demand.Before using the ARIMA (0,0,1) model that fits the data the best. To determine whether the data is stationary or not, we should run an autocorrelation test on it. The ACF function is used to determine the order of autocorrelation problems, such as pattern detection and noise analysis.

Table 1. Finding the best-fit model automatically using the Autoarima function.

ARIMA (2,0,2) (1,0,1) [12]with non-zero mean	Inf
ARIMA (0,0,0) with non-zero mean	2502.778
ARIMA (1,0,0) (1,0,0) [12] with non-zero mean	2499.73
ARIMA (0,0,1) (0,0,1) [12] with non-zero mean	2499.442
ARIMA (0,0,0) with zero mean	2557.236
ARIMA (0,0,1) with non-zero mean	2498.706
ARIMA (0,0,1) (1,0,0) [12] with non-zero mean	2499.18
ARIMA (0,0,1) (1,0,1) [12] with non-zero mean	2500.59
ARIMA (1,0,1) with non-zero mean	2500.702

ARIMA (0,0,2) with non-zero mean	2500.699
ARIMA (1,0,0) with non-zero mean	2499.669
ARIMA (1,0,2) with non-zero mean	2502.68
ARIMA (0,0,1) with zero mean	2531.978
Best model: ARIMA (0,0,1) with non-zero mean	

The order of the moving average is determined by partial ACF. The data has then been subjected to the Augmented Dickey-Fuller Test to determine whether or not the data is stationary by looking at the P-value. If the P-value is less than or equal to 0.05, then the data sample is stable. The P-value for the water usage data is 0.05698, as indicated in the ADF test results below in Fig. 7. All data points inside the blue line appear to provide information about the 95% confidence interval, which implies a significance threshold close to zero.

In the statistical model, the ACF and PACF functions are used for visualising and analysing the autocorrelation and partial autocorrelation functions of the residuals of a time series model shown in Figures 8, 9,10, 11, and 12. The ACF evaluates the degree to which a time series is correlated with its lagging values. The residuals of a time series model are investigated to see whether autocorrelation exists. The ACF plot shows the lag on the x-axis and the correlation coefficient on the y-axis. To account for the impact of temporal delays, the PACF calculates the correlation between a time series and its delayed values. In a time series model, it is used to determine the relative positions of the AR and MA terms. The precision of a time series model's forecasts may be enhanced by examining the ACF and PACF plots to decide in which order the AR and MA components should be included.

In particular, the Forecast function is used to generate forecasts using a statistical model called a "wavelet-based time series model with a non-zero mean." The argument specifies the level of confidence desired for the forecast, in this case, 95%, and another argument specifies the number of periods ahead to forecast, in this case, 5 years, and the time series has a monthly frequency.

```
        Augmented Dickey-Fuller Test

data:  wc
Dickey-Fuller = -3.4383, Lag order = 4, p-value = 0.05698
alternative hypothesis: stationary
```

Fig 7. Augmented Dickey-Fuller(ADF) Test to check for data stionarity

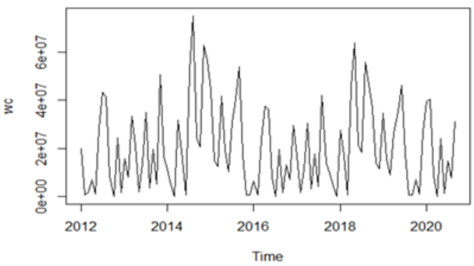

Fig. 4. Water Consumption pattern from 2012 to 2020

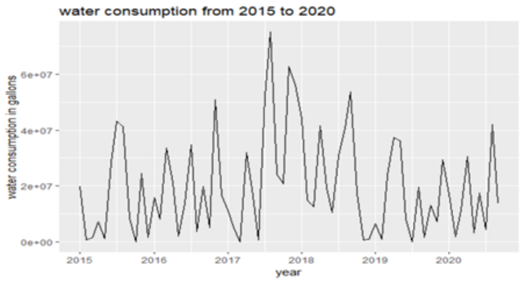

Fig. 5. Water Consumption pattern from 2015 to 2020.

