Modelling And Forecasting Small Haplochromine Species (Kambuzi) Production In Malaŵi – A Stochastic Model Approach

Wales Singini, Emmanuel Kaunda, Victor Kasulo, Wilson Jere

Abstract: The study aimed at forecasting small *Haplochromine* species locally known as Kambuzi yield in Malaŵi, based on data on Lake Malombe fish catches during the years from 1976 to 2011. The study considered Autoregressive (AR), Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) processes to select the appropriate stochastic model for forecasting small *Haplochromine* species yield in Lake Malombe. Based on ARIMA (p, d, q) and its components Autocorrelation function (ACF), Partial autocorrelation (PACF), Normalized Bayesian Information Criterion (NBIC), Box – Ljung Q statistics and residuals estimated, ARIMA (0, 1, 1) was selected. Based on the chosen model, it could be predicted that the small *Haplochromine* species yield would increase to 4,224 tons in 2021 from 93 tons in 1976.

Key words: Forecasting, ARIMA, NBIC, Lake Malombe, Haplochromine, Modelling, production

————————————————◆————————————————

1 INTRODUCTION

Malaŵi ranks among the world's most densely populated and least developed countries. The fishing sector is important to both Malaŵi's economy and its overall food security, providing 300,000 – 450,000 jobs and 4% of GDP (FAO, 2008). However, most natural fish stocks in Malaŵi are either fully or over exploited. Coupled with the highest population growth rate in Southern Africa of 2.8 % annually (NSO, 2008), this has reduced per capita consumption of fish in Malaŵi from 14 kg per person per year in the 1970s, to less than 4 kg by 2008 (Jamu and Chimatiro, 2005), (FAO, 2008). These declining trends in fish catches and per capita availability of fish are reflected across Africa (Brummett *et al*. 2008). Additionally, Malaŵian fish catches are increasingly dominated by commercially-less valuable fish such as usipa (*Engraulicypris sardella*) and utaka (*Copadichromis* spp), while harvests of the commercially important Chambo (*Oreochromis* spp.) have declined from 23,000 tons in 1984 to 7000 tons in 2001 (NASP, 2004). Despite the declining per capita access to fish, fish stocks of Malaŵi account for an estimated 28% of total animal protein consumed in Malaŵi (Jamu and Chimatiro, 2005), one of the highest dependencies on fish for animal protein in the Southern Africa. This is particularly true for poorer Malawians, for whom fish may be the only regularly available source of animal protein. The demand for fish in Malaŵi is very high, as a result all fish that is caught is consumed locally. Because fish is on high demand, it is easily traded in both rural and urban communities. Overall, Malawians today find it more difficult and more expensive to obtain fish than ever before (PIAD, 2006).

———————————————————

- Wales Singini is currently pursuing PhD degree program in Fisheries Science Mzuzu University Malawi, Private Bag 201, Mzuzu, Malawi
  E-mail: walessingini@yahoo.co.uk
- Emmanuel Kaunda is a Professor in Aquaculture and Fisheries Science at University of Malawi.
- Victor Kasulo is an Associate Professor of Economics at Mzuzu University, Malawi
- Wilson Jere has a PhD in Fisheries Science based at University of Malawi.

In this background, this study was conducted to forecast pattern of small *Haplochromine* species yield in Lake Malombe, so as to help in formulating the needed strategies for sustainable management and conservation of the stock.

2 Materials and methods

As the aim of the study was to forecast pattern of small *Haplochromine* species yield in Lake Malombe, various forecasting techniques were considered for use. ARIMA model introduced by Box and Jenkins (1970) was frequently used for discovering the pattern and predicting the future values of the time series data. Tsitsika *et al*. (2007) used univariate and multivariate ARIMA models to model and forecast the monthly pelagic production of fish species in the Mediterranean Sea during 1990 – 2005. Jai Sankar and Prabakaran (2012) also used ARIMA model to model and forecast milk production in Tamilnadu during 1978 – 2008. Jai Sankar *et al*. (2010) also used stochastic modeling for cattle production and forecast the yearly production of cattle in the Tamilnadu state during 1970 – 2010. Jai Sankar (2011) used a stochastic model approach to model and forecast fish product export in Tamilnadu during 1969 – 2008. Among the methods based on univariate techniques, the ARIMA models by Box and Jenkins (1970) stand out because of their wide range of application. ARIMA model involved four steps: Identification, estimation, diagnostic checking and forecasting. To check for stationarity of the catch data, graphical analysis method was used. Model identification involved examining plots of the sample autocorrelograms and partial autocorrelograms and inferring from patterns observed in these functions the correct form of ARMA model to select. Gujarati (2004) pointed out that when the PACF has a cutoff at p while the ACF tails off, it gives an autoregressive of order p (AR (p)). If the ACF has a cutoff at q while the PACF tapers off, it gives a moving-average of order q (MA (q)). However, when both ACF and PACF tail off, it suggests the use of the autoregressive moving-average of order p and q (ARMA (p, q)).

