

Modeling and Forecasting of Lake Malombe Fish Biomass and Catch Per Unit Effort (CPUE)

Rodgers Makwinja^{1,2*}, Seyoum Mengistou¹, Emmanuel Kaunda³, Tena Alemiew⁴, Titus Bandulo Phiri², Ishmael Bobby Mphangwe Kosamu⁵ and Chikumbusko Chiziwa Kaonga⁵

¹African Centre of Excellence for Water Management, Addis Ababa University, P.O. BOX 1176, Addis Ababa, Ethiopia, rodgers.makwinja@aau.edu.et/makwinjarodgers@gmail.com, seyoumeng@gmail.com/ Smeng091@bio.aau.edu.et

²Senga Bay Fisheries Research Centre, P. O. Box 316, Salima, Malawi, rodgers.makwinja@aau.edu.et/makwinjarodgers@gmail.com, titobandulo@gmail.com / 2020sfr@gmail.com

³ African Centre of Excellence in Aquaculture and Fisheries (AquaFish), Lilongwe University of Agriculture and Natural Resources, ekaunda@yahoo.com/ ekaunda@luanar.ac.mw

⁴Water and Land Resource Centre of Addis Ababa University, P.O. Box 3880, Addis Ababa, Ethiopia.

⁵Physics and Biochemical Sciences Department, University of Malawi, The Polytechnic, P/Bag 303, Chichiri, Blantyre 3, Malawi, ikosamu@poly.ac.mw, ckaonga@poly.ac.mw

The author to whom correspondence should be addressed:

makwinjarodgers@gmail.com/rodgers.makwinja@aau.edu.et, phone:+265888625533, ORCID, <https://orcid.org/0000-0002-0818-4727>

Keywords: ARIMA, CPUE, Fish biomass landings, Forecasting, Lake Malombe, Time series approach

Abstract

Lake Malombe fish stocks have been depleted by chronic overfishing. Various management approaches (co-management, command control, and ecosystem-based management to fisheries) have been used to manage the fishery. However, the lack of an accurate predictive model has hampered their success. Therefore, we developed and tested a time series model for Lake Malombe fishery. The seasonal fish biomass and CPUE trends were first observed and both were non-stationary. The second-order differencing was applied to transform the non-stationary data into stationary. Autocorrelation functions (AC), partial autocorrelation function (PAC), and Akaike information criterion (AIC) were estimated, which led to the identification and construction of autoregressive integrated moving average (ARIMA) models, suitable in explaining the time series and forecasting. The results showed that ARIMA (1,2,1) provided a better prediction than its counterparts. The model satisfactorily predicted that by 2032, both fish biomass and CPUE will decrease to 3204.6 tons and 59.672 respectively, signifying the potential threat to Lake Malombe fishery. The model justified the necessity of taking precautionary measures to avoid the total collapse of the fishery.

Introduction

Lake Malombe is located within the Upper and Middle Shire River basin in the Southern part of Malawi. The Lake is an important freshwater ecosystem especially for provisioning services such as fishing, culture, regulation, and others [1-3]. Over the past years, Lake Malombe has experienced a serious collapse of fish stocks due to, among other factors, increased depletion, and degradation of water quality, siltation, lake water level fluctuation, invasion of alien species such as water hyacinth, and over-exploitation. The recent increasing pressure from the rapidly growing population of fisherfolks [4-6] around the lake has subjected some of the major fish species such as *Oreochromis spp*, *Bagrus meridionalis*, *Clarias species*, large demersal *haplochromine* cichlid, *Labeo mesops*, *Opsaridium microcephalum*, *Trematocranus placodon*, and others into extirpate. Similar trends have also been reported elsewhere. For example, in Egypt [7], the evidence of over-fishing has been demonstrated by the loss of fish biomass from Bardawi Lagoon, North Sinai Port Said, and in the mediterranean sea [8]. The same collapse has also been reported in other African Lakes such as Lake Kariba and Lake Victoria [9]. In Bangladesh, Barua [10] also reported a gradual decline of high-value fish species as

the result of overfishing. In Estonian lakes, Kangur et al. [11] on the other hand reported a significant reduction in the fish population as the result of human activities.

The concept of stock production models has been applied in Malawi inland fishery to provide information for fisheries managers and biologists in achieving management goals and take appropriate management strategies for their sustainable exploitation. The concept is based on the assumption that the dynamic of lake fishery is directly linked to the concept of the precautionary approach- a basic concept in fish stock management [12]. Schaeffer and Fox production approaches -also known as surplus production models (a model associated with maximum sustainable yield (MSY)) have been commonly applied to depict the status of fishery in Lake Malawi and other inland lakes. Although these models are acceptable in fishery management, their applications are still questionable. For example, the models only depict the current picture of the fishery using the time series trend and lack the power of prediction. Other models such as bootstrap, state-space [13], and Bayesian [12,13,14,15] have also been applied with little success. The time-series approach, however, has been indispensable in understanding natural resources systems and the development of better management policies[16]. It is described as a new approach to fisheries modeling and stock assessment. The approach demands few biological assumptions than other traditional fisheries models [17] and with simple mathematical techniques and few assumptions, it can significantly reduce the modeling costs including research costs [16]. Time series modeling provides a feasible way to examine the time series data and provide a prediction [18-19]. Several researchers have recommended the time series approach as the most outstanding technique. Lake Malombe has a long run time series of detailed data records on fish biomass landings and CPUE. However, no research work has been undertaken to forecast the future status of the fishery. In other words, management approaches (co-management, command control, and ecosystem-based management to fisheries) have failed due to a lack of accurate predictive models in this lake. In this paper, we developed and tested a time series model as a tool for forecasting the status of Lake Malombe fish biomass landings and CPUE.

MATERIALS AND METHODS

The study area

Lake Malombe is located in the Southern part of Lake Malawi within coordinates 14°40'0"S 35°15'0"E. It is an impoundment of the outflow from Lake Malawi via the Upper Shire River [20] (Figure 1).

