NON-CLASSICAL APPROACHES TO THE STUDY OF THE ECOLOGY AND FISHERY OF *Rastrineobola argentea* (PELLEGRIN 1904) IN THE WINAM GULF OF LAKE VICTORIA, KENYA

BY:

MANYALA, JULIUS OTIENO

B. SC. FIRST CLASS (HONS.) – POONA UNIVERSITY

M. SC. (HYDROBIOLOGY) – UNIVERSITY OF NAIROBI

THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN FISHERIES AND AQUATIC SCIENCES OF MOI UNIVERSITY

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without the prior written permission of the author and/or Moi University.

Signature:

Manyala, Julius Otieno **DATE**
School of Natural Resource Management
Chepkoilel University College, P. O. Box 1125 - 30100
ELDORET, KENYA

This thesis has been submitted for examination with our approval as University Supervisors:

Signature:

Prof. Charles C. Ngugi PhD **DATE**
P.O. Box 58187 - 00200
NAIROBI, KENYA

Signature:

Prof. Mlewa Chrisestom Mwatete PhD **DATE**
Pwani University College, P.O. Box 195 - 80108
KILIFI, KENYA

Signature:

Prof. James M. Njiru PhD **DATE**
School of Natural Resource Management
Chepkoilel University College, P. O. Box 1125 - 30100
ELDORET, KENYA

ABSTRACT

Many schools of thought tend to suggest that the central assumption in classical fisheries models may not necessarily hold and thus there is need to explore new approaches such as Bayesian Belief Networks (BBN), Artificial Neural Networks (ANN), Nominal and Ordinal Logistic Regression. This study used non classical methods such as logistic regression, Bayesian Belief Network (BBN), Artificial Neural Network (ANN) and Weibull/Lognormal distribution to study food habits, production and recruitment of *R. argentea* in Lake Victoria for the first time. Significant ontogenic changes in stomach content was determined for *Thermocyclops oblingatus, Brachionus falcatus* and *Moina macrourus* ($p<0.0005$) as compared to the baseline (*Epiphanes spp.*) for the 30-50, 50 and 30-50 mm length classes respectively. The odds ratio was 10.25-11.42 times for T. oblingatus and *Moina macrurus* as compared to *Epiphanes*. The BBN show that the Root Mean Square (RMS) change for *Brachionus caudatus* (0.00221), *B. falcatus* (0.00217), *Epiphanes* (0.00207), *Keratella serrulata* (0.00268), *T. emini* (0.00233), *Bosmina longirostris* (0.00217) and *Daphnia lumholtzi* (0.00258) and *Trichocerca* (0.00207) had the highest sensitivity of food items in the stomach as compared to the environment while *B. calyciflorus, B. angularis* and *M. macrurus* had the lowest sensitivity. Maximum Spawning Biomass (SB) and egg production was at a size between 40 and 60 mm TL. Egg production was best explained by a polynomial relationship of the fourth order with r^2 of 0.959. Egg production, based on SB was significant for both Gamma and Weibull distribution ($p<0.00005$) according to the Shapiro-Wilks test. The location parameter was relatively consistent for both the Gamma (7,139) and Weibull (7,057) distributions, thereby providing similar recruitment threshold. Weibull distribution predicted a higher recruitment magnitude (scale parameter of 1,080,678) as compared to Gamma (354,600). The production modeling of *R. argentea* in Winam Gulf of Lake Victoria obtained the best ANN architecture of 10-9-1 based on environmental data and 12-6-1 based on fish catch statistics with 25 hidden layers and 30 hidden layers respectively, when the activation was based on the hyperbolic tangent function. Input importance analysis for environmental variables show that rainfall was the most significant variable (37%) followed by fisheries development classification (33%) and the lake level (17%) for environmental data. For fish catch statistics, the importance of fisheries development classification was 71.1%, *Lates* was 15.6%, *Haplochromis* was 6.6% and *Bagrus* was 4.2%. The actual catches versus output from the network had an average Absolute Error (AE) of 2,072 and 3,843 and an average Relative Absolute Error (RAE) of 14.2% and 20.7% for catch data and environmental data respectively. The ANN approach could be used to predict the catches of *R. argentea* in Lake Victoria during the different developmental stages of the fishery as well as projection of future production. Model data for both the environmental (r^2 =0.852) and fish catches ((r^2 =0.910) fitted well to the raw data. The non-classical methods offer robust alternatives for analysis of fisheries ecology data in light of data availability, nature of multispecies fishery and inadequacies of stock assessment models in tropical freshwater ecosystems. The study concludes that ordinal logistic regression best describes ontogenic changes in feeding while the BBN generated a stable feeding model for multiple food items. S-R relationship was best described by both Gamma and Weibull distributions for a given size at maturity, sex ratio, length-weight relationship and fecundity. The ANN consistently and adequately produced outputs that were consistent with target values from both environmental and catch data and could be used for predicting future values under varying fishing or environmental regime.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors Prof. Charles Ngugi, Prof. Mlewa Chrisestom Mlewa and Prof. James M. Njiru for their patience and guidance on this study. My thanks also go to Dr. John Gichuki who provided the environmental data on Lake Victoria. I am grateful to Mr. Peter Nzungi who provided fisheries catch data for Lake Victoria for the study. I thank my brother Prof. R. O. Manyala who gave me facilities for thesis writing in Zambia. I would like to thank all those who assisted me at various stages of this study who I have not mentioned by name but whose contribution were instrumental in the completion of this study.

