

Asian Journal of Applied Science and Engineering

Abbreviated key title: Asian j. appl. sci. eng.

Access this article online
<http://journals.abc.us.org/index.php/ajase/issue/archive>

Modeling spatial non-stationarity of Chambo in South-East Arm of Lake Malawi

Patrick Likongwe^{1*}, John Kihoro², Murimi Ngigi³, Daniel Jamu⁴

¹Research Analyst, Natural Resource Management, WorldFish, Zomba, **MALAWI**

²Associate Professor, E-learning Manager, Department of Acturial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, **KENYA**

³Senior Lecturer, GIS, Department of Remote Sensing, Jomo Kenyatta University of Agriculture and Technology, Nairobi, **KENYA**

⁴Ecologist, WorldFish, Zomba, **MALAWI**

ARTICLE INFO

Received: Oct 10, 2014

Accepted: Dec 14, 2014

Published: Apr 24, 2015

*Corresponding Contact

Email:

alikongwe@gmail.com

Cell Phone: +265888879113

 Prefix 10.18034

ABSTRACT

Complex patterns and processes in aquatic resources are mostly analyzed under the assumption that ecological relationships do not vary within management areas. This assumption was questioned by studying the spatial distribution of the target fish population, *Oreochromis karongae* (Chambo) from South East Arm (SEA) of Lake Malawi where it is in abundance. Survey data from 2007 is used in the modeling of the spatial distribution. Three models, global logistic regression (GLR), generalized additive logistic model (GAM) and geographically weighted logistic regression (GWR) were run to explore the best model that can understand spatial non-stationarity better and how it affects fish distribution. GWR (AIC = 18.62) model explained significantly more variability than the global models GLR (AIC = 40.84) and GAM (AIC = 40.22). Adjusted R² explained 62.8% in GWR against 41.4% for GAM model. GWR is a better model than GAMs in understanding the spatial distribution of fish species.

Keywords: geographically weighted regression, logistic, *Oreochromiskarongae*, spatial distribution, South East Arm of Lake Malawi

Source of Support: Nil, No Conflict of Interest: Declared.

How to Cite: Likongwe P, Kihoro J, Ngigi M and Jamu D. 2015. **Modeling spatial non-stationarity of Chambo in South-East Arm of Lake Malawi** *Asian Journal of Applied Science and Engineering*, 4, 81-90.

This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Attribution-NonCommercial (CC BY-NC) license lets others remix, tweak, and build upon work non-commercially, and although the new works must also acknowledge & be non-commercial.



INTRODUCTION

There has been a decline in *Oreochromis karongae* (locally known as Chambo) fishery resource over the years throughout the lake. More research has been done on biomass and stock assessment (Bell, et al., 2012), biology (Banda, et al., 2005) and general factors on the

diminishing of *O. Karongae* (Bulirani, 2005). However, there has been limited spatial coverage of research programs that can help to provide spatial reference to factors that contribute to the decline. Since early 2000, spatial research has been one of the top research targets according to the Malawi's Fisheries Resource Management, Sustainability and Conservation Act of 1997 (Kachinjika, 2001).

In aquatic systems, there is dynamic spatial interactions between biological and environmental variables and the fisheries are highly mobile (Rose, 2005; Ciannelli, et al., 2008). As a result, one cannot assume stationarity of the processes under study as a rule. The Fisheries Research Unit uses generalized additive models (GAMs) in related analysis of fish distribution and stock assessments. It is important to explore the spatial effects on fisheries resource using other models like global logistic regression and geographically weighted regression. The study therefore modeled the spatial distribution of *O. Karongae* in SEA of Lake Malawi. It tried to identify the best model between global logistic regression (GLR), binomial generalized additive model (GAM) and logistic geographically weighted regression (GWR). The parameter coefficients from the best model were mapped for spatial interpretation of the observed dynamics.

METHODS

Study Area

The study covered the South East Arm (SEA) of Lake Malawi that shares its boundary with Mangochi district. There is a fisheries research station in Monkey Bay which is responsible for fisheries research on fishery resource and associated limnological aspects of Lake Malawi. Other similar but smaller stations are in Salima and Karonga in the central and northern regions of Malawi respectively. The lake is divided into these sections - Lake Malombe, Upper Shire River, South East Arm (SEA) in Mangochi, South West Arm (SWA) in Mangochi and Salima, Domira Bay in Salima, Nkhotakota, Likoma and Chizumulu Islands, Nkhata Bay and Karonga. The SEA under study has three fishing zones. These were demarcated for management purposes and done according to depth and fish abundance, designated as A, B and C, which is shallow, medium and deeper waters respectively. Figure 1 is a map showing these demarcations including the SEA and its fishing zones plus areas for fishery data collection. There are a total of 50 small polygons or areas from where fisheries surveys on the SEA of the lake take place, here referred to as locations.