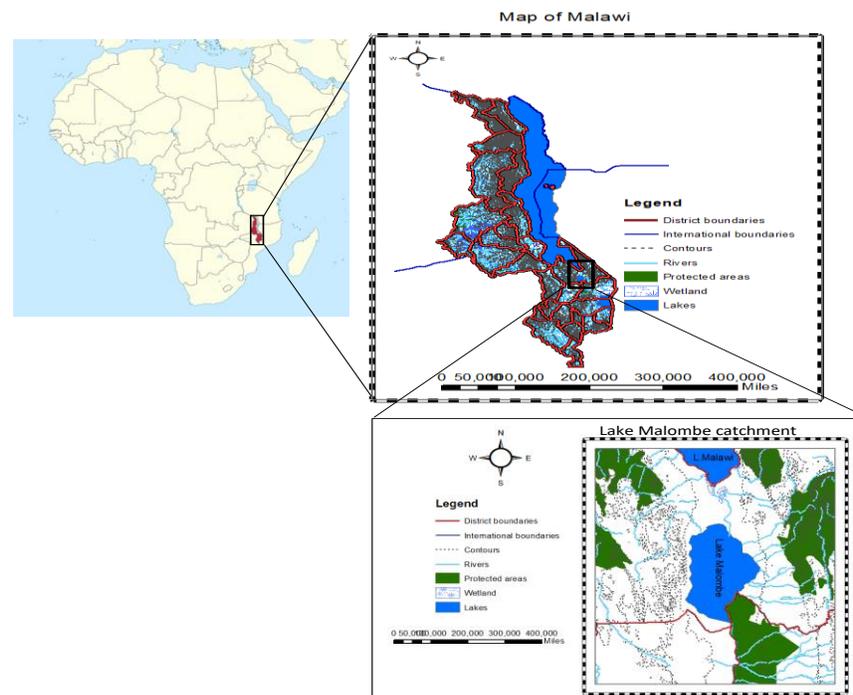


Figure 1: Map of Lake Malombe

The Lake is documented as the third-largest in Malawi with an estimated total area of 162 square miles (420km²), water surface area of approximately 390 km², length of 30km, the width of 17km, and water depth not exceeding 6m [21]. The communities around the lake are predominately fishers [22] and the lake has approximately 65 fishing beaches scattered over the three major administrative strata known as Lake Malombe East coded as 1.1, Lake Malombe West coded 1.2 and Upper Shire coded 1.3 [23]. The surrounding area of Lake Malombe is densely populated by the Yao ethnic tribe consisting of over 85% [22] of the fishing population. Few tribes such as Chewa, Lhomwe, and Nyanja are also found around the Lake.

Data collection

Department of Fisheries through its statistical office under the Malawi Fisheries Management Act (1997) regulations are mandated to collect monthly catch, fishing effort, and CPUE data from both large-scale and artisanal fishermen. The data collected using Malawi Traditional Fishery (MTF)-a computerized gear-based sampling technique are entered into catch and effort statistics database hosted by Monkey Bay Fisheries Research Division of the Department of Fisheries. In this study, 41 years of fish biomass and CPUE data were used to develop a forecast model for Lake Malombe fish biomass and CPUE.

Conceptual framework of the ARIMA model

To model the fish biomass and CPUE data, the traditional statistical models such as Autoregressive (AR), smoothing, Moving Average (MA) and ARIMA are applied. The autoregressive (AR (p)) model is expressed as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \quad (1)$$

where Y_t is the dependent variable y at time t , $Y_{t-1} + Y_{t-2} + \dots + Y_{t-p}$ are the lagged dependent variables while $\alpha_1 + \alpha_2 + \dots + \alpha_p$ are the unknown parameters of the model $\alpha_1 \neq 0$ and ε_t is the value of the disorders term at time t , i.i.d. $\varepsilon_t \sim (0, \alpha^2)$: p -the number of lagged values of y and represent the order of the process. The moving average (MA(q)) is defined by a function of its present and q -past disorders (lagged error) and is expressed as

$$Y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (2)$$

where Y_t is the dependent variable y at time t , $\varepsilon_{t-1}, \varepsilon_{t-2} \dots \varepsilon_{t-q}$ means lagged disturbances and $\beta_1, \beta_2 \dots \beta_q$ means unknown parameters of the model $\beta_q \neq 0$, $\varepsilon_t \sim N(0, \alpha^2)$ q the number of lagged values of y and represents the order of the process. The two-process (AR) and (MA) are combined to form an autoregressive moving average model (ARMA (p, q)) expressed as:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_2 \varepsilon_{t-2} \dots + \beta_q \varepsilon_{t-q} \quad (3)$$

However, the ARMA model works with stationary data which is not the case with fish biomass and CPUE data. Therefore, the non-stationary data can be modeled using a log transformation to linearize the series. The application of differencing to remove a mean trend from non-stationary series is also another procedure for transforming non-stationary data into stationary. The procedure has been advocated in the Box-Jenkins approach. In this paper, the autoregressive integrated moving average model (ARIMA (p, d, q)) is introduced to deal with non-stationary data. The general form of the ARIMA model is

$\alpha(L)\Delta^d \gamma_t = \beta(L)\varepsilon_t$ with ε_t and $\Delta^0 = 1$ and $\alpha(L)\Delta^d \gamma_t$ is a new time series obtained by differencing initial series (γ_t) d

