

ORIGINAL RESEARCH ARTICLE

A Comparative Study of Time Series and Decision Tree Models for Forecasting Water Levels on the River Benue, Nigeria

Nura Umar^{1,2*}  and Alison Gray² ¹Department of Statistics, Umaru Musa Yar'adua University, PMB 2218, Katsina, Nigeria²Department of Mathematics and Statistics, University of Strathclyde, Glasgow, United Kingdom

ABSTRACT

The river Benue is vulnerable to flood risks, partly due to the release of water from the Lagdo Dam in Cameroon into Nigeria, as well as high precipitation, resulting in substantial damage and economic losses. Improved flood event prediction is crucial for decision-makers and the population to effectively plan strategies for reducing flood-related losses. This paper presents a comparative study using time series SARIMA and decision tree models applied to monthly water level data for 2011-2016 from Ibi, Makurdi, and Umaisha water stations on the river Benue. Granger causality and correlation tests indicate that water levels at a station closer to the river source are significant in predicting water levels at a station downstream for the decision tree models. Two accuracy metrics, namely mean absolute percentage error (MAPE) and root mean square error (RMSE), were used to assess the models. The prediction results show that the SARIMA (4,0,2)(1,0,1) model is the best choice for forecasting the Ibi station water levels, closely followed by the decision tree. For the Makurdi water station, the decision tree model including the Ibi station water level among the predictors, is best. Finally, for predicting the Umaisha station water level, two decision tree models are best, including the Ibi water level or the Makurdi and Ibi water levels among the predictor variables.

ARTICLE HISTORY

Received February 03, 2025

Accepted April 30, 2025

Published June 02, 2025

KEYWORDS

decision tree, flooding, prediction, river Benue, time series, SARIMA, water level



© The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License [creativecommons.org](https://creativecommons.org/licenses/by-nc/4.0/)

INTRODUCTION

Flooding, one of the world's leading natural disasters, affects millions of lives and properties globally. Peduzzi *et al.* (2009) report that over 800 million people live in flood-prone areas worldwide, and about 70 million people are exposed to flooding annually. This leaves a significant population at risk and underscores the widespread vulnerability to flood-related hazards, often worsened by climate change, causing extreme rainfall events, rapid urbanisation, and insufficient infrastructure. Addressing this growing challenge requires global efforts to improve flood management strategies, implement resilient infrastructure, and mitigate the risks posed to vulnerable communities.

The impacts of natural disasters such as flooding, drought, earthquakes, and landslides are generally measured by quantifiable means. Drought, flooding, and landslides have devastating impacts in Africa (Lumbroso *et al.*, 2016). However, in the last decade or so, flooding has been the most frequently occurring natural hazard in Africa (Lumbroso, 2020). In the last 50 years, Nigerians have experienced notable flood incidences in 2012, 2018 (NIHSA, 2020), and 2022 (Adesola *et al.*, 2024); however,

this does not mean that flooding in other years was not also catastrophic, but rather that its impact was not as great.

The negative impacts of flooding on the people and economy of Nigeria are enormous. Umar and Gray (2022) found that between the years 2011-2020, Nigeria recorded about 1,187 deaths, comprising 15% of Africa's deaths due to flood incidences in that time period, while the value of damage to properties was recorded as \$904.5 million, 21% of flood-related property damage in Africa. The north of Nigeria is especially vulnerable to flooding as the rivers Benue and Niger are situated in the northern region, while Lagdo Dam in Cameroon releases water into Nigeria through the river Benue, which often causes flooding. The northern region has more and larger states than the south, i.e., 19 states and the Federal Capital Territory (FCT), together covering a larger area than the 17 states in the southern part of the country. From the 2016 population projections by the National Population Commission, derived from the 2006 census figures, it was reported that 54% of the country's population resides in the northern region (NBS, 2017), which is vulnerable to the risks of flooding.

Correspondence: Nura Umar. Department of Statistics, Umaru Musa Yar'adua University, PMB 2218, Katsina, Nigeria. ✉ umar.nura@umyu.edu.ng.

How to cite: Umar, N., & Gray, A. (2025). A Comparative Study of Time Series and Decision Tree Models for Forecasting Water Levels on the River Benue, Nigeria. *UMYU Scientifica*, 4(2), 122 – 134. <https://doi.org/10.56919/usci.2542.015>

Accurate prediction of water levels is a key requirement for authorities before any decision-making and planning, and the main purpose of water level prediction is to establish accurate prediction models and to reveal the changing nature of rivers/basins/lakes (Xu *et al.*, 2019). This can inform waterborne transportation, water resources management, environmental management, flood mitigation, and emergency response. Information provided by such water level forecasts has the potential to reduce the effects of floods and mitigate and prevent disasters through foreknowledge and the timely allocation of resources for flood prevention or management.

In recent years, the autoregressive integrated moving average (ARIMA) time series model has been applied to forecast water levels for some river water stations. Yu *et al.* (2017) proposed an ARIMA model for daily water level prediction at three Yangtze River stations in China. The performance of the model was measured using root mean square error (RMSE), mean absolute percentage error (MAPE), percent of bias, and index of agreement. While it was concluded that the proposed model is a good candidate for short-term forecasting, its accuracy in predicting water levels decreased as the forecasting period increased. Hence, the work suggested incorporating other algorithms to overcome this drawback for longer-term forecasting. Arbain and Wibowo (2012) compared the performance of ARIMA and Artificial neural network (ANN) methods to predict water levels for the Dungun River in Terengganu, Malaysia, using mean square error (MSE) as a measure of accuracy, and concluded that the ANN gives better forecasting results because it can identify patterns and nonlinear characteristics of time series.

Lin and Watanabe (2017) proposed and compared some machine learning models, including k nearest neighbour (kNN), support vector regression (SVR), and linear regression, for water level forecasts using two water stations from the Ayeyarwady River, Myanmar. For the three prediction performance measures used, i.e., correlation coefficients, mean absolute error (MAE), and RMSE, the kNN model achieved the best predictions. Choi *et al.* (2019) developed a water level forecast model using four machine learning approaches for Upo wetland, South Korea, namely an ANN, a decision tree, random forests (RF), and a support vector machine (SVM). Three variables, humidity, precipitation, and temperature were used in the study to predict water levels. Comparison between the machine learning techniques revealed that the RF gave the lowest RMSE, better than the ANN model which other authors adopted. The authors suggested that in the absence of real data on water levels, the study's results could be used to develop wetland management techniques.

Yan and Ma (2016) combined an ARIMA model and a radial basis function network (RBFN) to predict monthly groundwater levels using observations from two wells in the city of Xi'an, China. The combined model performed better than either of the separate ARIMA and RBFN models, using various accuracy measures, and hence was

recommended for fitting and predicting groundwater levels. Xu *et al.* (2019) also proposed a new water level prediction method by combining an ARIMA model, which accounts for the linear component of the data, and a Recurrent Neural Network (RNN), which covers nonlinear aspects of the data. The applicability of the model was shown using daily water level and environmental data for Taihu Lake, China. Using RMSE, the combined ARIMA-RNN model was found to be better than the individual ARIMA and RNN models. In a related work, Phan and Nguyen (2020) proposed a hybrid model to increase water level prediction accuracy by combining machine learning models and ARIMA. The model predicts water level by including the lags of the residuals from the ARIMA model among the independent variables in the machine learning models. Datasets from Vu Quang, Hanoi, and Hung Yen water stations on the Red River, Vietnam, were used to test the proposed model. It was concluded that the hybrid model performed better than the single models and also better than a different hybrid model of Zhang (2003).

For Nigeria, while several studies have examined time series models for rainfall, Nwobi-Okoye and Igboanugo (2013) for the first time predicted water levels at the Kainji Dam using ARIMA time series models and artificial neural networks. The ARIMA model gave the lowest prediction error and was recommended for future water level predictions. However, the study did not consider seasonal patterns in water levels and did not use SARIMA models, although these models can incorporate seasonal effects and, therefore, could lead to improved predictions.

With the hypothesis that seasonality in water levels is important, this work examines and compares time series SARIMA models and decision tree models for water level forecasting to inform flood mitigation efforts in Nigeria. The performance of these models will be examined using water level data for three water stations on the river Benue, namely Ibi, Makurdi and Umaisha water stations, in the vulnerable northern region of Nigeria.

METHODOLOGY

Study region

The study region, comprising Ibi, Makurdi, and Umaisha water stations on the river Benue (Figure 1), is located in northern Nigeria, and the river cuts across various states in the northeast and northcentral regions, before its confluence with the river Niger at Lokoja in Kogi state. There are 194 water monitoring stations in Nigeria, located along the rivers and dams, used to monitor water movements and levels. Data availability was examined for these 194 stations using the Nigeria Hydrological Services Agency (NIHSA) records. Most water stations have no records from 1980 onwards and some have very few years of records in total. The data gaps at these water stations are mostly caused by vandalised equipment and/or faulty recording instruments. With available data, Ibi, Makurdi, and Umaisha water stations were selected as suitable stations for this study.