CHAPTER ONE

1.0 INTRODUCTION

This chapter introduces the fishery of Lake Victoria in the light of introduced Nile perch (*Lates niloticus*) in Lake Victoria, its phenomenal growth in the fishery, dominance and decline for a period spanning about 40 years. The changes in species composition and various attempts to carry out stock assessment using various classical methods have yielded varying results. The chapter specifically gives an overview of stock assessment in the lake using classical methods, the challenges in using these methods and provides an insight into possible application of some non-classical methods that can be used in the analysis of ecology and fishery of *Rastrineobola argentea* in Lake Victoria. The advantages of these non-classical methods over the traditional and classical methods are provided as a basis for this study.

1.1 The Traditional Fisheries of Lake Victoria

Until the mid 1990s, the commercial fishery of Lake Victoria was dominated by high catches of the introduced Nile perch (*Lates niloticus* Linnaeus 1758) (CIFA, 1988; Getabu, 1988) which contributed about 95% of the total fish landing by weight. The trends of *Rastrineobola argentea* (Pellegrin 1904) fishery has seen a progressive increase as reflected in the catches from 4.5% in 1960s (CIFA, 1988) to 30% in 1980s (Ogari, 1985: Acere, 1988; Bwathondi, 1988) to about 40% in the 1990s (Asila *et al*., 1990) and over 50% in the 2000s (GoK, 2009). According to GoK (2009), *R. argentea* is currently dominating in catches by biomass (54%) as compared to Nile perch (42%) and Nile tilapia (*Oreochromis niloticus* Linn 1758).

The earliest indication of in drastic decline in catch rates in Lake Victoria was attributed to overfishing hence led to the introduction of minimum gillnet mesh size of 5” in 1931 (Graham, 1929). Later on, the decline in fish stocks of Lake Victoria was attributed to predation by Nile perch and overfishing by destructive fishing methods (Whitehead, 1958; 1959; van Someren, 1959); papyrus encroachment and habitat degradation (Balirwa and Bugenyi, 1980; Ochumba, 1984) and pollution (Ochumba, 1984; Ochumba and Kibaara, 1989). More recently, it has been suggested that the decline in Lake Victoria fishery is a result of inappropriate mechanism for controlling entry into the fishery of the primary elements of fishing effort, principally boats, gears and fishers; and the secondary factors like the fish processing factories, which influence fishing effort through market forces (Muhoozi, 2002; Tumwebaze *et al*., 2007)