Data Sources

Data used in this study was retrospective, collected from a multispecies fishery surveys conducted on the lake in October 2007 by the Monkey Bay Fisheries Research Unit. Data values for the shooting latitude and longitude were first collected at each trawl. After 30mins of continuous trawling, hauling latitude and longitude was also recorded including abundance of the various fish species available and depth of the catch. The fish were then sorted according to size and species, labeled and stored for further action like weighing and length measurements at the research station.

Sampling Method

The fish sampling surveys used MV Ndunduma vessel. It had these hauling specifications: velocity (V) of the trawl over the ground when trawling taken to be 3.5nm. The head rope (h, nm) length was 0.01242 and a trawling time (t) of 30mins per trawl. x_2 is that fraction (0.639) of the head-rope which is equal to the width of the path swept by the trawl, and the 'wing spread' is $h \times x_2$ (all these courtesy of Monkey-Bay Fisheries Research Unit). These are

the values used in the variogram when geostatistics is used to analyze the data. Attached to the vessel is a 38mm mesh size trawl net with a 38mm cod-end mesh size which is the minimum mesh size restriction for the trawl cod-end (Kachinjika, 2001). Sampling during the fishery survey was done in specific locations indicated above, guided by the coordinates that demarcate an ecological area. The whole SEA was represented by 50 different and independent samples, each denoted as a location with mean coordinates X and Y.