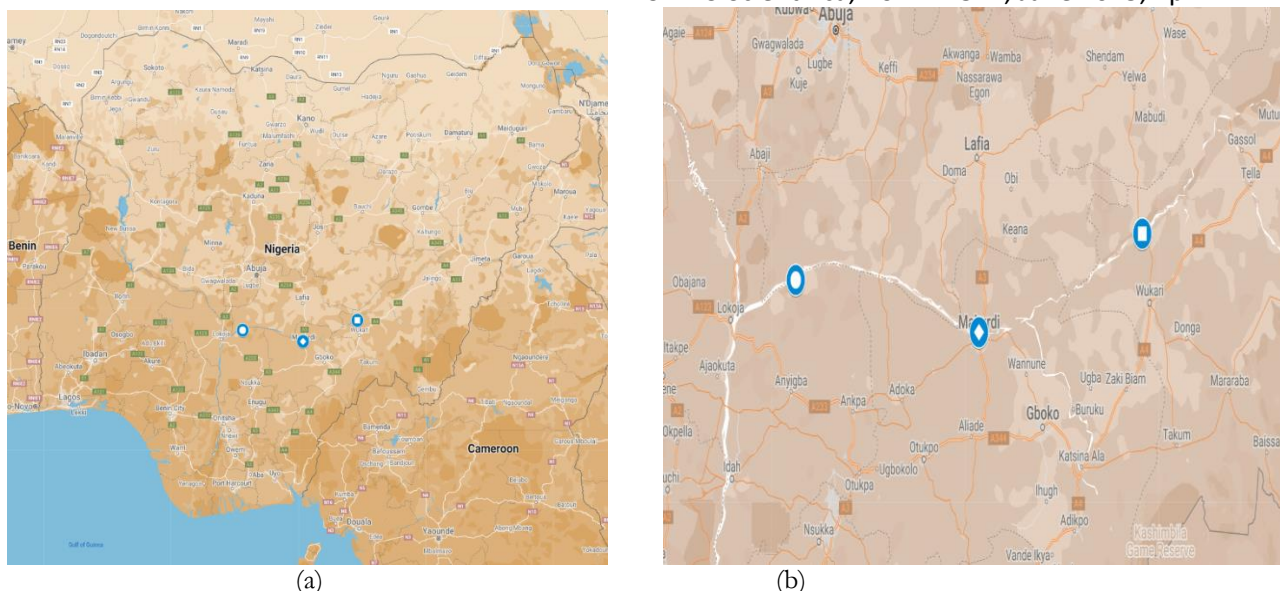


Figure 1: (a) Map of Nigeria showing the locations of the 3 studied water stations on the river Benue (marked left to right as: Ibi (circle), Makurdi (diamond), and Umaisha (square)); and (b) enlarged version of the map in (a) showing the locations of the stations on the river more clearly; Source: Google Maps, modified by the authors

Water level datasets

Water level data were obtained from NIHSA for the Ibi, Makurdi, and Umaisha stations. The data considered for the analysis is monthly from January 2011 to December 2016 (Table 1). The time periods were chosen to give an equally long time series for the data for the three water stations, for meaningful analysis. From the study period (2011-2016), Makurdi station has 49 missing monthly values, a proportion of 68% of observations missing, the highest among the three water stations, followed by Ibi and Umaisha stations with 39% and 32% of the data values missing, respectively (Table 1).

Due to the high proportions of missing data, missing values were imputed before further analysis. The predictive mean matching (PMM) imputation method is an attractive and popular imputation method for missing values, particularly when dealing with quantitative variables (Allison, 2015), and was used here. Oyerinde *et al.* (2021) used the PMM method to impute missing data from 22 water stations in the Niger basin and recommended it for imputing data gaps. Other studies where the PMM method outperformed competing imputation methods for missing data include Vink *et al.* (2014) and Akman *et al.* (2019).

The R software was used for data analysis (R Core Team, 2022). We imputed all of the missing values using the R package “mice” (van Buuren and Groothuis-Oudshoorn, 2011). Using the PMM method, we generated five multiple imputed values for each missing value, using 1000 iterations (repeatedly updating the imputed values for

missing data in a dataset until convergence is reached), having observed in experimentation that there was no difference in the results using 1000 iterations of the imputed values or more than 1000. The iterative imputation process refines estimates of missing values using available data and the current model. Increasing the number of iterations can enhance the algorithm's convergence and lead to improved imputation outcomes. The average of each of these sets of five values generated for a given missing value was chosen to represent this missing value in the same way for each station. The completed datasets for the Ibi, Makurdi, and Umaisha water levels are presented in Figure 2.

Figure 2 shows clear peaks and troughs in the water levels for the Ibi water station and a repeating pattern for each year. As well as this seasonal pattern, there is some slight evidence of a decreasing trend over time. Makurdi station has a similar repeating pattern, with high peaks in 2012, 2014, and 2016, but without an obvious trend. Similarly, for the Umaisha water station, a clear seasonal pattern is seen, with a more obvious decreasing trend. The water levels in each dataset take a similar range of values. These data series are seasonal, with a decreasing trend seen in at least the Umaisha dataset, so they are not stationary.

As a result of this clear seasonality, the SARIMA model, defined below, is expected to be more suitable for modelling these series than the more standard ARIMA time series model. The performance of SARIMA models and decision tree models is compared using the metrics below.

Table 1: Numbers and proportions of missing values in the provided data for the selected water stations.

Station	Records available	Study period in years	No. of months available within study period	No. of months missing within study period (%)
Ibi	1980-2016	2011-2016	44	28 (38.89)
Makurdi	2010-2019	2011-2016	23	49 (68.06)
Umaisha	2011-2019	2011-2016	49	23 (31.94)

Source of data: NIHSA, 2020.

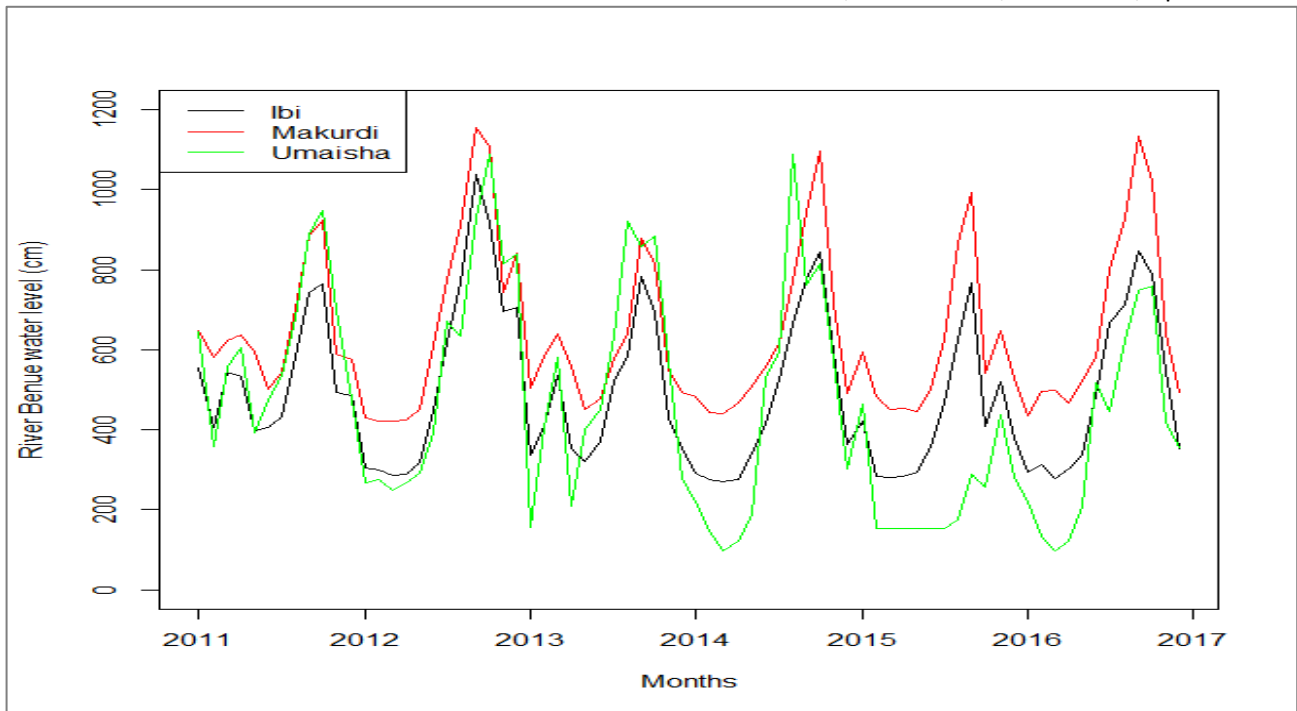


Figure 2: Time plots of monthly water levels at Ibi, Makurdi, and Umaisha water stations on the river Benue for the years 2011 to 2016, including imputed data values

Accuracy metrics

Mean absolute percentage error (MAPE) and root mean square error (RMSE) are both used to evaluate the performance of the fitted models, as using multiple accuracy metrics can help to draw more robust conclusions concerning performance of the studied models. These metrics are given as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100, \quad \text{and} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}, \quad (2)$$

where x_i and $\hat{x}_i$ are the actual or observed value and predicted value, respectively, for the i^{th} observation, and n is the sample size. Lower values of these metrics (closer to 0) correspond to greater prediction accuracy.