1.2 Changes in Lake Victoria Fisheries

The development in the Lake Victoria commercial fishery is reflected in the artisanal catch statistics (Manyala, 2006). Since comprehensive catch data collection started in Lake Victoria in 1968, the availabe catch data can be divided into five main periods of perturbation associated with ecological changes in the ecosystem according to Manyala (2006):

i) Pristine: Before Nile perch explosion when the catch composition consisted of less than 1% *Lates niloticus* (1968-1970 benchmark)

ii) Growth: When Nile perch became increasingly significant, and catches consisted of up to 35% *Lates niloticus* (1971 – 1980 benchmark)

iii) Dominance: After the Nile perch explosion when the catch composition consisted of up to 54% *Lates niloticus* (1981 – 1990 benchmark)

iv) Decline: The decline phase of Nile perch when the catch composition consists of 50% *Lates niloticus* (1990 – 2000 benchmark)

v) Collapse: Collapse period (2001 – 2007), recovery phase where Nile perch has reduced to less than 40% and the Native *R. argentea* has increased to more than 50% of the total annual landings

Whereas 1968-1970 forms the baseline period before Nile perch explosion, 1981-1990 forms the explosion phase of Nile perch and 1990-2000 is the start of decline in Nile perch (Manyala, 2006).

1.3 Classical Methods of Stock Assessment in Lake Victoria and Challenges

Stock Assessment in Lake Victoria has been carried out in the Winam Gulf of Lake Victoria from time to time using bottom trawls (Kudhongania and Cordone, 1974; Marten *et al.*, 1976; Benda, 1981; Muller and Benda, 1981), catch assessment survey on artisanal fishery (Rabuor, 1988), length-frequency analysis (Getabu, 1988; Asila and Ogari, 1988; Manyala *et al.*, 1995a) and fisheries hydroacoustics (Getabu *et al.*, 2003). The main objectives of these assessments were to estimate the biomass of fisheries resource in the environment. However, all these assessments never answered the question of fisheries outputs (harvesting), and inputs (effort) and targets which form the basis and core of fisheries management measures, inputs to decision support system and means of formulating alternative harvesting strategies.

1.4 Non-Classical Methods and their Advantages

Bayesian inference is an important statistical tool that is increasingly being used by ecologists (Ellison, 2004). In a Bayesian analysis, information available before a study is conducted is summarized in a quantitative model or hypothesis: the prior probability distribution. Bayesian inference uses the prior probability distribution and the likelihood of the data to generate a posterior probability distribution. Posterior probability distributions are knowledge based alternative to p-values and provide a direct measure of the degree of belief that can be placed on models, hypotheses, or parameter estimates (Ellison, 2004).

Based on the EDA approach, a number of computational routines can be used to plot frequency data and determine the underlying type of distribution (Taylor, 2007) other than direct fitting of S-R data to a pre-determined model.

To overcome difficulties of non linearity in ecological data, Artificial Neural Network (ANN), which are known to be efficient in dealing with heterogeneous data sets constitute a relevant alternative tool to traditional fisheries assessment and statistical methods (Lek *et al.*, 2000). ANN is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist approach to computation (Reuter and Möller, 2010; Li *et al.*, 2011).

Raymond *et al*. (1999) showed that Artificial Neural Network (ANN) could be used to predict fish yields in 59 African lakes using a three-layered feed-forward a ANN.

The use of non-classical methods in ecology and fisheries data analysis therefore offers a number of advantages over the classical methods:

i) Bayes models have the advantage of using full posterior probability distributions (van Gils *et al*. 2003) to analyze feeding models. According to Maynard-Reid and Chajewska (2001), Salamó and López-Sánchez (2011) and Yang *et al*. (2011), BBN accurately represents the interaction between food items and can be used to determine stability, assess extinction risk and resilience to perturbation in an ecosystem for any fish species. According to Ainsworth *et al*. (2010), attaching a confidence limit to diet estimates offer a heuristic advantage when evaluating seasonal or onto-genetic shifts in diet, as the degree of overlap between sizes can be used to objectively determine the differences. Bayesian information-theoretic methods provide robust measures of the probability of alternative models, and multiple models can be averaged into a single model that reflects uncertainty in model construction and selection (Dickson and Ellison, 1996; Wade, 2001).

ii) No assumptions are required for EDA and unlike classical analysis, EDA does not impose any model (normality, linearity, etc.) and the analysis, estimation, and testing that follows are not focused on the parameters of that model. For EDA, the data collection is not followed by a model imposition; rather it is followed immediately by analysis with a goal of inferring what model would be appropriate (Minitab, 1997).

iii) A typical ANN can exhibit complex global behaviour, determined by the connections between the processing elements and element parameters (Reuter and Möller, 2010; Li *et al*., 2011). While a neural network does not necessarily have to be adaptive they are designed to alter the strength (weights) of the connections in the network to produce a desired signal flow (Ruck *et al*., 1990; Bishop, 1992; Mohammadzaheri *et al*., 2012). They can be used to model complex relationships between inputs and outputs or to find patterns in data (Bishop, 1992; Ruck *et al*., 1990).