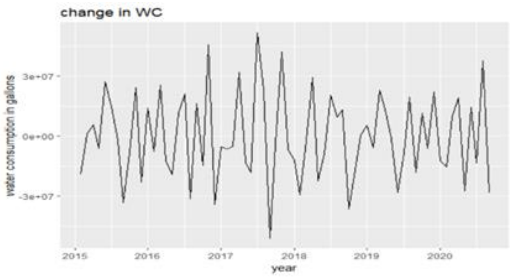

Fig. 6. Difference in Water Consumption pattern from 2015 to 2020

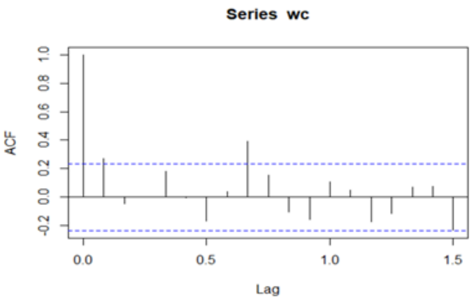

Fig 8. Checks the data's Auto correlation by ACF test.

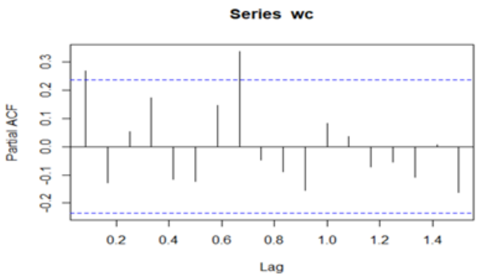

Fig 9. Checks the data's Auto correlation by Partial ACF test

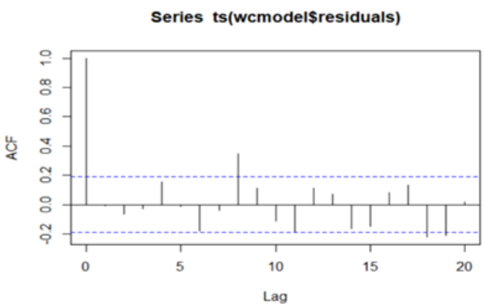

Fig 10. The Autocorrelation of residuals using ACF

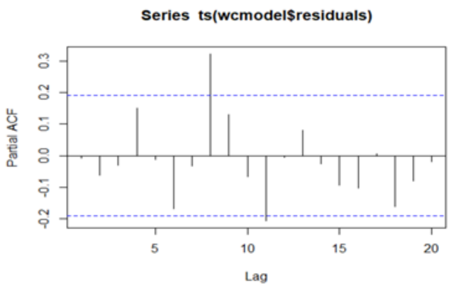

Fig 11. The Autocorrelation of residuals using PACF

The "non-zero mean" part of the model indicates that the time series being modelled has a non-zero average level. This is important to specify because it affects how the model generates forecasts. If the time series has a non-zero mean, the model will need to incorporate this information in its forecasts to avoid predicting unrealistic values. The following Table 2 displays the results of an ARIMA model used to forecast monthly water use from January 2021 through December 2025. and the graphic depicts a plot based on the expected value highlighted in the shaded part of Figure 12.The Box.test() function in R is used to perform the Ljung-Box test for autocorrelation in a time series to validate the ARIMA forecasting model.

The Ljung-Box test is a statistical test to determine if there is evidence of autocorrelation in a time series. The function takes three arguments. The residuals of the time series model are the errors or differences between the observed values and the values predicted by the model. The residuals are accessed through forecast1$residuals. The "lag" argument specifies the number of lags to include in the test. This lag of 5 means that the test will include autocorrelation up to 5 lags. The type argument specifies the type of test to be performed. The Ljung-Box test that will be run is shown in Figure 13, and if the p-value is less than 0.05, the data will be disregarded. The reliability of the model used for making predictions is also evaluated.

Table 3 displays the results of a validation of the enhanced ARIMA model to determine the precision with which the metrics listed below may be estimated. The mean absolute error (MAE) value assessed for the ARIMA model's projection of daily rainfall is 102.7644 [24], whereas the MAE value for the proposed model's projection of future water demand is 144.20105. The RMSE for a time series study of electrical energy use using an ARIMA model, which was shown to be valid, is 347.18 [25]; the RMSE for our suggested system is 169.103. The comparison of existing models' metrics with our model is shown in Table 4.