The autocorrelation (AC) and partial autocorrelation (PAC) functions are used to identify and construct ARIMA models suitable for explaining the time series and forecasting future values. The ACF is defined as the sequence of values:

$$\tau_t = \frac{c_t}{c_0} \quad T=0, 1, \dots, T-1 \quad (4)$$

where

$$c_T = \frac{1}{T} \sum_{t=\tau}^{T-1} (y_t - \bar{y})(y_{t-\tau} - \bar{y}) \quad (5)$$

is the empirical autocovariance at lag τ and C_o is the sample variance. The ACF reveals regular moving average spikes. For example, if the model has an MA (1) component, then there will be only one regular significant spike. If the model has an MA (2) component, then the model will have two regular spikes. The sample ACF ρ_τ at lag, τ is simply the correlation between two sets of residuals obtained from regressing the elements γ_t and $\gamma_{t+\tau}$ on the set of intervening values. The PACF reveals a spike at the lag of the interaction term. Different combinations of multiplicative parameters were estimated to determine whether the identified parameters were statistically significant. The t ratios of the parameters to their standard error were estimated. To evaluate the model adequacy, likelihood ratio and AIC were used. Ljung-Box Chi-squared test was also used to estimate whether the overall correlogram of the residuals displayed any methodical error. The Ljung-Box Chi-squared test is expressed as

$$Q = n(n+2) \sum (n-k)^{-1} r k^2 \quad (6)$$

where r_k ($k=1, \dots, m$) are residual AC and n is the number of observations used to fit the model. Akaike Information Criterion (AIC) is expressed as

$$AIC = -2 \ln f\left(\frac{y}{\phi}\right) + 2k \quad (7)$$

Where $-2 \ln f\left(\frac{y}{\phi}\right)$ means the goodness of fit while $+2k$ means model complexity and k are several model parameters. Note, the model with the smallest AIC is chosen as the finest. Once the suitable time series model estimated its unknown parameters and established that the model fits well, the next step is to forecast future fish biomass and CPUE values. In this context, the autoregressive model is represented as follows:

$$Y_t = \sum_{u=1}^{\infty} \pi_u \gamma_{t-u} + \varepsilon_t \quad (8)$$

The next observation beyond y_1, \dots, y_T is predicted using the model expressed below

$$\hat{Y}_{t+1} = \sum_{u=1}^{\infty} \hat{\pi}_u \gamma_{t+1-u} \quad (9)$$

Where the $\hat{\pi}_u$ are obtained by substituting the estimated parameters in the theoretical ones. The forecast γ_{t+1} is obtained and used to forecast γ_{t+2} which later is used to generate γ_{t+3} . The process is used to obtain a forecast out of any point in the future. The R statistical software version 3.6.3 has all required functions to fit the 41 years' time series of fish biomass and CPUE data using ARIMA models.

RESULTS

Modeling and forecasting of total annual biomass landings and CPUE were conducted based on the raw data. The precision and characteristics of ARIMA models were detailed studied following the Box-Jenkins approach. The model parameters were identified, estimated, and verified. The model shown in Figure 2 is based on the 41 years' time-series data from 1976 to 2017. Figure 2a depicts the instability of Lake Malombe's total annual biomass landings trend with the highest observed from 1980 to 1990. The same observation is made in the CPUE data plot Figure 2b which shows the highest within the period from 1980 to 1994.

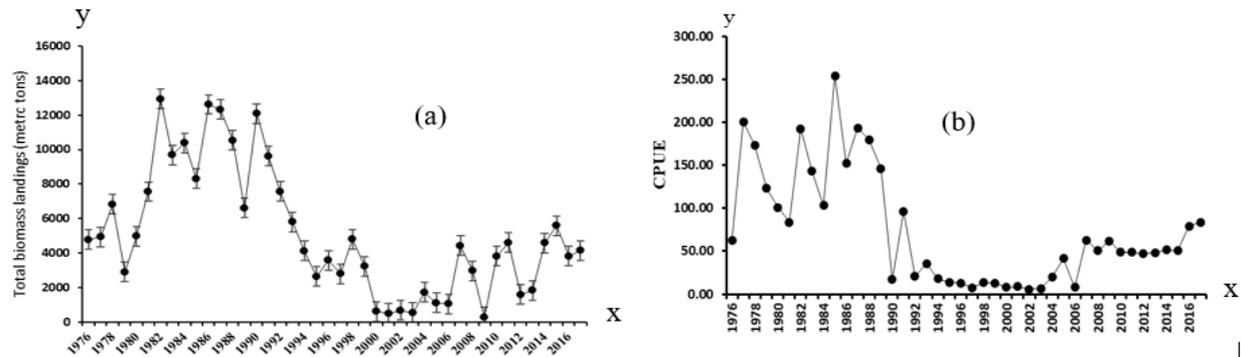


Figure 2: Total biomass landings (tons) (a) and catch per unit effort (b)

The first approach to time series modeling is data inspection. Time series data depicting seasonal spikes are usually described as non-stationary. As observed in Figure 2, the biomass landings and CPUE trends have been unstable from 1976 to 2017 indicating non-stationarity.

Model identification

The ARIMA model generally does not work with non-stationary data. Therefore, the first-order differencing was applied to transform non-stationary data into stationary.