1.3 Justification

There are a number of new approaches to dealing with various types of fish ecology, biology and fisheries data that seem to defy and negate the classical approaches. One of the central assumptions in classical fisheries data analysis is the steady state or equilibrium condition (Sparre *et al*., 1989); a condition that is both difficult to determine and is often violated in many classical models. Many schools of thought tend to suggest that the central assumption in classical fisheries models may not necessarily hold and there is need to explore new approaches such as Bayesian Belief Networks (BBN) (Ellison, 2004; Wade, 2001), Artificial Neural Networks (ANN) (Lek *et al*., 2000, Brosse *et al*., 2001), Nominal and Ordinal Logistic Regression (Manel *et al*., 1999; He *et al*., 2003), time series analysis like Auto Regressive Moving Average (ARIMA) (Manyala, 2000) and many other multivariate approaches (Grossman *et al*., 1998; Guégan *et al*., 1998; Manel *et al*., 1999). Since most of the information is *R. argentea* is available and the approaches to be used in collating, correlating and synthesis are model based, this study proposed and used an input-response or input-feedback mechanism in which any change in generated fisheries random variables can be simulated as many times as possible with an assumption that most of the data follows a stochastic process.

1.4 Problem Statement

Most of the stock assessment studies in Lake Victoria have been based on the use of classical models to predict possible biological and economic reference points. However, fisheries management, decision support system and management alternatives require analysis of input and outputs of the fishery in terms of effort corresponding yield. These classical models therefore requires reliable data on catch and the effort but unfortunately for Lake Victoria, there is a consistent collection of catch data but long term effort is difficult to get. Secondly, the classical models work best with single-species single-gear fisheries unlike the multi-gear multi-species nature of the Lake Victoria fisheries. The application of classical models makes very strong assumptions about the state of dynamic equilibrium in the fishery; a situation that is generally difficult to attain in the multi-gear multi-species fishery. This situation has resulted in resilience of the endemic cyprinid *R.argentea* in Lake Victoria that defies these classical models. In addition to the environmental changes observed in Lake Victoria in the last twenty years, it has not been possible for fisheries biologists working on stock asseement in the lake to provide unequivocal advice on stock status, biological reference points and optimal harvest strategies.

This study sets out to provide an alternative to analysis of ecology and fisheries data using non-classical methods which are known to have several advantages over their classical counterparts on the basis of making no assumptions, use of probability abbroaches and exploratory data analysis procedures.

1.5 Objectives of the Study

1.5.1 General Objective

The general objective of this study was to synthesize the available scientific information on the ecology and fishery of *R. argentea* based specifically on the available data in Kenya and generally in the region using non-classical modeling approaches.

Available datasets for the analyses included:

i) Count of food items in stomach and environment (Manyala, 1994)

ii) Data sources on sex ratio, fecundity, L-W relationship (Okedi, 1971; 1973; Wandera, 1992; Manyala *et al.*, 1992; Wanink, 1989; Manyala *et al.*, 1995a; 1995b; LVFO, 2005).

iii) Catch Assessment Survey (CAS) data (GoK, 2008) with catch data from 1968 to 2007, environmental variables such as average annual temperature, rainfall, river discharge and lake level (Mwirigi *et al.*, 2005) for 1950 to 2005 were obtained but only matching time series data from 1968 to 2005 (with projections for 2006 and 2007)

1.5.2 Specific Objectives

The specific objectives of this study were based on the following:

1.5.2.1 Food and feeding habits

i) Determine the relative change in each food items with size, using ordinal logistic regression.

ii) Develop a feeding model based on probabilities of food items in both the stomach and environment using Bayesian Belief Network.

1.5.2.2 Sexual stage of maturity, size at maturity and fecundity

i) Estimate recruitment threshold, magnitude and elasticity using information on sexual stage of maturity, sex ratio and fecundity to estimate recruitment threshold, magnitude and elasticity.

ii) Determine the best stock-recruitment relationship based on a family of distributions such as Log-normal, Gamma and Weibull.