THEORETICAL BACKGROUND

Time series models

This section will briefly describe time series models, which are parametric statistical methods, leading to the SARIMA model (Brockwell and Davis, 2016) that will be used in the analysis.

Autoregressive models

A time series $\{x_t\}$ is said to be an autoregressive (AR) series of order p or an $AR(p)$ process if it can be written as

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t, \quad (3)$$

where p is a nonnegative integer, $\phi_0, \phi_1, \phi_2, \dots, \phi_p$ are real numbers, and e_t is a white noise with mean zero and variance σ_e^2 . Equation (3) is similar to a multiple linear regression model, with lagged values serving as the independent variables (Tsay, 2010).

Moving Average models

A stationary time series $\{x_t\}$ is a moving average of order q or $MA(q)$ if

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}, \quad (4)$$

where q is a nonnegative integer, $\theta_1, \theta_2, \dots, \theta_q$ are constants, e_t is an error term and $e_t \sim N(0, \sigma_e^2)$. If $\{x_t\}$ has a non-zero mean, it is advisable to subtract the mean from the data before fitting the model. This adjustment is akin to including a constant/intercept term in equation (4) (Shumway and Stoffer, 2017).

Autoregressive Moving Average models

A stationary series $\{x_t\}$ is an autoregressive moving average of order p and q or $ARMA(p, q)$ if it can be written as

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}, \quad (5)$$

where p and q are nonnegative integers, ϕ_i and θ_i are the AR and MA coefficients, respectively, for the i^{th} observation, and e_t is an error term.

As for the MA model, Shumway and Stoffer (2017) indicate that when the mean of $\{x_t\}$ is non-zero, the mean of the data should be subtracted first, and this is equivalent to not subtracting the mean but including a constant/intercept in the ARMA model.

If $p = 0$, the ARMA series becomes MA (q), and if $q = 0$, the series is an AR (p) model as described earlier. One limitation of ARMA series is satisfying the stationarity condition, and if the series is nonstationary, it is recommended to difference the series to make it stationary (Zhang, 2018).

Autoregressive Integrated Moving Average models

A time series $\{x_t\}$ is an autoregressive integrated moving average (ARIMA) process or ARIMA (p, d, q) series if it becomes an ARMA (p, q) after differencing the original series d times (Brockwell and Davis, 2016). It is written in terms of the polynomials from the backshift operator B as follows:

$$x_t \sim \text{ARIMA}(p, d, q) \Leftrightarrow \phi(B)(I - B)^d x_t = \theta(B)e_t, \quad (6)$$

where d is the order of differencing and other parameters are defined as before (Carmona, 2014).

Seasonal Autoregressive Integrated Moving Average models

A time series $\{x_t\}$ is said to be a SARIMA (p, d, q) $\times$ (P, D, Q) process with period s if the differenced series $y_t = (I - B)^d(I - B^s)^D x_t$ is an ARMA process. The SARIMA model is an extension of the ARIMA model, while the ARIMA model is a special case of the SARIMA model with no seasonal effect (Brockwell and Davis, 2016). The SARIMA model is given as

$$\phi(B)\Phi(B^s)y_t = \theta(B)\Theta(B^s)e_t, \quad (7)$$

where $\phi(B)$ and $\theta(B)$ are the p th and q th degree polynomials of the non-seasonal components, $\Phi(B^s)$ and $\Theta(B^s)$ are the P th and Q th degree polynomials of the seasonal components, and d and D are the non-seasonal and seasonal orders of differencing respectively. If no

differencing is used, then the model becomes a SARMA(p, q) $\times$ (P, Q) model.

SARIMA models are used to analyse and forecast time series data exhibiting seasonal patterns, where the values tend to repeat in a regular cycle. The seasonal component represents the repeating pattern over fixed intervals, such as days, months, or quarters. In many real-world time series datasets, the seasonal patterns are not perfectly identical from one cycle to the next, due to various factors such as changes in consumer behaviour, weather conditions, or economic fluctuations. These models capture randomness and fluctuations in the seasonal pattern from one cycle to another by incorporating seasonal differencing and seasonal autoregressive and moving average terms. Using these seasonal terms, the model can account for variability in the seasonal patterns over time, making the forecasts more accurate and reflective of the actual data behaviour. Overall, SARIMA models are useful for understanding and predicting time series data with recurrent patterns, allowing for flexibility in modelling variations in the seasonal component (Brockwell and Davis, 2016).

Decision trees

A decision tree (DT) is a popular non-parametric (distribution-free) machine learning algorithm that can be used for both classification and regression modelling (James *et al.*, 2013; Maimon and Rokach, 2014), depending on the application. It has a tree-like branching structure, with three components, namely a root node, decision nodes and leaf or terminal nodes (Figure 3). A DT algorithm divides a training dataset into parts (branches), which further sequentially segregate into other branches. This sequential division continues until a leaf node is produced, which cannot be divided further.

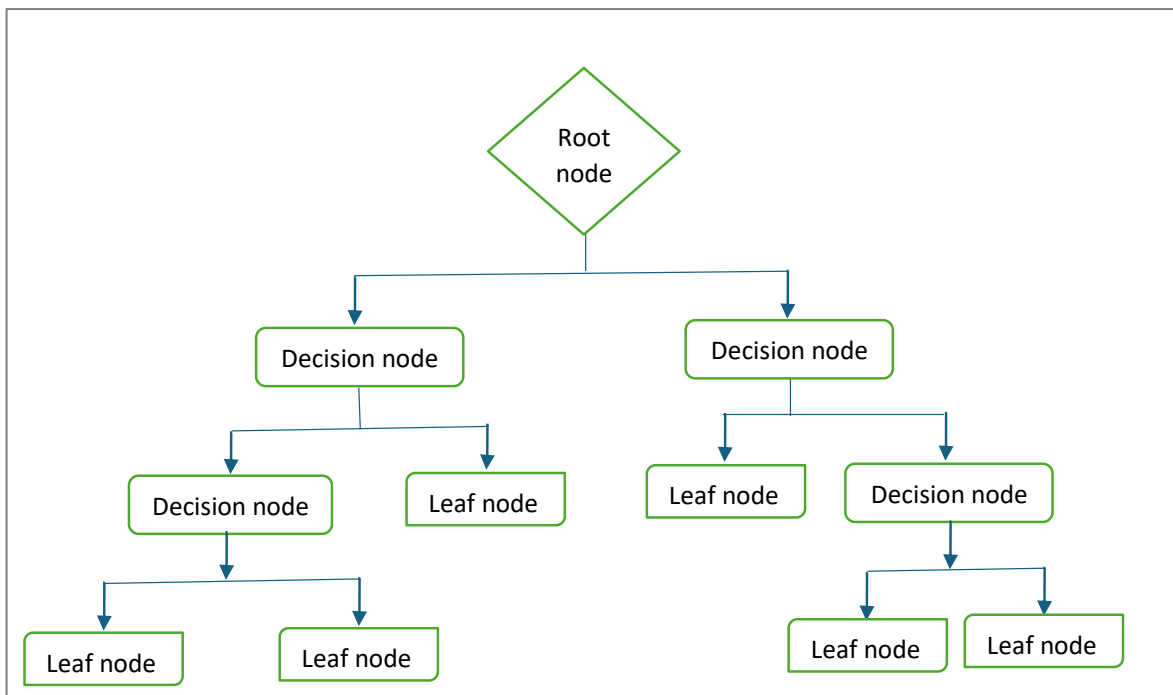


Figure 3: An example decision tree with its three different types of nodes

The topmost decision node in the tree is the root node, where the algorithm assesses a specific feature to partition the dataset into two subsets. Each subset corresponds to distinct values or ranges of values of the chosen feature. Branching out from the root node, smaller nodes represent subsets based on the initial decision and function as decision nodes each testing values of a feature variable. The tree continues to grow until it reaches a fixed value of a stopping criterion such as information gain or Gini impurity, a predefined tree depth or a minimum number of samples in a leaf node (Hastie *et al.*, 2009). Each leaf node contains a subset of training examples that have similar feature values, and to predict the target value for a new observation, the leaf node is found that the observation falls into, based on its values of the feature variables. When using a regression DT for the prediction of a continuous value, the mean of the target values of the training examples in the leaf node is used as the prediction (James *et al.*, 2013).