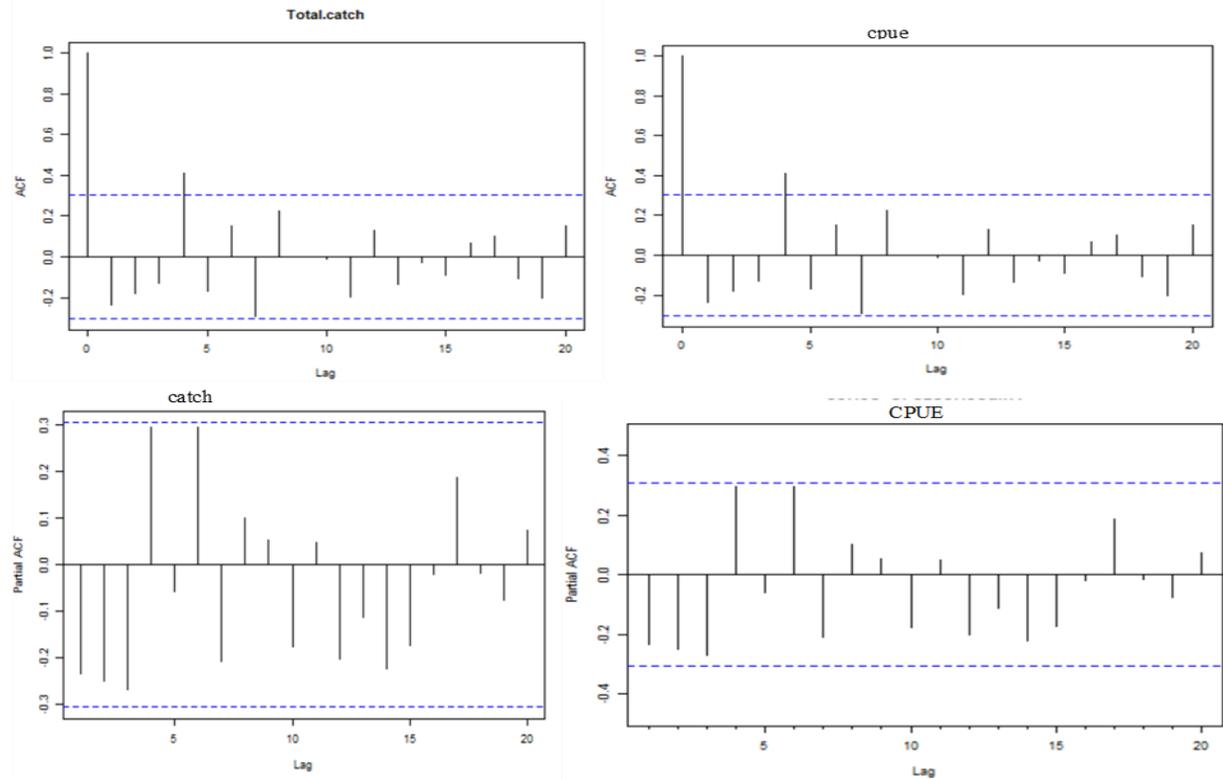


Figure 3: Autocorrelation (AC) and Partial Autocorrelation (PAC) functions used in the ARIMA model

However, it was noted that first-order differencing did not completely transform the data into stationary. Hence, second-order differencing was applied. An appropriate model was identified by examining ACF and PACF. Figure 3 shows the AC and PAC functions for both fish biomass landings and CPUE. The ACF plot (Figures 3) showed a significant spike only at lag 1 meaning that higher-order autocorrelation was explained by lag 1 AR [24]. In verifying the data stationarity, however, it was noted that AR or MA models were not pure as seen from ACF and PACF correlograms Figure 3. Therefore, several models had to be tested to identify the most suitable one for fish biomass and CPUE forecasting.

Model selection

The model selection was based on the minimization of the Akaike information criterion (AIC) and the Gaussian Maximum Likelihood Estimation algorithm. The ARIMA model in R statistical manual allows estimating the coefficient of the models that were previously identified by providing the parameters p , q , and d using a maximum likelihood estimation algorithm and Akaike information criterion (AIC). The procedure offers a new time series representing values adjusted by model residuals and confidence intervals of the adjustment at 0.05 level of significance. The best model displays the least AIC and high maximum likelihood estimation algorithm. Table 1 summarized the values of different competing model and proved the choice of the model on which fish biomass landings and CPUE prediction were based on. It is clear from Table 1 that the ARIMA model (1,2,1) is selected because it has the lowest AIC and highest maximum likelihood.

Table 1: Competing ARIMA models

Catch				CPU		
Model (P, D, Q)	AIC	-loglikelihood	Ljung-Box(P-value)	AIC	-loglikelihood	Ljung-Box(P-value)
ARIMA (1,0,1)	773.79	-382.89	0.84 ^{ns}	445.76	-218.11	0.73 ^{ns}
ARIMA (1,1,1)	755.18	-374.59	0.74 ^{ns}	434.12	-214.06	0.65 ^{ns}
ARIMA (1,2,1)	744.01	-362.02	0.54 ^{ns}	430.83	-210.42	0.91 ^{ns}
ARIMA (1,2,2)	744.07	368.03	0.18 ^{ns}	431.53	-211.76	0.49 ^{ns}
ARIMA (2,2,1)	744.31	368.06	0.68 ^{ns}	430.84	-211.42	0.12 ^{ns}

Note: ^{ns} indicates not significant while ** and * indicate significance at 0.01 and 0.05 probability level of confidence

Table 1 further shows that the p-value for the Ljung-Box test was greater than 0.05 in all competing models suggesting that there was very little evidence for non-zero autocorrelations. Both AR and MA models were greater than 0.05 suggesting that the residues of the model were independent at a 95% level of confidence and the ARMA model proved to be the best model fit.

Table 2 Estimated values of ARIMA (1,2,1) model

ARIMA Model Parameters							
			Estimate	SE	t	Sig.	
Fish biomass	Natural Log	Constant	4987.742	0.005	-5.693	0.000**	
		AR	Lag 1	-0.209	0.256	6.018	0.018*
			Lag 2	0.656	0.023	0.461	0.055
		Difference	2				
		MA	Lag 1	-1.000	0.031	5.005	0.003*
			Constant	79.660	0.023	7.432	0.000**
CPU	Natural Log	AR	Lag 1	-0.498	0.150	5.43	0.032*
			Lag 2	-0.369	0.025	3.62	0.041*
		Difference	2				
		MA	Lag 1	-1.000	0.010	4.342	0.023*

Note: ^{ns} indicates not significant while ** and * indicate significance at 0.01 and 0.05 probability level of confidence

Table 2 presents the results of the estimated values of the selected ARIMA (1,2,1) model. Based on the selected ARIMA (1,2,1) model presented in Table 2, the following model was developed.

$$y_t = \delta + \delta_1 y_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

with: y_t, y_{t-1} : fish biomass landings or CPUE period t and t-1, respectively, $\varepsilon_t, \varepsilon_{t-1}$: residuals of period t and t-1 constitute a white noise and δ_1 and θ_1 : coefficients of AR and MA processes respectively. From Table 2, the coefficient of AR and MA were extracted to develop the following forecasting models for fish biomass and CPUE.