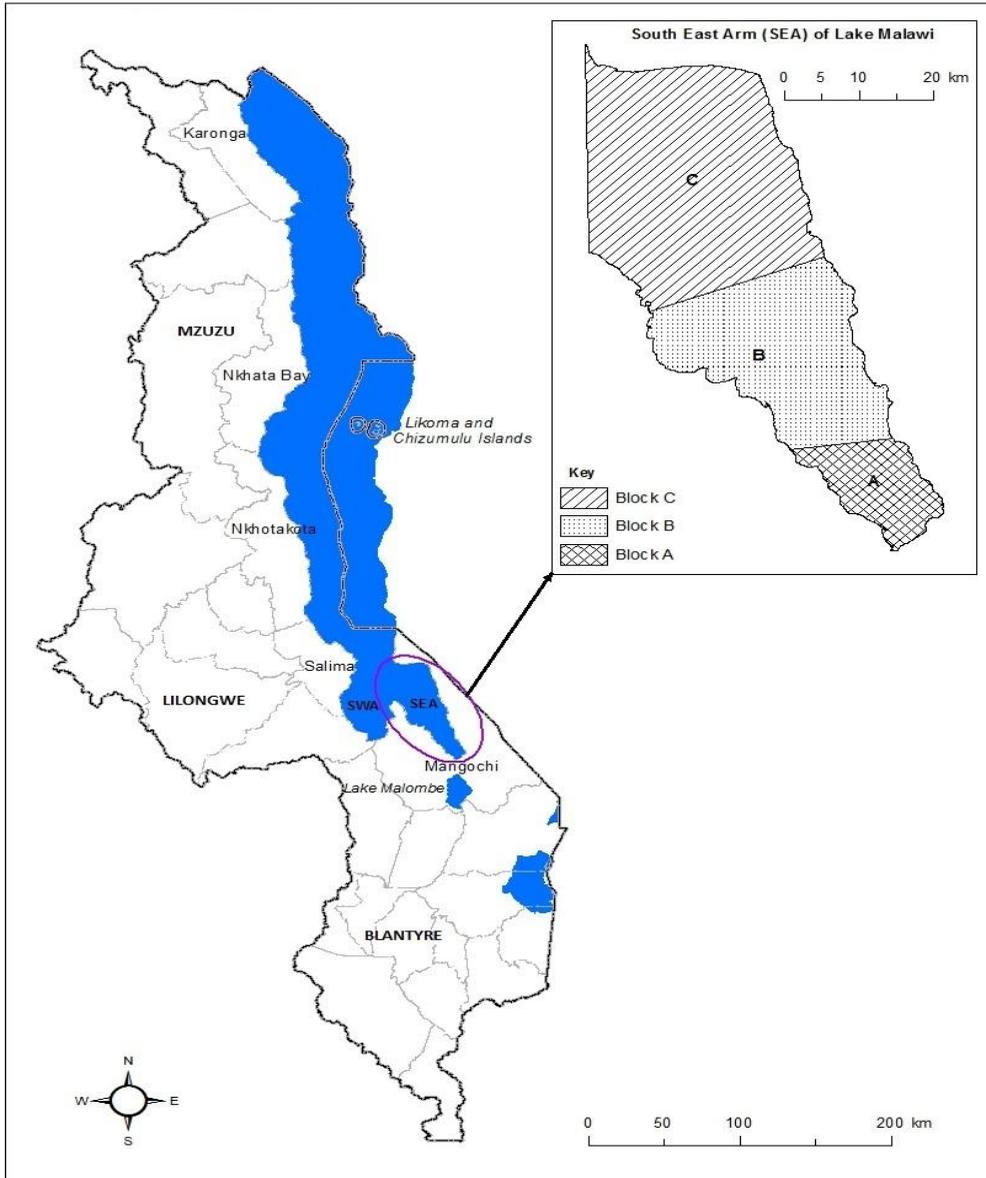


Figure 1: Map of Malawi showing the lake and the South East Arm where the study focused (Source: WorldFish, 2010)

Variables Definition

On modeling the presence and spatial distribution of *O. Karongae* in 2007, the dependent variable was either the presence or absence of *O. Karongae* (y_i) at a location based on data generated from the catch. If there was more than 0 kg in a catch, the generated dependent variable (CP07 in the data set) was coded as 1, zero otherwise. Primary and secondary data were used to analyze the probability of finding *O. Karongae* on a station. Depth (m), Distance (m) and Area covariates were used to explain the distribution of *O. Karongae* in the study area. Depth is the length into the water column from the surface to where the fish were caught within the water column, measured in meters. Distance is the shortest length measured from the nearest shoreline to the location where fish catches were done. Area is the zone that is composed of several locations. Table 1 show the variables used in the study.

Table 1: Description of variables used in the study

Variable	Description
Binary CP07	1 = Presence of <i>O. Karongae</i> at a station in 2007 Zero = Absence of <i>O. Karongae</i> at a station in 2007
Metrical Depth (m) Mdepth (m)	Distance from surface to where fish catches are done Mean depth from the 1999 and 2007 depth data
Spatial Location Area	50 stations where fish were caught, coordinate X,Y. 3 structured areas demarcated for fishery management

Modeling Approaches

Statistical approaches mostly used to analyze spatial distribution and abundance data can be loosely grouped into two categories. These are according to whether emphasis is either on the relationships among neighboring observations or on the relationship among the observations and the collected environmental variables (Ciannelli, et al., 2008). The first group is based on techniques developed for geographical analysis and mining resources (Matern, 1986) also known as geostatistical analysis. The second group is an extension of common regression techniques applied to spatial data (e.g. Guisan et al., 2002). Separately, the second technique captures important ecological processes of fish distribution. It is important to note that the analytical techniques above attempt to model the local species abundance (y_i) based on a similar underlying model of the type:

$$y_i = \mu_i + \varepsilon_i \quad \text{Equation 1}$$

where μ is a mean effect and ε_i is the error.

The following GLR, GAM and GWR models were fit and compared to find the best model that can explain the non-stationarity of *O. Karongae* much better.

$$M_{GLR}: \ln \left[\frac{p(x)}{1-p(x)} \right] = \alpha + \beta_1(\text{Depth}) + \beta_2(\text{Distance}) \quad \text{Equation 2}$$

$$M_{GAM}: \ln \left[\frac{p(x)}{1-p(x)} \right] = \alpha + \beta_1(\text{Depth}) + \beta_2(\text{Distance}) \quad \text{Equation 3}$$

$$M_{GWR}: \ln \left[\frac{p(x)}{1-p(x)} \right] = \alpha + \beta_1(\text{Depth}) + \beta_2(\text{Distance}) \quad \text{Equation 4}$$