DTs are intuitive to understand and interpret and can capture complex, nonlinear relationships in data. However, they can also be prone to overfitting, especially if the tree becomes too deep or the data are noisy, limiting the generalisability of the tree beyond the training data used to grow the tree. To mitigate this, techniques such as pruning, which removes parts of the tree that do not provide significant predictive power or generalisation to new, unseen data (Hastie *et al.*, 2009), can be used. A pruned tree with fewer splits may lead to lower variance and better interpretation, with lower bias.

The R software's tree package (Ripley, 2023) was used here to fit the DTs, with and without pruning. The deviance criterion was used to grow the tree. The integer parameter “best” in the R tree function sets the number of leaf/terminal nodes of a specific subtree in the cost-complexity sequence to be returned. If there is no tree in the sequence of the requested size, the next largest is returned (Hastie *et al.*, 2009).

Granger causality test

Time series models use past time series values to forecast future values. Machine learning models typically use many variables to predict future values of a target variable of interest, in this case, water level. Before choosing predictor variables for the DT, we consider relationships among the water levels at the three water stations.

The Granger causality test is a statistical hypothesis test mostly used in econometrics to determine whether one time series can be used to predict another in the sense that one time series variable (the potential cause or predictor) can be considered as a leading indicator of another series variable (the potential effect or outcome). The test is commonly used to analyse cause-and-effect relationships in time series data (White and Pettenuzzo, 2014).

Here we consider the Pearson correlation between the water level time series, but also use the Granger causality test to determine the presence and direction of causation between the water levels. In general, the direction of causality could be two-way (sometimes referred to as

“feedback”), one-way, or there could be no causation (Granger, 1969). The null hypothesis (H_0) in a Granger causality test states that the predictor variable does not Granger-cause the response variable, while the alternative (H_1) is that the predictor variable Granger-causes the response variable.

Granger (1969) examines a bivariate series of a vector autoregressive (VAR) model. Granger causality corresponds to nonzero entries in the autoregressive coefficients. The bivariate model consists of two stationary time series x_{t1} and x_{t2} ,

$$\begin{aligned} x_{t1} &= \sum_{k=1}^d a_k x_{t1-k} + \sum_{k=1}^d b_k x_{t2-k} + e_{x_{t1}}, \\ x_{t2} &= \sum_{k=1}^d l_k x_{t1-k} + \sum_{k=1}^d m_k x_{t2-k} + e_{x_{t2}}, \end{aligned} \quad (8)$$

where $e_{x_{t1}}$ and $e_{x_{t2}}$ are taken to be two independent white noise series, and d is presumed to be finite and smaller than the length of the given data. If any $b_k \neq 0$, the series x_{t2} is considered to Granger-cause the series x_{t1} . Similarly, if any $l_k \neq 0$, then x_{t1} Granger-causes x_{t2} . When both conditions are met, it indicates a feedback relationship between x_{t1} and x_{t2} . We test whether all of these coefficients are zero against the alternative hypothesis that at least one coefficient is non-zero by testing one nested linear model within another using an F-test implemented in the `grangertest` function from the R `lmtest` package, and specifying order 4 as the maximum number of lags to consider. While the water level datasets used here show evidence of non-stationarity, the correlation analysis gives a check on the results of the Granger test analysis.

RESULTS

Examining the data for fitting SARIMA models

The most suitable orders of time series models for each of the Ibi, Makurdi, and Umaisha water level datasets were determined using partial autocorrelation function (PACF) and autocorrelation function (ACF) plots (shown in Figure A.1 in the Appendix) for each dataset, using the complete datasets. The PACF plots suggest the order p of an AR model, while the ACF plots suggest the order q of a MA model. The ACF plots also show clear evidence of seasonality in all three datasets. On the basis of these plots, the following models were selected for fitting to the data: SARIMA(4,0,2)(1,0,1) and SARIMA (4,0,8)(1,0,1) for Ibi station, SARIMA(2,0,2)(1,0,1) and SARIMA(2,0,8)(1,0,1) for Makurdi station, and SARIMA(1,0,2)(1,0,1) and SARIMA(4,0,7)(1,0,1) for Umaisha station. In fact these are all SARMA models.

The identified SARIMA models and DT models will be fitted to the data from these water stations to identify the best fitting and best predicting models in each case.

Examining the data for fitting decision trees

Prior to fitting the DT model, the complete, whole datasets were used to calculate Pearson correlations and also carry out Granger causality tests from the R software

lmtest package (Zeileis and Hothorn, 2002) on the water levels from the Ibi, Makurdi and Umaisha water stations.

From the correlation results (Table 2), there are strong positive linear associations between the water levels from all three stations, meaning that as the water level for one station increases, so that of the other stations linearly increases, which makes sense in this context. The correlation between Ibi and Makurdi water levels is the highest (0.9713), and Makurdi station lies beyond Ibi station on the river Benue, while the correlation between Makurdi and Umaisha water levels is the lowest (0.7386), although Umaisha station follows Makurdi station on the map. Surprisingly, the correlation between Ibi and

Umaisha water levels is higher (0.8370) despite these water stations being further apart than Makurdi and Umaisha stations. The very low p-values (all less than 0.0001) of the correlations indicate that all the correlations are highly statistically significant (at a level of 5% or lower). Therefore, the water level at one water station nearer to the source of the river could be important for predicting the water level at another station further along the river, and so these are considered as potential predictor variables.

Table 3 shows the results of pairwise Granger causality tests.

Table 2: Pairwise Pearson correlations between the Ibi, Makurdi, and Umaisha water level datasets. In all cases the p-values are <0.0001.

Station	Ibi	Makurdi	Umaisha
Ibi	1.0000	0.9713	0.8370
Makurdi	0.9713	1.0000	0.7386
Umaisha	0.8370	0.7386	1.0000

Table 3: Granger causality test null hypothesis, F test statistic values and p-values for Ibi, Makurdi and Umaisha water level datasets. A p-value of 0.05 or less is taken as significant.

Null hypothesis (H_0)	F-values	p-values
Ibi does not Granger-cause Makurdi	3.4831	0.0128
Makurdi does not Granger-cause Ibi	1.4126	0.2410
Ibi does not Granger-cause Umaisha	1.5232	0.2071
Umaisha does not Granger-cause Ibi	0.9323	0.4515
Makurdi does not Granger-cause Umaisha	0.9849	0.4229
Umaisha does not Granger-cause Makurdi	1.6793	0.1669

Table 4: Accuracy measures of SARIMA and Decision tree (DT) models fitted using training data and evaluated on the training and test data, for the water stations on the river Benue; the bold text indicates the best model in each case. The order $s=12$ was used for the SARIMA models as the data is monthly.

Station	Model	Training dataset		Test dataset	
		MAPE	RMSE	MAPE	RMSE
Ibi	SARIMA (4,0,2)(1,0,1)	15.84	95.52	16.40	98.05
	SARIMA (4,0,8)(1,0,1)	13.70	79.81	14.15	78.83
	DT (unpruned)	13.73	92.14	19.90	132.12
	DT (pruned)	-	-	17.63	123.71
Makurdi	SARIMA (2,0,2)(1,0,1)	13.39	107.88	14.61	117.17
	SARIMA (2,0,8)(1,0,1)	12.82	101.23	12.99	123.94
	DT without Ibi	11.87	107.80	12.59	117.48
	DT with Ibi (unpruned)	6.32	51.12	5.22	48.25
	DT with Ibi (pruned)	-	-	5.76	49.79
Umaisha	SARIMA (1,0,2)(1,0,1)	38.02	147.79	93.82	276.55
	SARIMA (4,0,7)(1,0,1)	31.52	132.86	153.51	342.88
	DT without Ibi and Makurdi	22.94	114.38	46.86	160.34
	DT with Makurdi	26.45	114.07	42.70	189.34
	DT with Ibi	21.70	98.72	39.91	198.49
	DT with both (unpruned)	21.70	98.72	39.91	198.49
	DT with both (pruned)	-	-	43.89	193.69

From Table 3, the only significant p-value is $p=0.0128$, meaning that the corresponding H_0 is rejected and the Ibi water level Granger-causes the Makurdi water level, however the Makurdi water level does not Granger-cause the Ibi water level ($p=0.2410$). This means that the causality between these two variables is one-directional, which is expected as the Ibi station is nearer the river source than the Makurdi station (Figure 1). Table 3 also

shows that the Umaisha water level has no Granger-causal relationship with any of the water levels before it on the river (from Ibi and Makurdi stations), however the Ibi and Makurdi water levels also do not Granger-cause the Umaisha water level, which is a less expected result since the Umaisha station is further away from the source of the river than the Ibi and Makurdi stations.