Table 2. Sample Forecasts (JAN 2021 to DEC 2025) by the ARIMA models for water demand of Austin. city,US

Month Year	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct-20	15404379	-6588121	37396878	-18230247	49039004
Nov-20	19855683	-3145376	42856743	-15321402	55032768
Dec-20	19855683	-3145376	42856743	-15321402	55032768
Jan-21	19855683	-3145376	42856743	-15321402	55032768
Feb-21	19855683	-3145376	42856743	-15321402	55032768
Mar-21	19855683	-3145376	42856743	-15321402	55032768
Apr-21	19855683	-3145376	42856743	-15321402	55032768
May-21	19855683	-3145376	42856743	-15321402	55032768

Jun-21	19855683	-3145376	42856743	-15321402	55032768
Jul-21	19855683	-3145376	42856743	-15321402	55032768
Aug-21	19855683	-3145376	42856743	-15321402	55032768
Sep-21	19855683	-3145376	42856743	-15321402	55032768
Oct-21	19855683	-3145376	42856743	-15321402	55032768
Nov-21	19855683	-3145376	42856743	-15321402	55032768
Dec-21	19855683	-3145376	42856743	-15321402	55032768
Jan-22	19855683	-3145376	42856743	-15321402	55032768

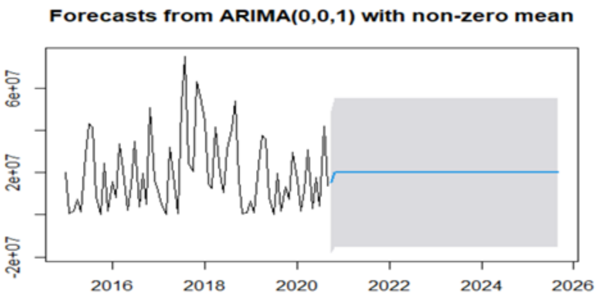

Fig.12. Forecast graph from 2021 to 2025.

```
        Box-Ljung test

data:  forecast1$residuals
X-squared = 3.0111, df = 5, p-value = 0.6983
```

```
        Box-Ljung test

data:  forecast1$residuals
X-squared = 17.712, df = 10, p-value = 0.06002
```

Fig. 13. The Ljung-Box test for different lag values.

Table 3. ARIMA model metrics value.

Metrics	ME	RMSE	MAE	MASE	ACF1
The training set	-14169.74	169.10302	144.20105	0.788321	0.0010364

Table 4. Comparison of our obtained RMSE with other existing analysis

Model References	Forecast	Method Used	RMSE
Mazwin Arleena Masngut1, Shuhaida Ismail, Aida Mustapha3 Suhaila Mohd Yasin[24]	Rainfall	ARIMA and ANN	34.674
Nahid Ferdous Aurna, Md. Tanjil Mostafa Rubel, Tanveer Ahmed Siddiqui, Tajbia Karim [25]	Electrical Energy Consumption	Holt Winters model	184.12
Redha Ali Al-Qazzaz, Suhad A. Yousif [26]	Oil Prices	Auto ARIMA	12.5539
Model we used to analyze	Water Consumption	Enhanced ARIMA Model	169.103

5. Conclusion

This article uses the ARIMA prediction framework to project water use in the years 2020–2025. The mode's accuracy was measured using the mean absolute scaled error (MASE), which was calculated to be 0.788321, and the RMSE value of 169.103. The ARIMA model was used to predict since it was shown to be the most effective. To better plan for the expected increase in demand for water in the future, most water systems use forecasting methodologies. The system's precision is the criterion that must be considered when engaging in fast and accurate decision-making for an IoT system. The practical findings demonstrate that when the MASE value is less than 1, the ARIMA model produces better, more precise predictions of future water requirements. The proposed ARIMA model has many advantages over other approaches, including its ability to handle non-stationary data by differencing the data to eliminate the trend and/or seasonality, its ability to capture trends and seasonality in the time series data, and its flexibility since multiple orders of the ARIMA model may be used to capture different patterns in the data.

References

[1] O'Connell, E. Towards Adaptation of Water Resource Systems to Climatic and Socio-Economic Change. *Water Resour Manage* **31**, 2965–2984 (2017). https://doi.org/10.1007/s11269-017-1734-2

[2] Afshar, A.; Miri Khombi, S.M. (2015). Multi objective Optimization of Sensor Placement in Water Distribution Networks; Dual Use Benefit approach. *International Journal of Optimizationin Civil Engineering*, 5(3), 315-331, 2015.

[3] Amatulla, P. H.; Navnath, B.P.; Yogesh, B.P. (2017). IoT based water management system for smart city. *International Journal of Advanced Research , Ideas and Innovations in Technology*, 3(2), 319-383, 2017.