$P(x)$ is the probability of finding *O. Karongae* while $1-P(x)$ is the probability of not finding it. In $[P(x)|1-P(x)]$ is the odds of finding *O. Karongae* given the covariates of depth and distance as the determinants of its presence or absence. The intercept is α while β_1 and β_2 are parameter coefficients for depth and distance variables used in GLR model (M_{GLR}). Furthermore f_1 and f_2 are smooth parameter coefficients for depth and distance used in GAM model (M_{GAM}) while β_1 and β_2 are parameter coefficients with j^{th} location for depth and distance as used in GWR model (M_{GWR}). Note that in logistic models, there is no error term as in linear models. In the GWR model, the t-statistic for each coefficient and local regression was run to inform us if there are places in our study area where the coefficient for a given variable is significantly different from the expected value. A Bonferroni test was run to check on the significance of the variation. The model run was tested using the Brudson, Fotheringham and Charlton F test (BFC) and Leung, Mei and Zhang F3 test (LMZ) which diagnoses the null hypothesis that the set of parameters tend to be constant over a region. BFC compares OLS model fit to GWR model fit using ANOVA test and the LMZ test tests for spatial variability of the parameter estimates. R software version 2.15 with relevant packages was used for the analysis.

RESULTS

The first statistical model (Equation 2) modeled the log odds of finding *O. Karongae* given depth and distance as the independent variables. The model produced these results (Table 2).

Table 2: Logistic regression coefficients and log odds

Coefficients	Estimate	Std. Error	Z value	Pr(> z)	Log Odds
(Intercept)	3.009	1.148	2.622	0.009**	20.276
Depth	-0.040	0.017	-2.299	0.022*	0.961
Distance	-0.0005	0.0003	-1.966	0.049*	0.999

Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1

The model and its explanatory variables explained 59.30% of the deviance. All parameters are significant (p-values<0.05) for depth and distance while the intercept is highly significant (p-value<0.009). The depth has an effect on the presence or absence of *O. Karongae* in the areas A, B and C followed by distance.

The GAM model was run (Equation 3) using the mgcv package, on the same depth and distance variables, now smoothed with the additive nature of a GAM model. The logit operation on the same random sample data set showed the intercept to be significant. Of the two smoothed variables, only distance was significant with (p-value<0.038)(Table 3). This model explained 44.8% of the deviance in the presence or absence of *O. Karongae* as compared to GLR model whose deviance is higher at 59.3%.

Table 3: GAM model results of *O. Karongae* presence/absence and log odds

Coefficients	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.353	0.773	-3.043	0.002**
Approximate significance of smooth terms:				
	Edf	Ref.df	Chi.sq	p-value
s(Depth)	1.749	2.194	3.879	0.165
S(Distance)	1.000	1.000	4.305	0.038*

Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1

GAMs are non-parametric extensions of linear model regressions. They apply non-parametric smoothers to each predictor variable and they additively calculate the response. When an ANOVA test for model evaluation was run, it gave a statistically significant difference in the models. The additive model described the relationship between presence/absence of *O. Karongae* and depth and distance much better. It used the Chi-square test for linearity which yielded a value of 0.28.

The GAM model plots for depth and distance is outlined in Figure 2. We see from these figures that the odds of the presence of *O. Karongae* is highest at depth below 55m and distance below 5,000m and reduces as these parameters increase in magnitude.

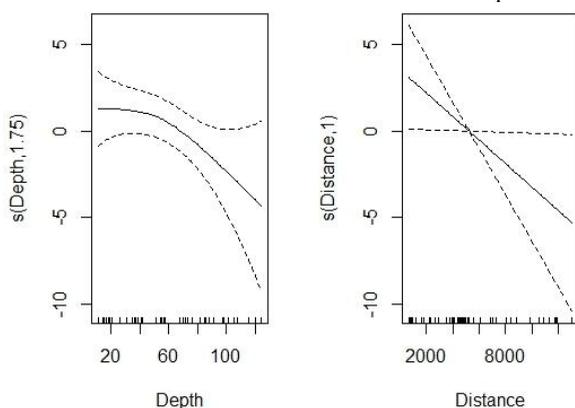


Figure 2: Smoothed GAM plots for Depth and Distance

Lastly, geographically weighted regression model (Equation 6) was run on the same variables of depth and distance. This time the inclusion of coordinates for location explained more of the deviance as compared to the two models of GLR and GAM. The GWR gave an R^2 value of 62.8, more than the GAM model which had an adjusted R^2 value of 41.4. The intercept and depth were significant whereas distance was not (range of values from the minimum to maximum includes a zero, Table 4). The odds of the presence of *O. Karongae* in SEA increased by 45.2% and reduced with an increase in depth and distance.