Machine learning models such as DTs generally use many predictor variables to predict the value of a different dependent variable. Here we use time series data of a single variable, which is water level. For the DT, variables used here to predict the water level at a given station at a given time are Month, Quarter, Year, and the lagged water level for lags 1-4 for each water level to be predicted. The maximum lag 4 was selected because it is the highest order of p from the SARIMA models fitted for all the water levels. Hence, for all the water stations, lag 1 up to and including lag 4 of the water levels will be included among the independent variables (Phan and Nguyen, 2020). Using more lags is recommended; however, using more lags reduces the available sample size. In some trees, the water level(s) from earlier water stations are also included as potential predictors based on the results from the Pearson correlations and the Granger causality tests (Tables 2 and 3).

Ibi water station is nearer to the river source than Makurdi and Umaisha water stations (Figure 1), so Ibi has no water level from a previous station included among its predictors. Makurdi station has the Ibi water level included among its predictors, and for Umaisha station, the water levels from both Ibi and Makurdi stations are included among the predictors. Ibi water station has only one model to be fitted. Makurdi station has two models to be fitted, one with and the other without the water level from Ibi station among the predictors. Finally, Umaisha has four models to be fitted, i.e., including none of the water levels from previous water stations among the predictors, including either of the Ibi and Makurdi water levels on its own, and including the water levels from both these stations as predictors.

Before model fitting and forecasting the water levels for each water station, we split the datasets into training and testing sets for the two rivers. We now use the four years of data from January 2011-December 2014 as training data, to develop suitable models for each water station, and use the two years of data from January 2015-December 2016 for testing. However, as a result of using the 4 lagged water level variables for each station, the datasets become imbalanced, so to have a balanced dataset for water level prediction, January-April 2011 (four rows) were lost across all variables for analysis. These results used May 2011-December 2015 as training data and January 2016-December 2016 as test data. For better comparison of the models, these same reduced datasets were also used for the time series models.

Table 4 gives the results comparing the time series models and the various DT models for all three water stations, using the training and test data.

For Ibi station, from Table 4, for the training data, the SARIMA (4,0,8)(1,0,1) model has the lowest values for both accuracy measures, i.e. 13.70 and 79.81 for MAPE and RMSE respectively, followed by the DT model with 13.73 and 92.14 for MAPE and RMSE respectively, and SARIMA (4,0,2)(1,0,1) is poorer for both accuracy metrics. For the test data, again the SARIMA (4,0,8)(1,0,1) model is best in terms of MAPE (14.15) and

RMSE (78.83), better than the SARIMA (4,0,2)(1,0,1) model, with MAPE=16.40 and RMSE=98.05, and the DT model is poorer. Pruning improves the performance of the DT model, but the pruned tree is still poorer than the SARIMA models. The SARIMA (4,0,8)(1,0,1) model performs best overall, for forecasting the Ibi station test data, and is by far the best model for prediction of both the training data and test data. However, as models with higher lag length suggest an increase in mean-square forecasting errors and model overfitting (Lütkepohl, 1993), in practice the SARIMA (4,0,2)(1,0,1) may be the better choice for forecasting the Ibi water level data. We now examine the fitted DT for the Ibi water level (Figure 4(a)), to identify the most important variables used in the construction of the tree.

The primary split is month <6.5, which indicates that month is the most important factor for predicting the Ibi water level, and months in the first half of the year give a lower water level than months in the second half, which may be attributed to the pattern of the raining season. Among months in the first half of the year, lag1 does affect the water level, where lag1>=311.85 leads to a higher predicted water level. For months in the second half of the year, lag4 <458.13 and lag1<523.41 give higher predicted water levels than for the earlier months, while lag1>=523.41 leads to a higher predicted water level, especially for lag3<458.13 (which is the highest level for this tree). For lag4>=458.13, years up to 2012 tend to give a higher water level than later years. The lag2 variable was not used in the tree. The Ibi unpruned tree (Figure 4(a)) has 7 leaf nodes, and it was pruned (Figure 4(b)) to have 5 leaf nodes (using parameter best =5).

Figure 4(b) shows the pruned tree for the Ibi station, which has similar features to the unpruned tree, with month <6.5 as the primary split. The major difference is that lag3, which was used for the unpruned tree (Figure 4(a)), is not used for the pruned tree.

For Makurdi station (Table 4), for the training and test data, the DT model, including the Ibi water level among the predictors, has much the lowest MAPE and RMSE (6.32, 51.12, respectively for the training data and 5.22 and 48.25, respectively for the test data). For the training data, the DT model without Ibi is second best for MAPE (11.87), although the SARIMA (2,0,8)(1,0,1) model is second best for RMSE (101.23). For the test data, the results are similar to these. It is clear that adding the Ibi water level among the variables to predict the Makurdi water level considerably improves the DT model performance for both metrics, reflecting the Granger-causal relationship (Table 3) and correlation (Table 2) above. Hence, the DT with Ibi is the best choice to model the Makurdi water levels. The unpruned tree for Makurdi with Ibi among the predictors (Figure 5(a)) has 5 leaf nodes, and it was pruned to have 4 leaf nodes (Figure 5(b)). For Makurdi, the pruned tree does slightly worse than the unpruned one, but the difference is small, and the tree is still the best prediction approach.

We now consider the unpruned fitted tree for the Makurdi water level with the Ibi water level among the predictors,

the better model in this case, and the pruned tree in Figures 5(a) and 5(b), respectively. The Ibi water level in fact is the only variable used in both trees for predicting

the Makurdi water level, and the higher the water level at the Ibi station, the higher the predicted Makurdi water level, which makes sense.

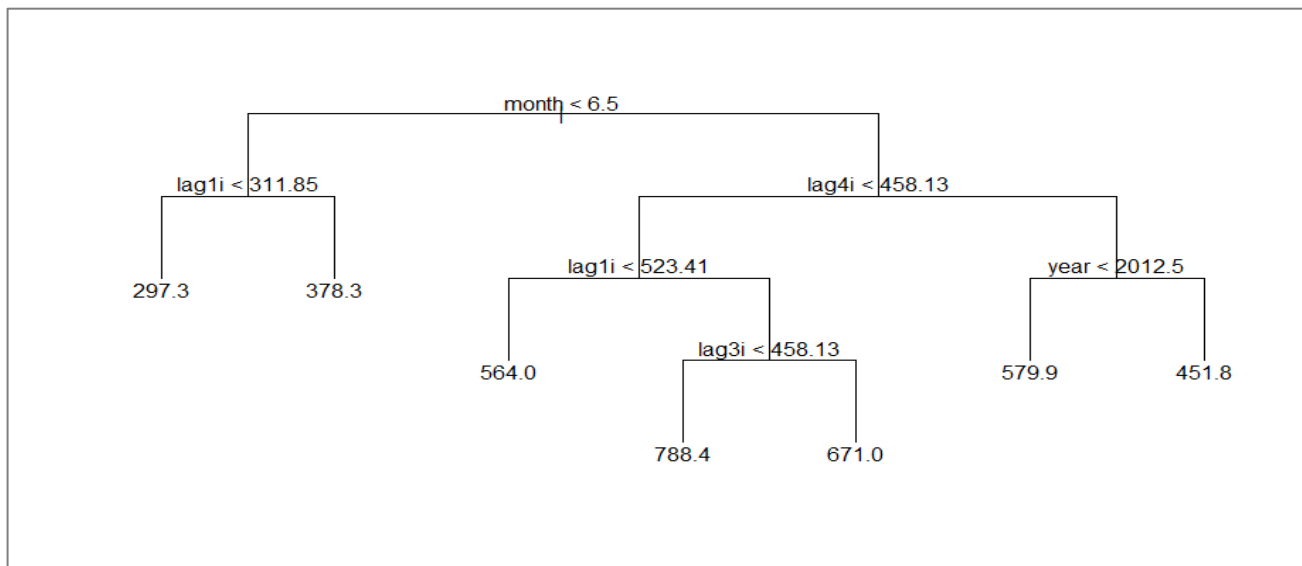


Figure 4(a): unpruned fitted regression tree for predicting the Ibi water level, using month, quarter, year, and lag 1-4 water levels as potential predictor variables.