[4] Rahim, M.S.; Nguyen, K.A.; Stewart, R.A.; Giurco, D.; Blumenstein, M. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review. *Water* 2020, *12*, 294. https://doi.org/10.3390/w12010294

[5] Lakshmi, K. N., Sankaranarayanan, S., Joel, J. P. C. R., & Kozlov, S. (2020). Water demand forecasting using deep learning in IoT enabled water distribution network. *International Journal of Computers, Communications and Control*, *15*(6) doi:https://doi.org/10.15837/ijccc.2020.6.3977

[6] Lakshmi Kanthan Narayanan, Suresh Sankaranarayanan; IoT-based water demand forecasting and distribution design for smart city. *Journal of Water and Climate Change* 1 December (2020); 11 (4): 1411–1428. doi: https://doi.org/10.2166/wcc.2019.019

[7] Pereira, L.; Aguiar, V.; Vasconcelos, F. FIKWater: A Water Consumption Dataset from Three Restaurant Kitchens in Portugal. *Data* 2021, *6*, 26. https://doi.org/10.3390/data6030026

[8] Norzanah Md Said; Zalhan Mohd Zin; Ismail, Mohd Nazri; Termizi Abu Bakar. International Journal of Advanced Technology and Engineering Exploration; Bhopal Vol. 8, Iss. 76, (Mar 2021): 473-483. DOI:10.19101/IJATEE.2020.762165.

[9] Hui Wang, Wenjun Wang, Zhihua Cui, Xinyu Zhou, Jia Zhao, Ya Li,A new dynamic firefly algorithm for demand estimation of water resources,Information Sciences,Volume 438,2018, Pages 95-106, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2018.01.041.

[10] Praveen Vijai, P Bagavathi Sivakumar, Performance comparison of techniques for water demand forecasting, Procedia Computer Science,Volume 143, 2018,Pages 258-266, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.394. (https://www.sciencedirect.com/science/article/pii/S187705091832091X)

[11] Antonio Candelieri, Ilaria Giordani, Francesco Archetti, Konstantin Barkalov, Iosif Meyerov, Alexey Polovinkin, Alexander Sysoyev, Nikolai Zolotykh,Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization,Computers & Operations Research,Volume 106, 2019,Pages 202-S209, ISSN 0305-0548,https://doi.org/10.1016/j.cor.2018.01.013. (https://www.sciencedirect.com/science/article/pii/S0305054818300133)

[12] Salah L. Zubaidi, Khalid Hashim, Saleem Ethaib, Nabeel Saleem Saad Al-Bdairi, Hussein Al-Bugharbee, Sadik Kamel Gharghan,A novel methodology to predict monthly municipal water demand based on weather variables scenario,Journal of King Saud University - Engineering Sciences, Volume 34, Issue 3, 2022, Pages 163-169,ISSN 1018-3639, https://doi.org/10.1016/j.jksues.2020.09.011. (https://www.sciencedirect.com/science/article/pii/S1018363920303111)

[13] Jorge E. Pesantez, Emily Zechman Berglund, Nikhil Kaza, Smart meters data for modeling and forecasting water demand at the user-level, Environmental Modelling & Software, Volume 125, 2020, 104633, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2020.104633.(https://www.sciencedirect.com/science/article/pii/S1364815219303457)

[14] Rafael González Perea, Emilio Camacho Poyato, Pilar Montesinos, Juan Antonio Rodríguez Díaz, Optimisation of water demand forecasting by artificial intelligence with short data sets, Biosystems Engineering, Volume 177, 2019, Pages 59-66, ISSN 1537-5110,https://doi.org/10.1016/j.biosystemseng.2018.03.011.(https://www.sciencedirect.com/science/article/pii/S1537511017311807)

[15] Wenyan Guo, Ting Liu, Fang Dai, Peng Xu,An improved whale optimization algorithm for forecasting water resources demand,Applied Soft Computing,Volume 86,2020,105925,ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.105925. (https://www.sciencedirect.com/science/article/pii/S1568494619307069).

[16] Menapace, A.; Zanfei, A.; Righetti, M. Tuning ANN Hyperparameters for Forecasting Drinking Water Demand. *Appl. Sci.* 2021, *11*, 4290. https://doi.org/10.3390/app11094290

[17] Zubaidi, S.L., Gharghan, S.K., Dooley, J. *et al.* Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resour Manage* 32, 4527–4542 (2018). https://doi.org/10.1007/s11269-018-2061-y.