1.5.2.3 Catch Analysis and Production

i) Develop artificial neural network to needed predict production of *R. argentea* in Lake Victoria, based on specific environmental and catch data.

CHAPTER TWO

2.0 LITERATURE REVIEW

Chapter 2 of this study reviews literature on Lake Victoria, its limnology, the general biology of *R. argentea,* classical and non-classical methods in fisheries, covering the analysis of food and food habits, recruitment and production modeling.

2.1 Geographical Location and Size of Study Area

The data used in this study were gathered from various studies and reports on Lake Victoria and its fisheries. The major portion of Kenyan waters of Lake Victoria is a narrow gulf, known to various authors by several names such as the Victoria Nyanza (Graham, 1929), Kavirondo Gulf (Copley, 1953; Muller and Benda, 1981), Nyanza Gulf (Ogari and Dadzie, 1988) and the Winam Gulf (Okach and Dadzie, 1988). The Winam Gulf has an area of approximately 1920 km^2 with a length of about 60 km and width varying between 6 and 30 km. The Winam Gulf lies between 34° 13' and 34° 52' East of latitude 0°, 0° 4' and 0° 32' South of the equator. The gulf has a mean depth of 6 m and a maximum depth of 43 m while its surface is at an elevation of 1136 m above sea level. Its irregular shoreline is about 300 km, with several large bays. The major affluent rivers include the Kibos and Nyando to the East and Sondu, Awach, Mogus and Lambwe to the South (Okach and Dadzie, 1988).

Water exchange with the rest of the lake takes place through the Mbita channel while the major outflow from the lake is the river Nile. Bottom deposits found within Winam Gulf include hard substrates of sand, gravel and bedrock in exposed areas and mud, silt and clay deposits in areas adjacent to the river mouths. Large quantities of both living and dead gastropods and bivalves are common in sheltered bays (Ogari and Dadzie, 1988).

2.2 Limnology of Lake Victoria

Limnological research in Lake Victoria was mainly done in the northern part near Jinja (Fish, 1957; Talling, 1957; 1966; 1969), and in the Mwanza Gulf (Akiyama *et al*., 1977); the brief visits of Worthington (1930); Melack (1976; 1979) and Kalff (1983) in the Nyanza Gulf and observation in Kenyan open Lake Victoria by Ochumba and Kibaara (1989). Apart from the works of Fish (1957); Newell (1960); Kitaka (1971) and Kite (1981), no detailed hydrographic description of the lake over a long period of time covering dry and rainy seasons is available. Although the above works provide a baseline for the limnological characterization of the lake, they are short of indicating hydrological factors affecting biological productivity on a long term basis.

The first limnological studies (Talling, 1957, 1966, 1969) were focused on the seasonality of phytoplankton photosynthesis and abundance, and their relationship to thermal and oxygen regimes. These classic studies provide the historic benchmark against which many ecological changes in Lake Victoria has been gauged. In the 1980s and 1990s the Kenya Marine and Fisheries Research Institute carried out extensive research and laboratory tests in the Winam Gulf of Lake Victoria (Ochumba and

Kibaara, 1989; Calamari *et al.*, 1995). Ochumba and Kibaara (1989) observed that the blue-green algal blooms in the open waters of Lake Victoria were caused by a combination of high temperature, release of nutrients from river inflows, upwelling and from sediment re-suspension into the euphotic zone. The results of this investigation also showed that the blooms declined as a result of physical flushing, temperature reduction associated with rainy season and nutrient exhaustions.

Hecky (1993) examined changes in the lakes environmental parameters – temperature, oxygen, chlorophyll, silicon, nitrogen and phytoplankton biomass. His findings indicated that environmental degradation resulted from high human population in the catchment, biomass burning, shallow mixing depths as a result of changing regional climate, and low flushing times. This study concluded that enormous effort, social transformation and investment from the international community would be required to stem the damage. Preliminary assessment of pollution levels in Winam Gulf conducted by Calamari *et al*. (1995), quantified urban industrial and agricultural loads, and related these to geographic and climatic condition. Extensive measurements of water currents, temperature dissolved oxygen and winds on the Kenyan waters of the lake were done by Worthington (1930), Fish (1957) and Talling (1966) have shown low hypolimnion temperatures below 24ºC. Hecky *et al*. (1994) concluded that low oxygen conditions are now more extensive and persistent than previous investigators had found.