Table 4: Summary statistics of the logistic GWR parameter estimates

Variable	Min	1st Quart.	Median	3rd Quart.	Max	Global
(Intercept)	0.081	0.279	0.570	0.940	1.340	0.822
Depth	-0.0248	-0.0086	-0.0041	-0.0018	-0.0005	-0.0059
Distance	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000

From the three models on logistic regression, GAM and GWR; the GWR had a better AIC and AICc values as compared to GLR and GAM (Table 5) as applied to the distribution of *O. Karongae* in the SEA of Lake Malawi.

Table 5: Model comparison based on AIC and AICc values

Model	N	Ke	AIC	AICc
GLR	50	3	40.84	41.4
GAM	50	3	40.22	41.0
GWR	50	6.3	18.62	29.84

COMPARISON OF FIT FOR THE GLR, GAM AND GWR MODELS

GLR and GAM are not better as compared to GWR model. The difference in AIC values between GAM and GLR to GWR is 22, which signifies the significant difference in the models. A difference of >3 between AIC values from two competing models is assumed to represent significant differences between them (rule of thumb). The model with the smallest AIC provides the closest approximation to reality. The AIC is a relative goodness-of-fit statistic for comparing two or more competing models.

The value of AIC for GWR (AIC = 18.62) is lower than those of GAM (40.22) and GLR (40.84), indicating that GWR resulted in a significantly better fit for both variables. GWR was therefore used to explain the results of the coefficients from this model, followed by mapping of the coefficients. Effective number of parameters for GWR based models is a more general measure of model complexity unlike in OLS regression which is simply the number of linear coefficients in the model (Fotheringham et al., 2002).

GEOGRAPHICALLY WEIGHTED REGRESSION

The parameter value (Table 4) of the local parameter coefficients from descriptive statistics produced by GWR revealed much variance in the parameter value. This confirms the presence of spatial non-stationarity in the relationships between *O. Karongae* distribution and the explanatory variables. The depth and distance variables have negative parameter value, and the intercept is positive. As the depth and distance increases, chances of finding *O. Karongae* decreases.

The mean intercept value (0.569) gave an odds value of 1.77 while for depth it is 0.99. The odds of the presence of *O. Karongae* are high in SEA with 77% but decreases with an increase in depth. Related probabilistic studies on the presence of *O. Karongae* for the SEA have never been done, only descriptive studies on its presence presented as percentages.

MAPPING LOCAL GWR PARAMETERS

In GWR, the regression is re-centered many times (on each observation) to produce local GWR parameter results (Fotheringham, et al., 2002). When these local GWR results are combined, they generate a complete map of the spatial variation of the parameter estimates. This facilitates for ease of mapping GWR results unlike from global model results. Focus is on maps of parameter estimates and t-values as these are the most commonly reported maps in research using GWR. Data classification for t-values should account for certain exogenous criteria that are of importance to the variable being mapped, especially the threshold values that distinguish parameter estimates that are significant from those that are not (Fotheringham, et al., 2002; Mennis, 2006).

The statistical output of GWR typically includes a baseline global model result (parameter estimates), GWR diagnostic information, and a convenient parameter 5- number summary of parameter estimates. The 5-number summary of parameter estimates defines the extent of the variability in the parameter estimates. It is a summary based on the minimum, lower quartile, median, upper quartile, and maximum local parameter estimates reported in the GWR model (Table 4) and Monte Carlo test result for non-stationarity in each parameter. It is necessary to visualize the local parameter estimates and their associated diagnostics to better understand the 5-number summary and the Monte Carlo test. GWR models estimate local standard errors, derive local t-statistics, calculate local goodness-of-fit measures (e.g., R^2), and calculate local leverage measures. The output from GWR provides data that can be used to generate surfaces for each model parameter that can be mapped and measured. Once

mapped, each surface depicts the spatial variation of a relationship with the outcome variable (Fotheringham, et al., 2002; Mennis, 2006; Matthews & Yang, 2012).

Since depth is significant in the model, parameter estimates were plotted (Figure 3) followed by t-values (Figure 4) to visualize the variations in depth and their effect in determining the availability of *O. Karongae*.