$$y_{fish\ biomass}^t = 4987.742 - 0.209 - 1.000\varepsilon_{t-1}$$

$$y_{cpue}^t = 79.660 - 0.498 - 1.000\varepsilon_{t-1}$$

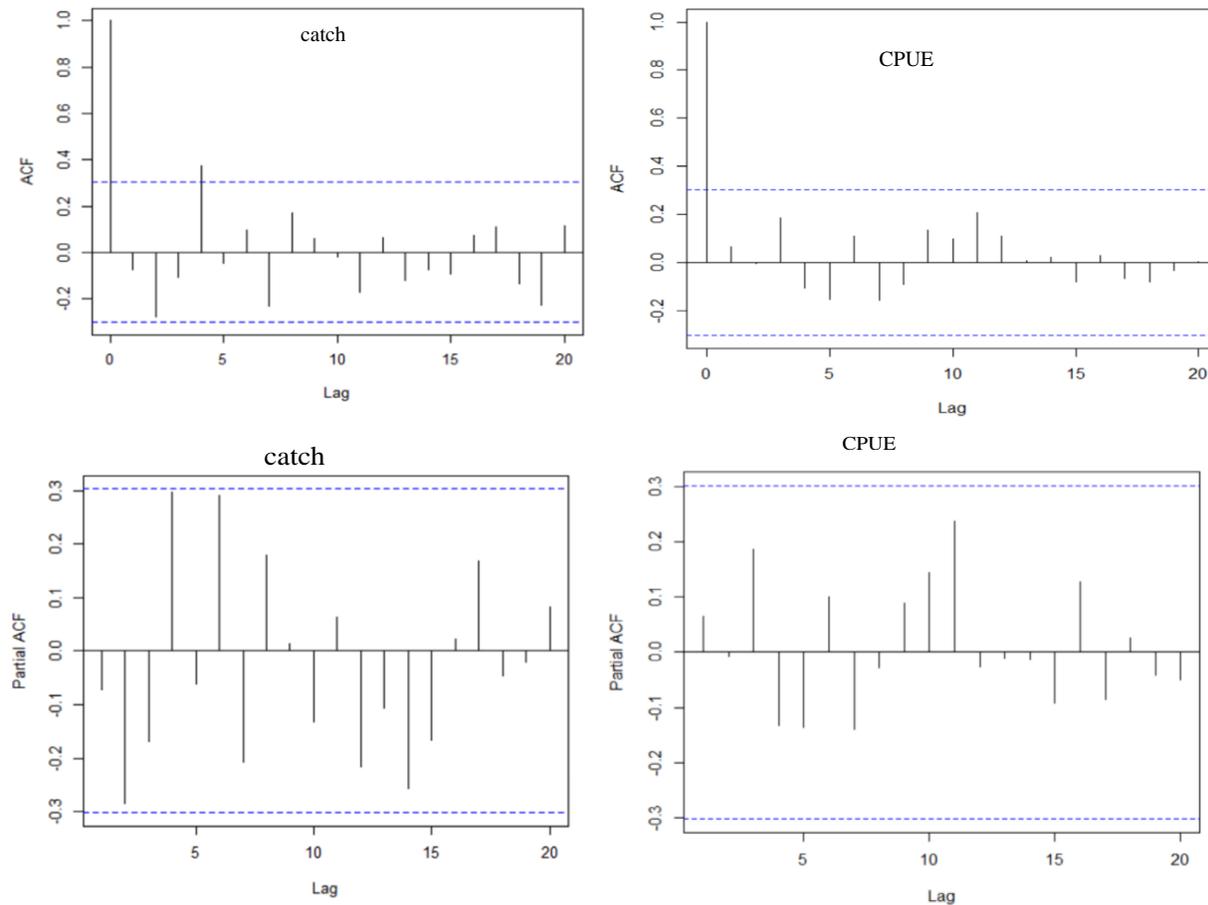
Accuracy of ARIMA (1,2,1) model

Before forecasting, the residuals were checked through the Box-Ljung test, ACF, and PACF to see if there was any systematic pattern that needs to be eliminated to improve the accuracy and performance of the selected model. The Ljung-Box test (Table 3) for CPUE had a p-value of 0.911 while fish biomass was 0.103 and both were not significant suggesting that there was little evidence of non-zero AR in the sample forecast errors at lags 1-20. Therefore, ARIMA (1,2,1) provided an adequate predictive model which probably could not be improved.

Table 3: The results of the Ljung-Box test

	X-squared	df	P.value
CPUE	12.13	20	0.911
Fish biomass	28.29	20	0.103

Besides, the ACF and PACF residuals plot (Figure 4) showed that none of the AC was significantly different from zero at 95% confidence intervals.

**Figure 4:** Autocorrelation and partial autocorrelation functions residuals

This proved that the selected ARIMA (1,2,1) model was an appropriate model for forecasting Lake Malombe fish biomass and CPUE values and could not be improved.

Forecast

After defining the most appropriate model of fish biomass and CPUE values, the forecasting was made. Table 4 shows that the noise residuals were a combination of positive and negative errors and falling within 95% confidence intervals indicating that the model had a good performance of forecasting.