Autoregressive process of order (p) is,

$$Y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \qquad (1)$$

Moving Average process of order (q) is,

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots, + \theta_v \varepsilon_{t-v} + \varepsilon_t \qquad (2)$$

Autoregressive moving average: ARMA of order (*p, q*) is,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots, + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots, + \theta_p \varepsilon_{t-p} + \varepsilon_t$$

$$\qquad (3)$$

The general form of ARIMA model of order (p, d, q) is

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + \varepsilon_t \qquad (4)$$

where Y_t is the observation at time t, ϕ and θ are coefficients and ε is an error term.

Model fitting consisted of finding the best possible estimates for the parameters of the tentatively identified models. In this stage, maximum likelihood estimation (MLE) method was considered to estimate the parameters. (MLE) method for estimation of ARIMA was applied in SPSS version 16.0. MLE runs an algorithm several times, using as the starting point the solution obtained in the previous iteration/run. Basically SPSS maximizes the value of a function by choosing the set of coefficient estimates that would maximize the function. Each time, it uses the estimates obtained in the previous iteration/run. In model diagnostics, various diagnostics such as the method of autocorrelation of the residuals and the Ljung-Box-Pierce statistic were used to check the adequacy of the identified models. If the model was found to be inappropriate, the process was returned back to model identification and cycle through the steps until, ideally, an acceptable model was found. Plots of autocorrelation and partial autocorrelation of the residuals were used to identify misspecification. For evaluating the adequacy of ARMA and ARIMA processes, various statistics like Correlogram of the residuals; Normalized Bayesian Information Criterion (BIC), R-square, Stationary R-square, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Maximum Absolute Percentage Error (MaxAPE), Mean Absolute Error (MAE) and Maximum Absolute Error (MaxAE) were used. In this study data on fish catch in Lake Malombe was generated from a computer based programme called Traditional Fishery Data Base (TFDB) from the Department of Fisheries Research Unit at Monkeybay, Government of Malawi from 1976 to 2011. The data were used to fit the ARIMA model to predict the future yield of small *Haplochromine* species in Lake Malombe.

3 Results And Discussion

Model Identification: ARIMA model was designed after assessing that transforming the variable under forecasting was a stationary series. The stationary series was the set of values that varied over time around a constant mean and constant variance. A dimension of the preliminary analysis for examining stationarity of the data was by considering the time series plots of small *Haplochromine* species as shown in Fig 1.
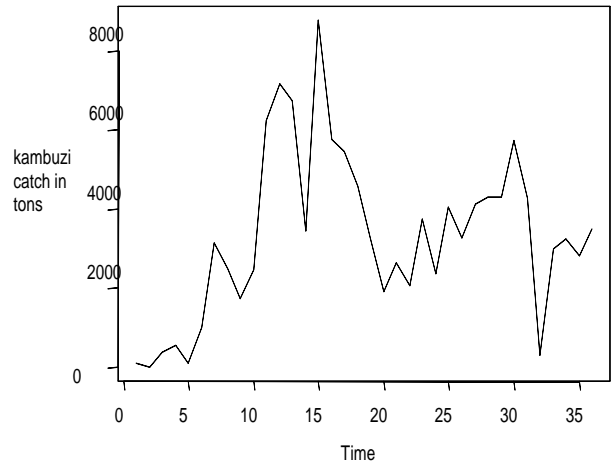