The significant variable (Depth) is being mapped and from the residuals, the first quartile represents area A where the difference is minimal (0.0026). Connected to it is the second set which is also getting into area B as a transition between the two areas. The last quartile has residual values for the upper and right side of area C, each quartile sharing the same characteristics in the area of the same dotted colour. There is more variation in the upper quartile (area C) where there is no fish available.

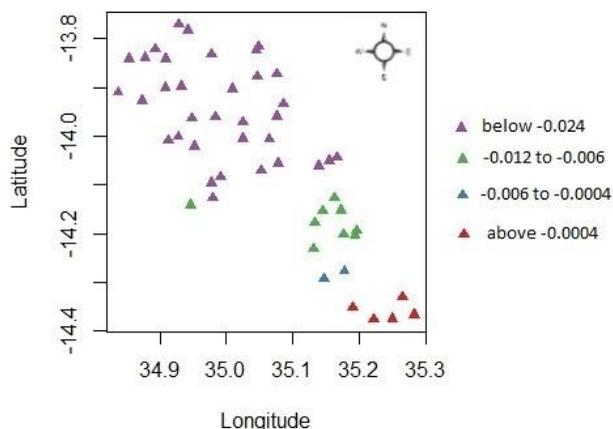


Figure 3: Map showing the geographical patterning of the depth parameter estimates

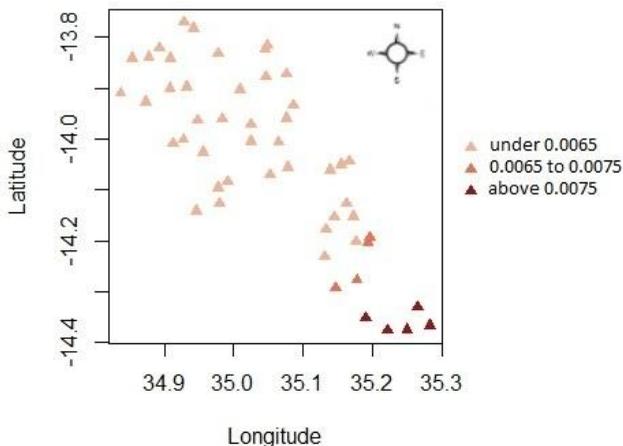


Figure 4: Estimated t-values for depth from the GWR model.

Figure 4 shows the t values that are not significant, in this case shown in faint red while those in light and dark shades of red are significant. With reference to area demarcation, it means areas A and part of B have their t values from the GWR model significant. The relationship between the variables in the model and the areas are significant, *O. Karongae*

being available in the far south and not available in the area C. Though the depth is transforming, with other areas having *O. Karongae* and others not, the same shade signifies the non-significance of the model coefficient residuals where there is no *O. Karongae*. If it were present, the modeling would also show the different shades of colours to signify the spatial non-stationarity of the variable explaining the behavior of the dependent variable.

DISCUSSION

Performance of the global models (GLR and GAM) is not superior to GWR. From the GAM model results and the plots, (Hastie & Tibshirani, 1990) outlines that the largest partial residual in the figures. This is seen as a potentially valuable observation since it corresponds to a presence (1) in a region with very low predicted probability. Geographically weighted regression is another recent technique which provides a method for understanding how regression model parameters vary across space. This is spatial non-stationarity in the process under study and spatial dependence. It further represents spatial modification to normal techniques, such as ordinary least-squares regression (Brunsdon, et al., 1996). Despite that GWR is a better model for exploring *O. Karongae* presence as compared to others; it still can do much better if other factors that influence presence of *O. Karongae* are considered like plankton abundance or lake levels. As Tweddle & Magasa, (1989); Bell, et al., (2012) reiterated that *O. Karongae* catch was related to lake height three years prior. Also productivity of all fish in the lake was related to primary productivity which is a function of the wind velocity and thermal structure of the water column. These factors were not included in the analysis. They further suggested that periods of declining lake heights occurred under conditions that enhanced nutrient upwelling and provided more food for the juvenile *O. Karongae*.

However, Bell et al. (2012) noted that there is a change in lake heights from three to two years due to a decline of larger fish such that the bulk of the catch becomes age-1 individuals. The total biomass however is a function of the environmental conditions as well as the anthropogenic factors around the lake. These anthropogenic factors in turn are tied to changes in the climate and the economics of the country. The lake height in turn affects the distance as well as depth at which *O. Karongae* can be caught. Apart from the lake heights, Bell et al. (2012) realized that the main driver of *O. Karongae* biomass was fishing pressure which was higher, making it almost impossible to achieve maximum sustainable yield during the entire time series studied (1976 to 2003). Despite lack of other data parameters to explain the relationship; GWR model has proved to be a far much better model than GAM and GLR. These results concur with (Windle et al., 2010) who also found that geographically weighted regression was superior in performance as compared to GAMs and GLR.