Table 4: Forecast biomass and CPUE from 2014 to 2032

Year	Fish biomass (metric ton)				CPU			
	Actual	Predicted	UB	LB	Actual	Predicted	UB	LB
2014	4796.02	5620.03	6780.16	-462.12	92.65	72.43	112.13	-12.64

2015	5587.78	5097.64	1640.49	-234.13	86.67	76.66	140.3	-18.34
2016	3819.99	4819.2	4216.8	-170.4	78.91	78.97	110.14	-20.17
2017	4151.56	4104.24	8170.4	-442.17	83.44	73.44	160.13	-19.16
2018	-	4060.97	8699.16	-577.22	-	71.73	160.41	-16.94
2019	-	4058.81	10040.49	-1922.86	-	75.02	171.46	-19.42
2020	-	4038.14	11270.45	-3104.15	-	74.169	181.57	-33.23
2021	-	4021.35	12339.49	-4297.25	-	72.165	193.09	-48.76
2022	-	4003.75	8699.16	-577.22	-	71.748	202.68	-59.19
2023	-	3986.31	10040.49	-1922.86	-	70.371	212.01	-71.26
2024	-	3951.38	11270.45	-3104.15	-	69.214	221.11	-82.68
2025	-	3916.46	15118.86	-8052.09	-	66.977	220.8	-93.49
2026	-	3898.97	16762.32	-8894.47	-	65.849	236.27	-104.31
2027	-	3881.54	17545.79	-9712.866	-	64.719	246.64	-114.85
2028	-	3664.01	18309.25	-10511.26	-	63.571	254.64	-125.2
2029	-	3584.04	19055.82	-11292.75	-	62.446	254.64	-125.2
2030	-	3446.04	19787.91	-120559.77	-	61.311	270.4	-145.5
2031	-	3329.14	20507.49	-12814.27	-	60.175	278.09	-155.47
2032	-	3204.6	21216.132	-13557.84	-	59.672	285.69	-165.34

Note: CV means confidence intervals, LB means lower bound, means upper bound

Figure 5 presents the results of fish biomass and CPUE forecasts obtained after applying the ARIMA model (1,2,1) for the period of 15 years from 2017 to 2032.

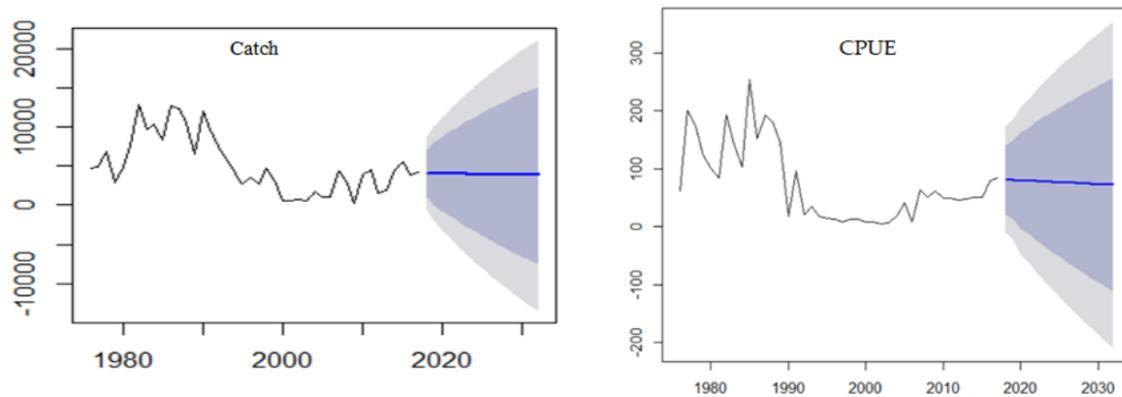


Figure 5: Fish biomass and CPUE Forecasted values

As seen from Table 4 and Figure 5, the model satisfactorily predicted that by 2032, both fish biomass and CPUE will decrease to 3204.6 tons and 59.672 respectively.

Discussion

The abundance of fish species in Lake Malombe is directly linked to CPUE though common criticism of CPUE is that the relationship between abundance and CPUE is more complex [25-26]. Both the catch and CPUE appeared to increase from 1980 to 1984 and remained constant with a slight fluctuation from 1984 to 1990s and then collapsed drastically from 1994 to 2000. Several researchers had also similar observations. For example, Weyl et al [27], evidenced an increase in biomass and CPUE within the period of 1992/1993 in the Southern part of Lake Malawi after subsequent closure of the fishery. Alexander et al [28] on the other hand reported a strong relationship between whole lake CPUE to relative fish biomass and abundance. Maynou et al [29] further noted that CPUE series reflect the general abundance of species and catch fluctuation. It was noted that the decrease in CPUE and fish biomass trends within the period of

1990s to 2005 was an indicator of an increase in the diversity of gears, population of fisherfolks, and overexploitation. According to Weyl et al [27], CPUE variation is strongly linked to the difference in the number of fishers, man-hours, and the categories of fishing gears. Low CPUE indicates a relatively low abundance of fish which results in prolonged man-hours and an increase in the number of fishers and low catches [30]. The option to model both catches and CPUE was based on the fact that these two indicators can be used by regulators to monitor for a potential change in fish population related to the effects of human exploitation and other anthropogenic factors [31].

In this study, the first approach to time series modeling was data inspection using a plot. If time-series data depicted seasonal spikes, it means that the data is non-stationary and needs differencing. The fish biomass and CPUE series appeared to be stationary after second-order differencing. The method of modeling time series of fish biomass and CPUE addressed the problem of non-stationarity and gave good predictions. In other words, the conditional variance of the model was found to be volatile with time series and the ARIMA model well addressed this phenomenon. The study findings showed that it is reasonable to use history as a basis for prediction because of the steady decreasing trend in the time series and the slowly diminishing noise. The time series for the fish biomass and CPUE reflected a common scenario in Lake Malombe. Evidence has shown that various fish species in the Lake have been targeted by fisherfolks in recent years and its stocks are overexploited. A similar situation occurs in Lake Chilwa, Chiuta, and Shire river [32-33]. We noted that several other factors such as fishing effort, gear modification, climate variability, increased in fisherfolks population, and change in fishing gears and technology had a strong influence on fish biomass and CPUE trend though further research is required to support this claim. The ARIMA (1, 2,1) model predicted that both fish biomass and CPUE will continue declining suggesting that the fish stocks are continuously being overexploited. The model predictions agree with Kanyere et al [34] who also observed that the relative biomass index of Lake Malombe fishery in Figure 6a has been fluctuating with the highest registered in the mid-1970s and lowest (less than 0.3 units) in 2017. Using Schaefer Biomass dynamic model, Kanyere et al further demonstrated that the fish biomass (Figure 6b) gradually increased from about 30,000 tons in the early 1970s to a maximum of 44,000 tons by 1981 and rapidly declined to half the carrying capacity ($K/2$) by 1986. After that, the biomass continued to decline with an increase in fishing effort reaching the lowest level of just above 3,000 by the year 2000 thereafter slight increase in the biomass became noticeable and by 2018, the biomass had reached about 6,000 tons.