Fig 1: Time series plot of small *Haplochromine* species yield

The plot gives an initial clue about the likely nature of the time series. It is shown that the time series is non–stationary due to an unstable means which increase and decrease at certain points throughout the 1976 to 2011 period. The observations made by Pierce and Boyle (2003) and Georgakarakos *et al.* (2002) that use of ARIMA models require stationary time series but most fisheries time series are non-stationary support the preliminary analysis of this study. The non–nonstationarity in mean was corrected through first differencing of the data. The newly constructed variable Y_t was then examined for stationarity. Since, Y_t was stationary in mean, the next step was to identify the values of p and q. For this, the autocorrelation and partial autocorrelation coefficients (ACF and PACF) of various orders of Y_t were computed and represented in Fig 2 and Table 1. The process of making the series stationary agrees with Gutí´errez-Estrada *et al.* (2004) who showed that a good autoregressive model of order p (AR (p)) has to be stationary, and a good moving average model of order q (MA (q)) has to be invertible. Invertibility and stationarity will give a constant mean, variance, and covariance
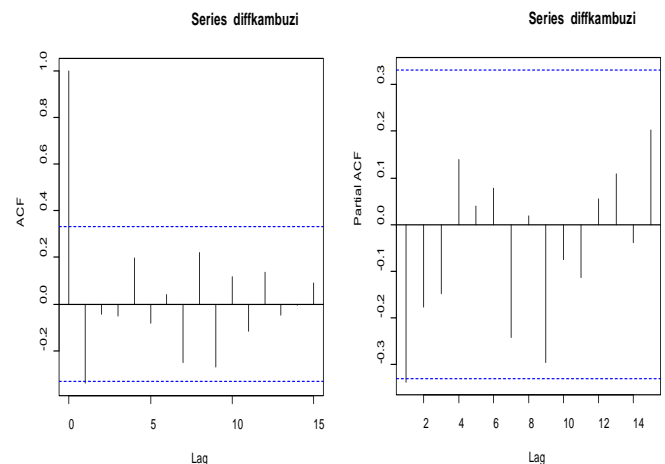

Fig 2. Autocorrelograms and partial autocorrelograms of first differenced data

Table 1: ACF and PACF for first order differenced scores of small Haplochromine species

| Lag | ACF | Std. Error | Box-Ljung Statistic | | | PACF | Std. Error |
			Value	df	Significance		
1	-.338	.162	4.357	1	.037	-.338	.169
2	-.042	.160	4.427	2	.109	-.177	.169
3	-.051	.157	4.531	3	.210	-.148	.169
4	.198	.155	6.173	4	.187	.139	.169
5	-.081	.152	6.457	5	.264	.039	.169
6	.042	.150	6.534	6	.366	.078	.169
7	-.251	.147	9.453	7	.222	-.243	.169
8	.219	.144	11.764	8	.162	.018	.169
9	-.270	.142	15.385	9	.081	-.297	.169
10	.118	.139	16.102	10	.097	-.075	.169
11	-.116	.136	16.827	11	.113	-.115	.169
12	.138	.133	17.900	12	.119	.054	.169

The tentative ARIMA models are discussed with values differenced once (d=1) and the model which had the minimum Normalized BIC was selected. The various ARIMA models and the corresponding Normalized BIC values are given in Table 2. The value of the Normalized BIC of the selected ARIMA was 15.09.

Table 2: Model selection criteria

Model	SR^2	R^2	RMSE	MAPE	MaxAPE	MAE	MaxAE	NBIC
Kambuzi (0,1,1)	0.146	0.849	1708.19	160.32	2488.31	1194.92	4487.98	15.09
Kambuzi (1,1,1)	0.146	0.853	1734.53	160.14	2486.80	1196.94	4467.05	15.20

Model Estimation: Model parameters estimated are presented in Table 3. Having obtained some suggested models the best possible estimates for the parameters were found by considering the final estimates of parameters and the model selection criteria. Gutiérrez-Estrada *et al.* (2004) indicated that quality of the coefficients has to meet the following requirement; it must be statistically significant for each coefficient of the estimated model. Czerwinski *et al.* (2007) indicated that the Normalized BIC test reveals that the model with the least Normalized BIC is better in terms of forecasting performance than the one with a large Normalized BIC. The most suitable model for small *Haplochromine* species yield was ARIMA (0, 1, 1), as this model had statistically significant coefficient, the lowest Normalized BIC, good R^2 and model fit statistics (RMSE and MAPE).