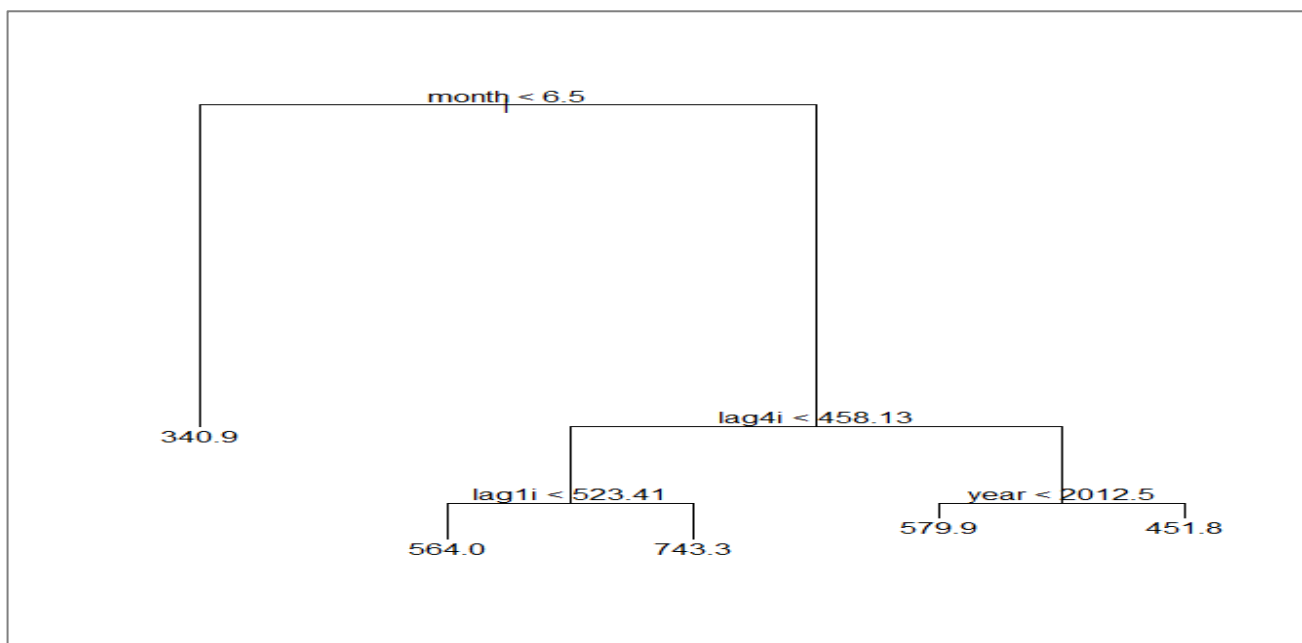


Figure 4(b): pruned fitted regression tree for predicting the Ibi water level, using month, quarter, year, and lag 1-4 water levels as potential predictor variables.

Finally, for the Umaisha station (Table 4), for both the training and test data, the MAPE and RMSE from the four DT models are notably better than those from the SARIMA models. Among these DT models, the models including Ibi or both Ibi and Makurdi water levels have the same lowest MAPE (21.70) and RMSE (98.72) for the training data. For the test data, the DT models including Ibi only and with both Makurdi and Ibi also have the same MAPE (39.91) and RMSE (198.49), which is the best in terms of MAPE, but the DT with Makurdi and not Ibi is slightly better in terms of RMSE and the DT without either is best for RMSE (160.34). Overall, including Ibi or both Ibi and Makurdi water levels among the independent variables is slightly better than including only the Makurdi

water level or neither, however for the test data this depends on whether RMSE or MAPE is considered. Among the time series models, the SARIMA (4,0,7)(1,0,1) model has the better MAPE (31.52) and RMSE (132.86) values for the training data, but the simpler SARIMA (1,0,2)(1,0,1) is clearly better for the test data, with a MAPE value of 93.82 and RMSE of 276.55. Neither time series model competes with the trees. The DT with Ibi or both Ibi and Makurdi included among the predictors is recommended for modelling the Umaisha water level data.

Next we consider the fitted tree for the Umaisha water level with Ibi and Makurdi water levels among the potential predictors, as adding the Ibi water level,

especially, to the predictor variables improves the performance.

Figure 6(a) shows that Ibi (water level) is the most important variable for predicting the Umaisha water level. For $Ibi < 368.045$ and years up to 2013, the predicted water level tends to be higher than for later years. For $368.045 \leq Ibi < 489.675$, lag3 does affect the Umaisha water level, and $lag3 \geq 457.838$ gives a higher predicted water level. For $Ibi \geq 489.675$, where $lag2 \geq 425.67$, and $lag4 < 502.01$, the predicted water level is higher and much the highest for the tree. The model with only the Ibi water

level added to the predictor variables and omitting the Makurdi water level gives the same tree as in Figure 6(a); therefore, the Makurdi level is not used. For Umaisha, the DT with Ibi and Makurdi among the predictor variables (Figure 6(a)) is considered for pruning. The tree has 7 leaf nodes and is pruned to have 5 leaf nodes, as shown in Figure 6(b). The year and lag3 variables are not used in the pruned tree. In this case, pruning the tree with both Ibi and Makurdi included slightly improves the RMSE accuracy (193.69) compared to the unpruned tree (198.49), but gives a slightly poorer MAPE (43.89 compared to 39.91), so overall, it makes little difference.

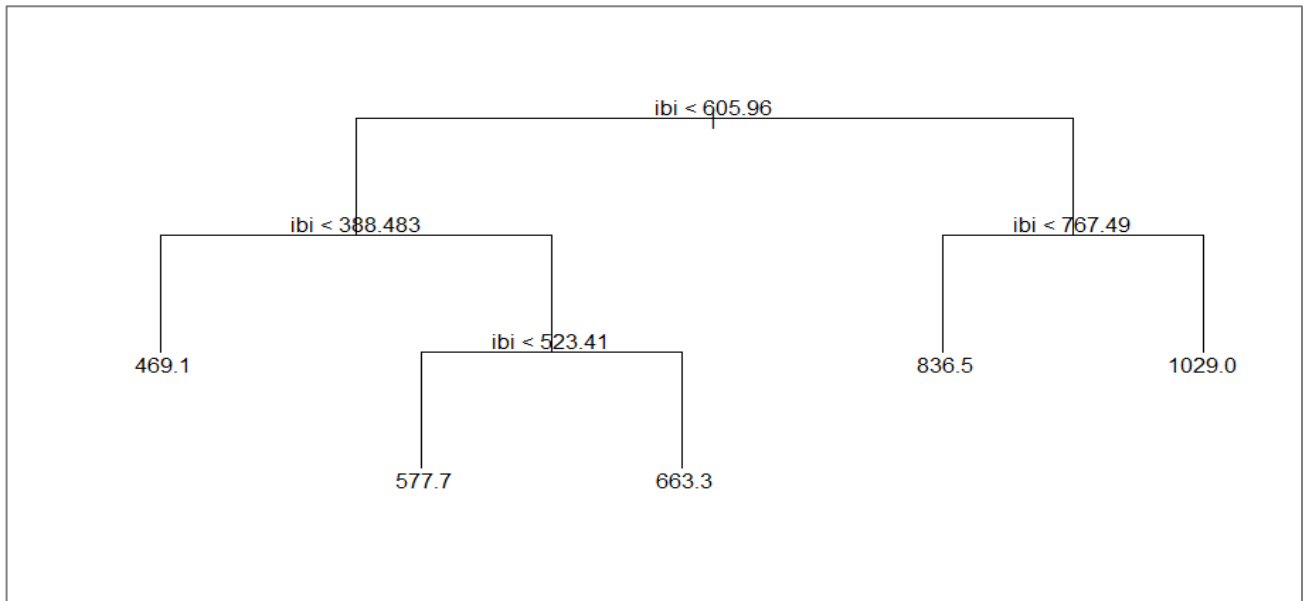


Figure 5(a): unpruned fitted regression tree for predicting the Makurdi water level, using month, quarter, year, Ibi water level, and lag 1-4 water levels as potential predictor variables.