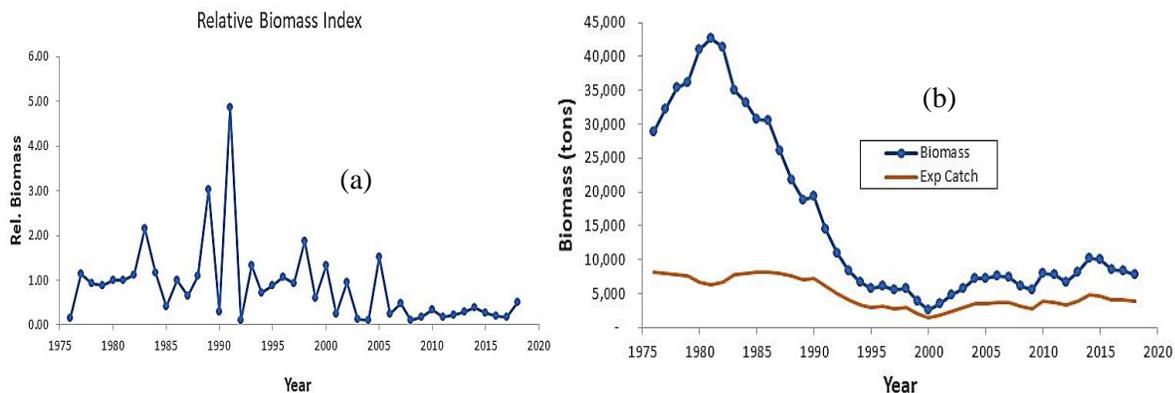


Figure 6: Relative biomass index (a), biomass and exponential catch (b) [34]

Kanyere et al further attempted to calculate the current fishing mortality for Lake Malombe and Upper Shire river using the average for the last three years and reported to be at 0.49 ± 0.01 . Comparing this value to the fishing mortality sustainable yield (FMSY) (0.28 ± 0.04), they concluded that the current mortality in Lake Malombe is over 75% above the optimum long-term fishing mortality suggesting that the Lake Malombe fishery has experienced a catastrophic decline. In Lake Nasser, Egypt, the time series of size and CPUE also showed a negative trend indicating the high exploitation rate of the most important commercial fish species in the lake by the fishing gears [35]. Therefore, this study finding present a wakeup call to the fisheries managers and shows that, if the over-exploitation of Lake Malombe fishery is to be reduced without alternative livelihood option, the most vulnerable riparian population living at the margins of the cash economy will have limited choices for generating income and sustaining their livelihoods [5,2,6].

However, if the population of indigenous fish species in Lake Malombe collapse, immigration from elsewhere will do little to help the population recover [36]. This can also be critical if a restocking program is attempted using the population from other localities. Given the failure of previous efforts to manage Lake Malombe fishery, this research suggests that there is a need to review Malawi fisheries regulations to balance between conservation and the demand of the population of fisherfolks otherwise the CPUE and fish biomass models developed in this study indicate the potential threat to Lake Malombe fishery.

Conclusion

In this context, an ARIMA model was developed to model the fish biomass and CPUE forecasting using the Box-Jenkins time series approach. Out of five competing models (ARIMA (1,0,1), ARIMA (1,1,1), ARIMA (1,2,1), ARIMA (1,2,2) and ARIMA (2,2,1)), the ARIMA model (1,2,1) was selected and the results proved that this model can be used for modeling and forecasting the future status of Lake Malombe fish biomass and CPUE trends accurately.

Declaring a conflict of interest

The authors of this paper declare that there is no conflict of interest.

Supplementary Materials: The time series data for Lake Malombe biomass and CPUE are available online at www.mdpi.com/xxx/s1

Author Contributions: RM. Currently, a Ph.D. candidate at Addis Ababa University conceptualize the study, developed the methodology, sourced the data, analyzed the data, and developed the original manuscript. Authors SM, EK, TA supervised the study, reviewed and edited the manuscript, visualized, and validated the study. Authors, TB, IBMK, and CCK reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Source funds

This research was funded by the Federal Democratic Republic of Ethiopia through the African Center of Excellence for Water Management, Addis Ababa University, Grant number GSR/9316/11.

Acknowledgment

The authors wish to thank the Department of Fisheries particularly the Mangochi District Fisheries Office and Monkey Bay Fisheries Research for providing fish biomass landings and CPUE data. M'balaka Mwamad is further appreciated for his technical advice particularly in statistical analysis.

References

1. Fa, J, and D Brown. "Impacts of hunting on mammals in African tropical moist forests: A review and synthesis ." *Mammal Review* (2009): 39 (2009) 231–264
2. Le Billon, P. "The political ecology of war: natural resources and armed conflicts. ." *Political Geography* (2001): 20 (2001) 561- 584.
3. Makwinja, R, I.B.M Kosamu, and C.C Kaonga. "Determinants and Values of Willingness to Pay for Water Quality Improvement: Insights from Chia Lagoon, Malawi." *Sustainability* (2019): 11, 4690; DOI:10.3390/su11174690.
4. Nasi, R, T Christophersen, and C Belair. "Ending empty forests: Management and sustainable use of wildlife in tropical production forests." *ITTO Tropical Forest Update* (2010): 20(2010) 19–21
5. Nagoli, J, et al. " Conflicts over Natural Resource Scarcity in the Aquatic Ecosystem of the Lake Chilwa." *Environment and Ecology Research* (2016): 4(4)207-216.
6. Wainger, L, and M Mazzotta. "Applying an ecosystem service framework to wetlands management and policy: what is needed from interdisciplinary research?" *Environmental Management* (2011): 48(4):71-73
7. El-Aiatt, A.A.O, K.A.S Shalof, and M.M Saber. "Bio-economic studies on the catch of Bardawi Lagoon, North Sinai, Egypt." *Egyptian Journal of Aquatic Research* (2019): 45(2019)59-65.