Table 3: Final estimates of small *Haplochromine* species ARIMA models

Model	Model type	Coefficient	SE Coefficient	T Value	P Value
Kambuzi(0,1,1)	Constant	92.215	174.06	0.53	0.600
	MA 1	0.409	0.16	2.56	0.015
Kambuzi(1,1,1)	Constant	92.107	174.59	0.526	0.602
	AR 1	0.031	0.43	0.944	0.004
	MA 1	0.434	0.393	0.278	0.002

Diagnostic checking: The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the selected ARIMA, which was done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. Abrahart and See (2000) argued that in order to achieve an acceptable model there is need to test whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent from each other and constant in mean and variance over time. For this purpose, various autocorrelations up to 12 lags were computed and are provided in Fig 3. The plots of ACF and PACF residuals show that the sample autocorrelation coefficients of the residuals are extremely low and lie within the limit of ±0.327 (1.96/square root 36) showing that none of the autocorrelations was significantly different from zero at any reasonable level. This proved that the selected ARIMA model was an appropriate model for forecasting small *Haplochromine* species yield. Hence, the fitted ARIMA model for small *Haplochromine* species yield data was:

$$Y_t = 92.215 + 0.409\varepsilon_{t-1} + \varepsilon_t \qquad (5)$$
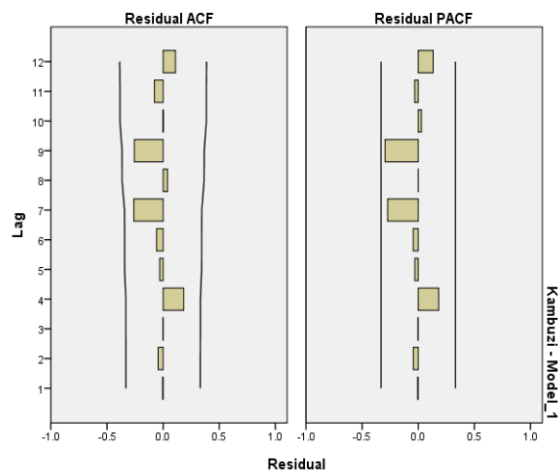

Fig 3. Residuals of ACF and PACF

Forecasting: To assess the forecasting ability of the fitted ARIMA model, the measures of the sample period forecasts' accuracy were computed. This measure also indicated that the forecasting inaccuracy was low. Table 4 shows the forecasted value of for small *Haplochromine* species yield (with 95%

71

confidence limit). Czerwinski *et al*. (2007) argued that a good model has a low forecasting error, therefore when the magnitude of the difference between the forecasted and actual values are low then the model has a good forecasting power and if the difference is high, then the model has a low forecasting power. This study noted that the magnitude of the difference between the forecasted and actual values were low for the selected model. The noise residuals were combinations of both positive and negative errors which shows that, the model is not forecasting too low on the average or too high on the average. Having this positive behaviour of the model, the model has outperformed as far as the forecasting power of the model is concerned. Figure 4 shows the actual and forecasted value of small *Haplochromine* species yield (with 95% confidence limit).

Table 4: Forecast of small *Haplochromine* species yield (in tons) in Lake Malombe

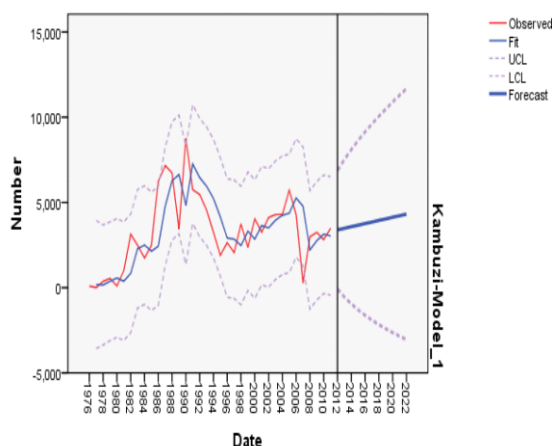| Year	Forecasted catch	95% confidence bounds
2011	3038.92	(-436.32,6514.16)
2012	3394.08	(-81.16,6869.31)
2013	3486.29	(-550.11,7522.70)
2014	3578.51	(-950.06,8107.07)
2015	3670.72	(-1301.52,8642.96)
2016	3762.94	(-1616.51,9142.39)
2017	3855.15	(-1902.78,9613.08)
2018	3947.37	(-2165.66,10060.39)
2019	4039.58	(-2409.01,10488.18)
2020	4131.80	(-2635.75,10899.34)
2021	4224.01	(-2848.12,11296.14)