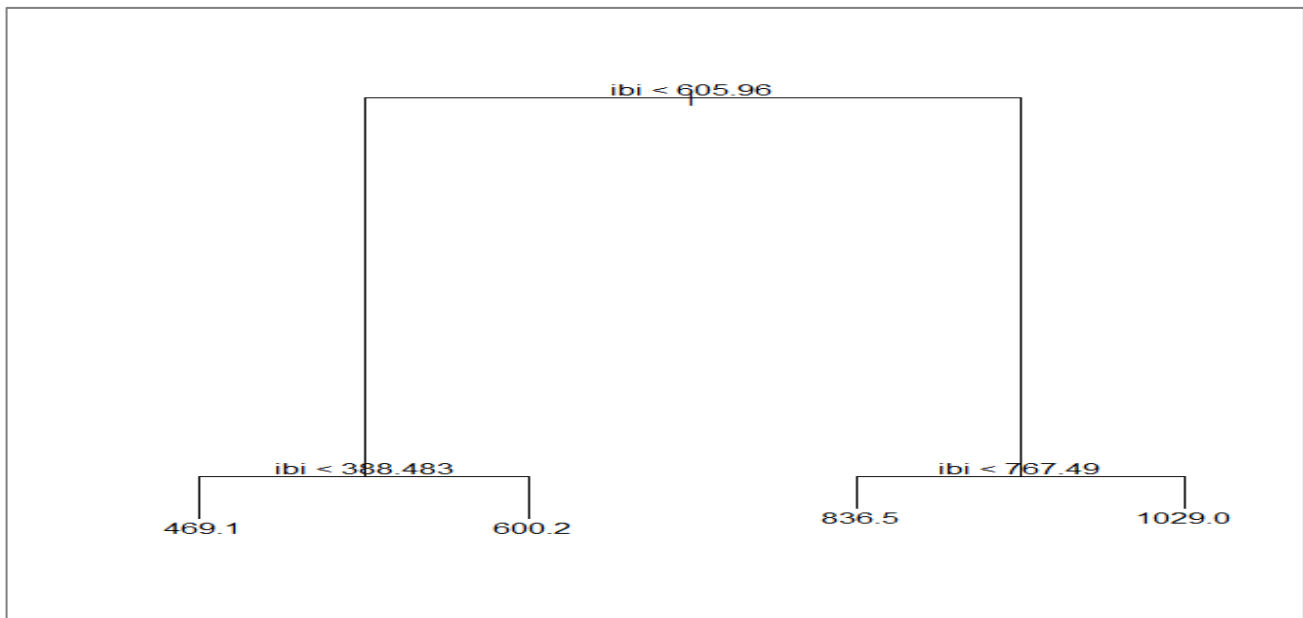


Figure 5(b): Pruned fitted regression tree for predicting the Makurdi water level, using month, quarter, year, Ibi water level, and lag 1-4 water levels as potential predictor variables.

CONCLUSIONS

In this work, we examined the water level forecasting performance of SARIMA time series models and decision tree models, using water level datasets from Ibi, Makurdi

and Umaisha water stations on the river Benue in Nigeria. Results from the fitted prediction models show that the SARIMA model is best for forecasting for Ibi water station, closely followed by the decision tree. Decision trees are by far the best for Makurdi and Umaisha water

stations. The results confirm that including the Ibi water level among the predictors for the Makurdi water level decision tree model is important. For predicting the Umaisha water level on unseen data, including the Ibi or Makurdi water levels or both among the predictors only improved prediction in terms of one accuracy measure considered (MAPE). As it makes sense to consider water levels nearer to the river source when predicting water

levels further from the source, this finding would be worth exploring further with other training and test sets and other accuracy measures. In view of the superior results overall for decision trees, it would also be useful to investigate further machine learning models to compare their predictive performance with that of the decision tree approach used here.

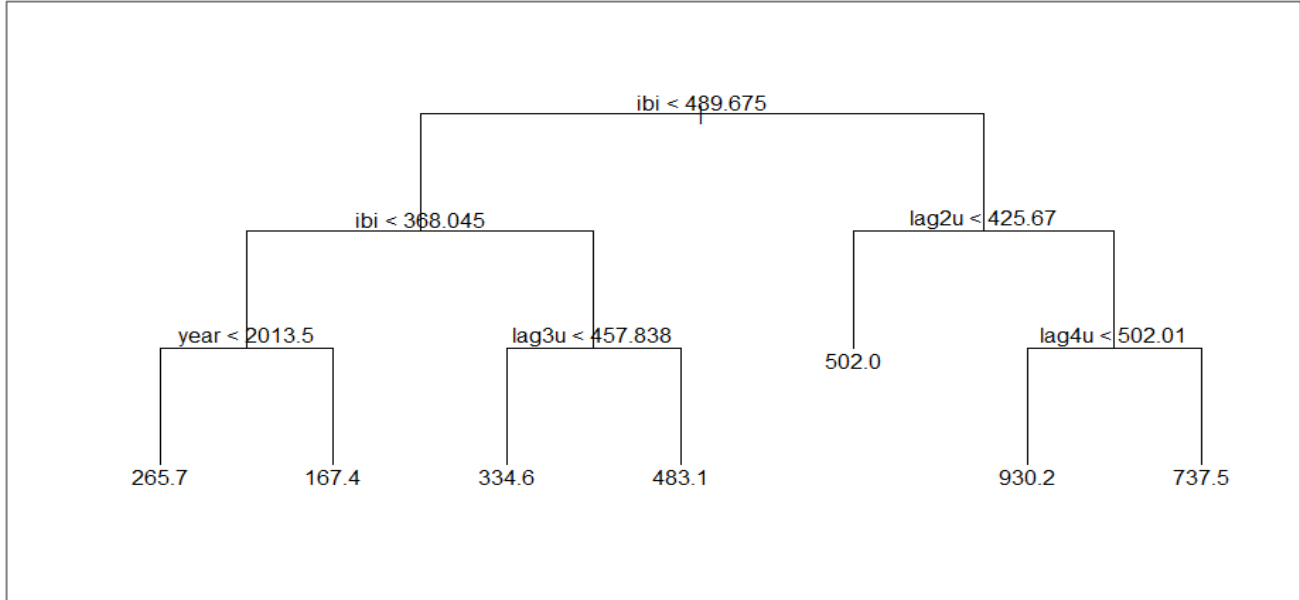


Figure 6(a): Unpruned fitted regression tree for predicting the Umaisha water level, using month, quarter, year, Ibi water level, Makurdi water level, and lag 1-4 water levels as potential predictor variables.

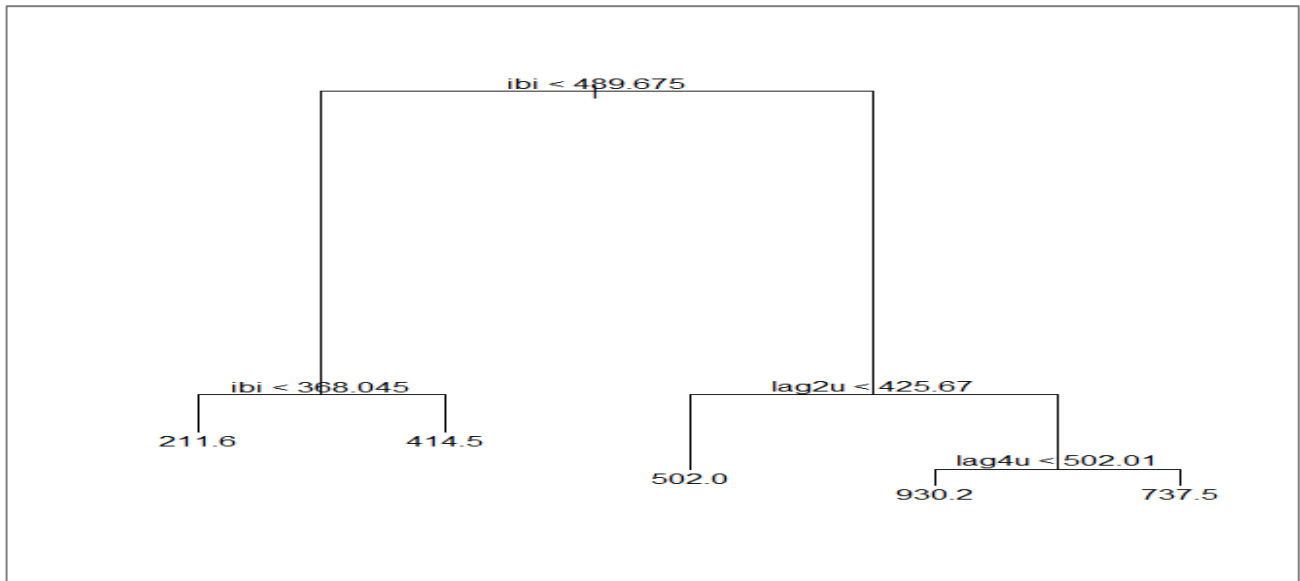


Figure 6(b): Pruned fitted regression tree for predicting the Umaisha water level, using month, quarter, year, Ibi water level, Makurdi water level, and lag 1-4 water levels as potential predictor variables.

ACKNOWLEDGEMENTS

This work was carried out while the first author was at the University of Strathclyde, UK for his PhD studies. The first author is grateful to the Petroleum Technology Development Fund (PTDF), Nigeria, for generously funding his PhD research. The authors gratefully acknowledge the Nigeria Hydrological Services Agency (NIHSA), especially Mr. Bamgbose Taiwo A. for his assistance in supplying the water level data used in this study.