8. Darasi, F, and M Aksissou. "Longline, trawl and purse seine in coastal fishing of Tangier port in North-West of Morocco." *Egyptian Journal of Aquatic Research* (2019): 45(2019)381-388.
9. Njaya, F. "Governance challenges for the implementation of co-management: experiences from Malawi." *International Journal of the Commons* (2007): 1(2007)137-15
10. Barua, S. "Maximum Sustainable Yield (MSY) estimates for industrial finfish fishery in marine waters of Bangladesh using trawl catch log." *Bangladesh J.Fish* (2009): 32(2)313-324
11. Kangur, A, P Kangur, and E Pihu. "Long-term trends in the fish communities of Lakes Peipsi and Vortsjarv (Estonia)." *Aqua Ecosyst Health Manage* (2002): 5(2002)379-389
12. Lewy, P, and Nielsen. "Modelling stochastic fish stock dynamics using Markov Chain Monte Carlo." *ICES Journal of Marine Science* (2003): 60(2003) 743-752
13. Gudmundsson, G. "Time series analysis of catch-at-age observations. ." *Applied Statistics* (1994): 43(1994)1117-1126
14. Virtala, M, S Kuikka, and E Arjas. "Stochastic virtual population analysis." *ICES Journal of Marine Science* (1998): 55 (1998): 892-904
15. Shabri, A, and R Samsudin. "Fishery Landing Forecasting Using Wavelet-Based Autoregressive Integrated Moving Average Models." *Mathematical Problems in Engineering* (2015): 2015(1)1-9
16. Craine, M. *Modelling Western Australian Fisheries with techniques of time series analysis: Examining data from a different perspective*. North Beach, Western Australia: Department of Fisheries Research Division, Western Australian Marine Research Laboratories., 2005
17. Scandol, J. "Use of cumulative sum (CUSUM) control charts of landed catch in the management of fisheries." *Fisheries Research* (2003): 64 (2003):19-36
18. Makwinja, R *et al*. Stochastic modeling of Lake Malawi Engraulicypris sardella (Gunther, 1868) catch fluctuation. *International Journal of Fisheries and Aquaculture*, 2018, 10(4),34-43
19. NRC, *Improving fish stock assessments*. Washington, D.C: National Academy Press, 1998
20. Fisheries Department. *Status of the fisheries of Lake Malombe and Upper Shire River*. Lilongwe: Department of Fisheries, 2019.
21. Dulanya, Z, *et al*. "Palaeolimnological reconstruction of recent environmental change in Lake Malombe (S.Malawi) using multiple proxies." *Water SA* (2014): 40(4)1-12
22. Matiya, G, and Y Wakabayashi. "Small Scale Fisheries of Malawi: An outline of Lake Malombe Fisheries." *Mem.Fac.Agr. Ehime Univ* (2005): 50(2005)3-9
23. Department of Fisheries. *Annual Frame Survey report*. Monkey Bay, Mangochi: Fisheries Research Unit, 2010
24. Mgaya, J.F. "Application of ARIMA models in forecasting livestock product consumption in Tanzania." *Cogent Food & Agriculture* (2019): 5(1) 1607430
25. Harley, S.J, A.M Ransom and A Dunn. "Is catch-per-unit-effort proportional to abundance? " *Canadian Journal of Fisheries and Aquatic Sciences* (2001): 58 (2001)1760-1772
26. Guzzo, M.M, M.D Rennie, and P.J Blanchfield. "Evaluating the relationship between mean catch per unit effort and abundance for littoral cyprinids in small boreal shield lakes." *Fisheries Research* (2014): 150 (2014) 100- 10
27. Weyl, O., Mwakiyongo, K., & D.Mandere. An assessment of the nkacha net fishery of Lake Malombe Malawi. *African Journal of Aquatic Science*, . (2004) 29(1), 47-55
28. Alexander, T.J, *et al*. "Estimating whole-lake fish catch per unit effort." *Fisheries Research* (2015): 172 (2015) 287-302
29. Maynou, F, M Demestre and P Sánchez. "Analysis of catch per unit effort by multivariate analysis and generalized linear models for deep-water crustacean fisheries off Barcelona (NW Mediterranean)." *Fisheries Research* (2003): 65 (2003) 257-269
30. Solomo, S.G, *et al*. "Catch per unit effort and some water quality parameters of Lake Kalgwai Jigawa state, Nigeria." *Food Science and Nutrition* (2018): 6(2018)450-456
31. Kidd, K.A, *et al*. "Collapse of a fish population after exposure to synthetic estrogen." *Proceedings of the National Academy of Sciences of the United States of America* (2007): 104(2007)8897-8901
32. Njaya, F, *et al*. "The natural history and fisheries ecology of Lake Chilwa, southern Malawi." *Journal of Great Lakes Research* (2011): 37 (2011) 15-25

33. Makwinja, R, et al. "Location and Roles of Deep Pools in Likangala River during 2012 Recession Period of Lake Chilwa Basin." *International Journal of Ecology* (2014): vol. 2014, Article ID 294683, 4 pages, 2014. <https://doi.org/10.1155/2014/294683>
34. Kanyere, G.Z, et al. *Status of the Fisheries of Lake Malombe and Upper Shire River, FAO/Technical Report*. Lilongwe: Fisheries Department, 2019
35. El-Far, A, et al. "Fisheries management based on gear selectivity of a tropical reservoir, Lake Nasser. Egypt." *Egyptian Journal of Aquatic Research* (2020): 46(2020)71-77
36. Turner, G.F. "Fisheries Management, Biodiversity Conservation, and Genetic Stock Structure." *Lake Malawi Fisheries Management Symposium - Proceedings*. Lilongwe: Department of Fisheries, 2001. 31-41.