Fig 4. Actual and forescated small Haplochromine species yield.

Conclusion: The most appropriate ARIMA model for small *Haplochromine* species yield forecasting was found to be ARIMA (0, 1, 1). From the forecast available from the fitted ARIMA model, it can be found that forecasted small *Haplochromine* species yield would increase to 4,224 tons in 2021 from 93 tons in 1976. That is, using time series data from 1976 to 2011 on small *Haplochromine* species yield, this study provides evidence on future small *Haplochromine* species yield in Lake Malombe, which can be considered for future policy making and formulating strategies for augmenting and sustaining small *Haplochromine* species yield.

Acknowledgments

REFERENCES

[1] R.J. Abrahart, and L. See, "Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments". Hydrol. Process. 14, 2157–2172, 2000.

[2] G.E.P. Box, and G.M. Jenkins, "Time Series Analysis - Forecasting and Control". Holden-Day Inc., San Francisco, 1970.

[3] R.E. Brummett, J. Lazard, and J. Moehl, "African aquaculture:" Realizing the potential. Food Policy 33, 371–385, 2008.

[4] I.A. Czerwinski, J.C. Juan Carlos Guti´errez-Estrada, and J.A. Hernando-Casal, "Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches." Fisheries Research 86: 120–128, 2007.

[5] FAO, "Malawi Fishery and Aquaculture Country Profile." Food and Agriculture Organization of the United Nations (FAO-UN), Rome, 2008.

[6] S. Georgakarakos, J. Haralabous, V. Valavanis, and D. Koutsoubas, "Loliginid and ommastrephid stock prediction in Greek waters using time series analysis techniques." Bull. Mar. Sci. 71, 269– 288, 2002.

[7] D.N. Gujarati, "Basic Econometrics." The McGraw−Hill Companies, 4th Edition, pp 792 - 866, 2004.

[8] J.C. Guti´errez-Estradade Pedro-Sanz, E. L´opez-Luque, and I. Pulido-Calvo, "Comparison between traditional methods and artificial neural networks for ammonia concentration forecasting in an eel (Anguilla Anguilla L.) Intensive rearing system." Aquat. Eng. 31, 183–203., 2004.

[9] Jai Sankar, and R. Prabakaran, "Forecasting milk production in Tamilnadu." International Multidisciplinary Research Journal 2012, 2(1):10-15 ISSN: 2231-6302, 2012.

[10] Jai Sankar, R. Prabakaran, K. Senthamarai Kannan, and S. Suresh, "Stochastic Modeling for Cattle Production Forecasting." Journal of Modern Mathematics and Statistics 4(2): 53-57, 2010.

[11] Jai Sankar, "Stochastic Modeling Approach for forecasting fish product export in Tamilnadu." Journal of Recent Research in Science and Technology 3(7): 104-108, 2011.

[12] D. Jamu, and S.K. Chimatiro, "Sustainable agro-pisciculture in Malawi." In: Agriculture and Rural Development: Contributing to International Cooperation, 12/2, 45- 46, 2005.

[13] National Statistical Office (NSO), "Population and Housing Census". National Statistical Office, Government Print, Zomba, Malawi, pp 3 – 7, 2008.

[14] G.J. Pierce, and P.R. Boyle, "Empirical modelling of interannual trends in abundance of squid (Loligo forbesi) in Scottish waters." Fish. Res. 59, 305– 326, 2003.

[15] Presidential Initiative on Aquaculture Development (PIAD), Government of Malawi Department of Fisheries, 2006.

[16] E V Tsitsika, C D Maravelias, and J. Haralabous, "Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models."Fisheries Science 73: 979-988, 2007.