CONCLUSION

Among the competing models used in the analysis of *O. Karongae* distribution and its non-stationarity in the variables affecting its presence or absence, geographically weighted regression provided better results. The AIC from GWR is lower (18.62) against 40.84 for GLR and 40.22 for GAM. Of the parameters from the models used in the study, depth explained better confirming the availability and spatial distribution of *O. Karongae* that is mostly found at low depth and not in deeper waters. Again, the goodness-of-fit measure from adjusted R^2 explained 62.8% in GWR model against 41.4% from GAM. Depth in this case was significant and was further analyzed by mapping the parameter coefficients and related t-values for visualization of the non-stationarity aspect of a significant variable.

REFERENCES

- Banda M., Kanyerere G. & Rusuwa B.(2005)The status of the Chambo in Malawi: Fisheries and biology. In: Banda M., Jamu D., Njaya F., Makuwila M and Maluwa A (eds)*The Chambo Restoration Strategic Plan*. 13-16 May 2005. WorldFish Center Conference Proceedings, pp1-7.
- Bell R., Collie J., Jamu D. & Banda M. (2012) Changes in the biomass of chambo in the southeast arm of Lake Malawi: A stock assessment of *Oreochromis* spp.. *Journal of Great Lakes Research* **38** (4), 720-729.
- Brunsdon C., Fotheringham A. & Charlton M.(1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis***28**, 281-298.
- Bulirani A. (2005)Observation on the factors behind the decline of the Chambo in Lake Malawi and Lake Malombe.In: Banda M., Jamu D., Njaya F., Makuwila M and Maluwa A (eds)*The Chambo Restoration Strategic Plan*. 13-16 May 2005. WorldFish Center Conference Proceedings, pp 8-11.
- Ciannelli L., Fauchald P., Chan K. A. V. & Dingsor G. (2008) Spatial fisheries ecology: recent progress and future prospects. *Journal of Marine Systems***71**, 223-236.
- Cox O. R. (1970). *Analysis of Binary Data*. Meihuen.
- Fotheringham A. S., Brunsdon C. & Charlton M. (2002). *Geographically Weighted Regression: The analysis of spatially varying relationships*. New York: Wiley.
- Hastie T. & Tibshirani R. (1990). *Generalized Additive Models*. s.l.:Chapman & Hall.
- Hosmer D. & Lemeshow S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.
- Kachinjika O. (2001)A general overview of fisheries research and development in Malawi. In: Wely O.L.F. & Weyl M.V. (eds) *Proceedings of the Lake Malawi Fisheries Management Symposium*, 4-9 June 2001, Lilongwe: Department of Fisheries, pp 11-19.
- Matern B. (1986). *Spatial Variation [Early book providing mathematical detail for correlation functions]*. New York: Springer.
- Matthews S. & Yang T.C.(2012) Mapping the results of local statistics: Using geographically weighted regression. *Demographic Research* **26** (6), 151-166.
- Mennis J. (2006) Mapping the Results of Geographically Weighted Regression. *The Cartographic Journal***43** (2), 171-179.
- Mvula P., Njaya F. & Nkoko B. (2003). *Socio-economic factors influencing the decline of the Chambo*. In: Banda M., Jamu D., Njaya F., Makuwila M and Maluwa A (eds)*The Chambo Restoration Strategic Plan*. 13-16 May 2005. WorldFish Center Conference Proceedings, pp 30-37.
- Rose G. A. (2005) On distributional responses of North Atlantic fish to climate change. *ICES Journal of Marine Science***62**, 1360-1374.
- Tweddle D. & Magasa J. (1989) Assessment of multispecies cichlid fisheries of the southeast arm of lake Malawi, Africa.*International Council for the Exploration of the Sea***45**(2), 209-222.
- Windle M. J. S., Rose G. A., Devillers R. & Fortin M.J. (2010) Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science* **67**, 145-154.

AJASE!!!

"Speedy publication service, Online archives, Paperless, web-based peer review system, Open access policy, Indexing in world known citation databases, Global circulation, Broad international readership and authorship, Online submission system, Minimum publication charge"