[18] Huang, H., Zhang, Z. & Song, F. An Ensemble-Learning-Based Method for Short-Term Water Demand Forecasting. *Water Resour Manage* 35, 1757–1773 (2021). https://doi.org/10.1007/s11269-021-02808-4

[19] S. L. Zubaidi, H. Al-Bugharbee, Y. R. Muhsen, K. Hashim, R. M. Alkhaddar and W. H. Hmeesh, "The Prediction of Municipal Water Demand in Iraq: A Case Study of Baghdad Governorate," 2019 12th International Conference on Developments in eSystems Engineering (DeSE), Kazan, Russia, 2019, pp. 274-277, doi: 10.1109/DeSE.2019.00058.

[20] Haque, M.M.; Rahman, A.; Hagare, D.; Chowdhury, R.K. A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting. *Water* 2018, *10*, 419. https://doi.org/10.3390/w10040419

[21] Wu, S.; Han, H.; Hou, B.; Diao, K. Hybrid Model for Short-Term Water Demand Forecasting Based on Error Correction Using Chaotic Time Series. *Water* 2020, *12*, 1683. https://doi.org/10.3390/w12061683

[22] Zubaidi, S.L.; Ortega-Martorell, S.; Al-Bugharbee, H.; Olier, I.; Hashim, K.S.; Gharghan, S.K.; Kot, P.; Al-Khaddar, R. Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study. *Water* 2020, *12*, 1885. https://doi.org/10.3390/w12071885

[23] Zubaidi, S.L.; Abdulkareem, I.H.; Hashim, K.S.; Al-Bugharbee, H.; Ridha, H.M.; Gharghan, S.K.; Al-Qaim, F.F.; Muradov, M.; Kot, P.; Al-Khaddar, R. Hybridised Artificial Neural Network Model with Slime Mould Algorithm: A Novel Methodology for Prediction of Urban Stochastic Water Demand. *Water* 2020, *12*, 2692. https://doi.org/10.3390/w12102692.

[24] Mazwin Arleena Masngut , Shuhaida Ismail , Aida Mustapha, Suhaila Mohd Yasin. Comparison of daily rainfall forecasting using multilayer perceptron neural network model. IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 9, No. 3, September 2020, pp. 456~463 ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i3.pp456-463.

[25] Nahid Ferdous Aurna, Md. Tanjil Mostafa Rubel, Tanveer Ahmed Siddiqui, Tajbia Karim, Sabrina Saika, Md. Murshedul Arifeen, Tasmima Noushiba Mahbub, S. M. Salim Reza, Habibul Kabir.Time series analysis of electric energy consumption using autoregressive integrated moving average model and Holt Winters model. TELKOMNIKA Telecommunication, Computing, Electronics and Control Vol. 19, No. 3, June 2021,DOI: 10.12928/TELKOMNIKA.v19i3.15303.

[26] Redha Ali Al-Qazzaz, Suhad A. Yousif. High performance time series models using auto autoregressive integrated moving average. Indonesian Journal of Electrical Engineering and Computer Science Vol. 27, No. 1, July 2022, pp. 422~430 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v27.i1.pp422-430.

[27] Asmaa Y. Fathi1, Ihab A. El-Khodary, Muhammad Saafan. Integrating singular spectrum analysis and nonlinear autoregressive neural network for stock price forecasting, IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 11, No. 3, September 2022, pp. 851~858 ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i3.pp851-858.

[28] Rani Puspita, Lili Ayu Wulandhari. Hardware sales forecasting using clustering and machine learning approach, IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 11, No. 3, September 2022, pp. 1074~1084 ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i3.pp1074-1084

[29] Kwame Boateng, Machine Learning in Cybersecurity: Intrusion Detection and Threat Analysis , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[30] Christopher Davies, Matthew Martinez, Catalina Fernández, Ana Flores, Anders Pedersen. Machine Learning Approaches for Predicting Student Performance. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/174

[31] Kumbhkar, M., Shukla, P., Singh, Y., Sangia, R. A., & Dhabliya, D. (2023). Dimensional reduction method based on big data techniques for large scale data. Paper presented at the 2023 IEEE International Conference on Integrated Circuits and Communication Systems, ICICACS 2023, doi:10.1109/ICICACS57338.2023.10100261 Retrieved from www.scopus.com