REFERENCES

- Adesola, R. O., Okeke, V. C., Chris, N. P., Dike, C. K., Olughu, O. I., & Vermilye, A. (2024). Public health impacts of flooding: A case study of 2022 flood outbreak in Nigeria. *International Journal of Travel Medicine and Global Health*, 12(3), 145–153. [\[Crossref\]](#)
- Akmam, E. F., Soemartojo, S. K., Siswantining, T., & Sarwinda, D. (2019). Multiple imputation with predictive mean matching method for numerical

- missing data. In *3rd International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE. [\[Crossref\]](#)
- Allison, P. (2015). Imputation by predictive mean matching: Promise & peril. *Statistical Horizons*. statisticalhorizons.com
- Arbain, S. H., & Wibowo, A. (2012). Time series methods for water level forecasting of Dungun river in Terengganu Malaysia. *International Journal of Engineering Science and Technology*, 4(4), 1803–1811. [ijest.info](#)
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer Nature. [\[Crossref\]](#)
- Carmona, R. (2014). *Statistical analysis of financial data in R* (2nd ed.). Springer. [\[Crossref\]](#)
- Choi, C., Kim, J., Han, H., Han, D., & Kim, H. S. (2019). Development of water level prediction models using machine learning in wetlands: A case study of Upo wetland in South Korea. *Water*, 12(1), 93. [\[Crossref\]](#)
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. [\[Crossref\]](#)
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer. [\[Crossref\]](#)
- Lin, T., & Watanabe, H. (2017). Water level prediction for disaster management using machine learning models. In *16th Information Science and Technology Forum*. Waseda University. waseda.ac.jp
- Lumbroso, D. (2020). Flood risk management in Africa. *Journal of Flood Risk Management*, 13, 1–5. [\[Crossref\]](#)
- Lumbroso, D., Brown, E., & Ranger, N. (2016). Stakeholders' perceptions of the overall effectiveness of early warning systems and risk assessments for weather-related hazards in Africa, the Caribbean and South Asia. *Natural Hazards*, 84(3), 2121–2144. [\[Crossref\]](#)
- Lütkepohl, H. (1993). *Introduction to multiple time series analysis* (2nd ed.). Springer-Verlag. [\[Crossref\]](#)
- Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: Theory and applications* (1st ed.). World Scientific.
- National Bureau of Statistics. (2017). *Demographic statistics bulletin*. nigerianstat.gov.ng
- Nigeria Hydrological Services Agency. (2020). *Annual flood outlook*. nihsa.gov.ng
- Nwobi-Okoye, C. C., & Igboanugo, A. C. (2013). Predicting water levels at Kainji Dam using artificial neural networks. *Nigerian Journal of Technology*, 32(1), 129–136. ajol.info
- Oyerinde, G. T., Lawin, A. E., & Adeyeri, O. E. (2021). Multi-variate infilling of missing daily discharge data on the Niger basin. *Water Practice and Technology*, 16(3), 961–979. [\[Crossref\]](#)
- Peduzzi, P., Dao, H., Herold, C., & Mouton, F. (2009). Assessing global exposure and vulnerability towards natural hazards: The disaster risk index. *Natural Hazards and Earth System Sciences*, 9(4), 1149–1159. [\[Crossref\]](#)
- Phan, T.-T.-H., & Nguyen, X. H. (2020). Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. *Advances in Water Resources*, 142, 103656. [\[Crossref\]](#)
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. r-project.org
- Ripley, B. (2023). *tree: Classification and regression trees* [R package version 1.0-44]. [\[Crossref\]](#)
- Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications* (3rd ed.). Springer-Verlag. [\[Crossref\]](#)
- Sims, C. A. (1972). Money, income, and causality. *The American Economic Review*, 62(4), 540–552. jstor.org
- Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley. [\[Crossref\]](#)
- Umar, N., & Gray, A. (2022). Flooding in Nigeria: A review of its occurrence and impacts and approaches to modelling flood data. *International Journal of Environmental Studies*, 80(3), 540–561. [\[Crossref\]](#)
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. [\[Crossref\]](#)
- Vink, G., Frank, L. E., Pannekoek, J., & van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, 68(1), 61–90. [\[Crossref\]](#)
- White, H., & Pettenuzzo, D. (2014). Granger causality, exogeneity, cointegration, and economic policy analysis. *Journal of Econometrics*, 178(2), 316–330. [\[Crossref\]](#)
- Xu, G., Cheng, Y., Liu, F., Ping, P., & Sun, J. (2019). A water level prediction model based on ARIMA-RNN. In *5th IEEE International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE. [\[Crossref\]](#)
- Yan, Q., & Ma, C. (2016). Application of integrated ARIMA and RBF network for groundwater level forecasting. *Environmental Earth Sciences*, 75(5), 396. [\[Crossref\]](#)
- Yu, Z., Lei, G., Jiang, Z., & Liu, F. (2017). ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River. In *4th International Conference on Transport Information and Safety (ICTIS)*. IEEE. [\[Crossref\]](#)
- Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3), 7–10. r-project.org
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [\[Crossref\]](#)
- Zhang, M. (2018). Autoregressive models AR, MA, ARMA, ARIMA. *Time series*. University of Pittsburgh. pitt.edu

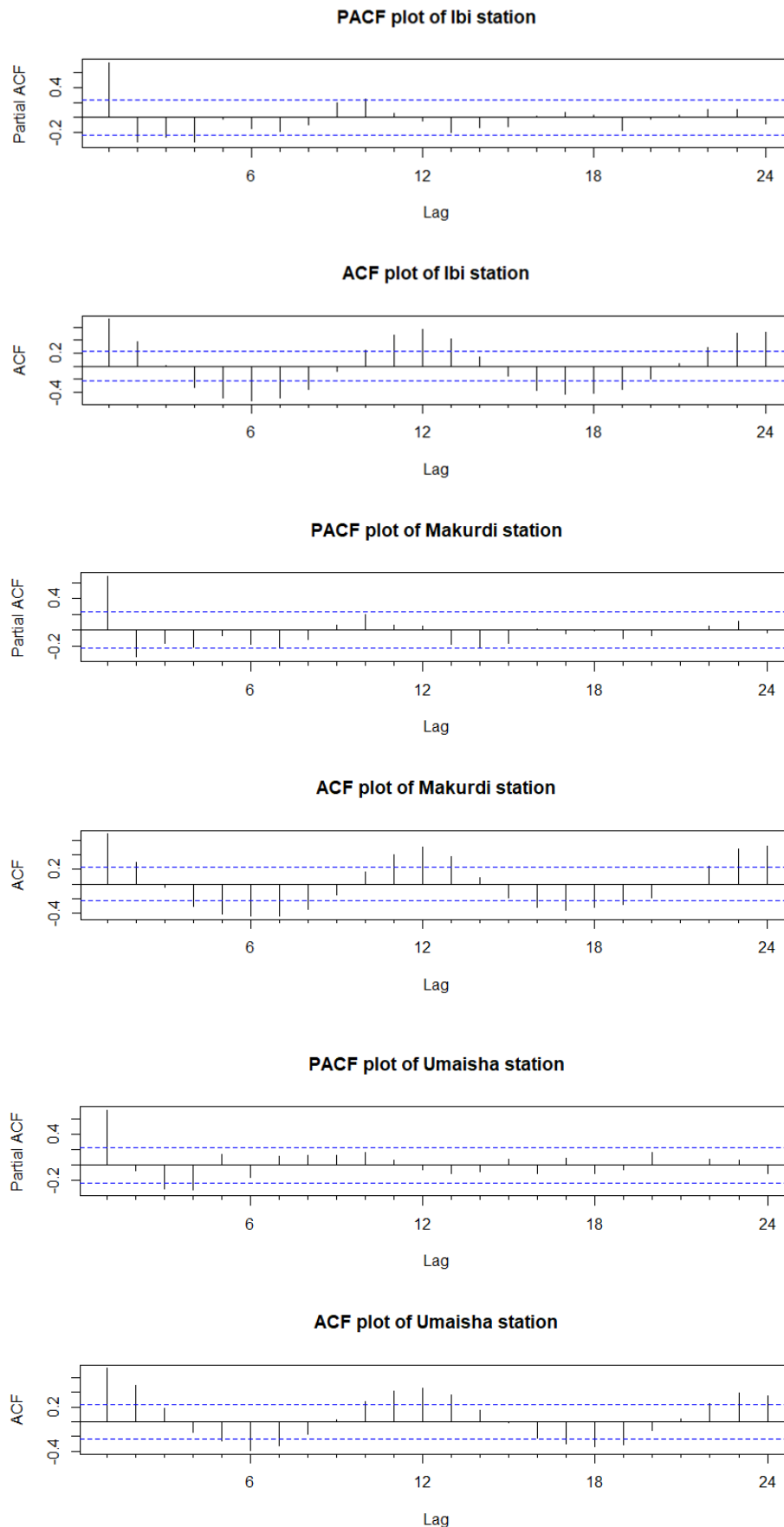


Figure A.1: PACF and ACF plots of the water level datasets from the Ibi, Makurdi and Umaisha water stations on the river Benue.