Zooplankton dynamics in Lake Victoria were outlined by Bransrator *et al*. (1996), who suggested that the composition of cladocerans, calanoid copepods and cyclopoid capepods in the modern community were largely unchanged from historical conditions although the proportions may have changed. The results of this study also showed that changes in the fish community of Lake Victoria may have led to the establishment of *Daphnia lumnoltzi* var. *monacha* in the zooplankton community.

The phytoplankton community structure in the Kenya waters of Lake Victoria have been described and related to environmental conditions (Lung'ahyia *et al*., 2000). These authors identified 103 species of phytoplankton with blue-green algae (Cyanophyceae) being the most diverse, followed by diatoms (Bacillariophyceae), green algae (Chlorophyceae) and dinoflagellates (Dinophyceae). Seasonal variations in the gulf and open lakes were observed. Chlorophyll concentrations confirmed increasing phytoplankton biomass in Lake Victoria since the 1960s.

The ecosystem changes in Lake Victoria as reflected in sedimentary lithostratigraphic units and anthropogenic organic compounds were studied by Hecky (1984). The study showed that organic rich sediments have been deposited for the last 200 years but that the nature of the organic matter and diatom microfossils had changed over the past 40 years likely due to eutrophication affecting the lake. The detailed palaeolimnological records (Hecky, 1984) showed that increases in phytoplankton production developed from the 1930s onwards, which parallels human-population growth and agricultural activity in the Lake Victoria drainage basin. Hecky *et al*. (1994) noted a dominance of bloom-forming cyanobacteria since the late 1980s which coincided with a relative decline in diatom growth, which could be attributed to the depletion of dissolved silica resulting from 50 years of enhanced diatom growth and burial (Hecky *et al*., 1994). Further, eutrophication-induced loss of deep water oxygen started in the early 1960, and may have contributed to the 1980s collapse of indigenous fish stocks by eliminating suitable habitat for certain deep water cichlids (Gikuma-Njuru and Hecky, 2005; Stager *et al*., 2009).

According to Gor *et al*. (2005), Abira *et al*. (2005) and Tamatamah *et al*. (2005), the main lake (including the littoral stations) had higher PO_4-P (56.5 µg l^{-1}) as compared to the gulf (19.4 µg l^{-1}) while the spatial analysis of NO_3-N concentration revealed three zones with similar concentration (88.9 to 90.8 µg l^{-1}). SiO_2-Si average concentration decreased along the gulf from 4.51mg l^{-1} to 1.28 mg l^{-1} while the values in the main lake ranged between 0.6 mg l^{-1} to 0.84 mg l^{-1}. Total nitrogen and total phosphorus showed opposite spatial variation along the gulf into the main lake while Dissolved Organic Phosphorus (DOP) ranged between 0.022 and 0.046 mg l^{-1} and was higher in the main lake than in the gulf but Dissolved Organic Nitrogen (DON) was higher in the gulf than in the main lake and had values ranging between 0.34 and 0.47 mg l^{-1} (Gikuma-Njuru and Hecky, 2005).

Limnological studies in the 1980s (Hecky, 1984) and early 1990s (Hecky, 1993) showed that conductivity was higher in the gulf than in the main lake, showing a decreasing trend from 161.8 µS cm^{-1} to 98.2 µS cm^{-1} but remained at an average value of 101.1 µS cm^{-1} in the main lake. The sudy by Hecky *et al*. (1994) showed that water transparency (Secchi depth) was higher in the open water (2 - 3.5m) than in the inshore areas (0.4 – 1.6 m) and varied exponentially with Total Suspended Solids (TSS) and chlorophyll-a. In the shallow littoral stations (<20 m), dissolved oxygen was higher throughout the water column (>5mg l^{-1}) than in the deep pelagic stations, where it reduced with depth and was anoxic below 30m during stratification period (January-March and August-November). Many limnological studies in the 21st Century have concentrated on nutrient loading from urban and municipal environments (Abira *et al.*, 2005), non-point loading (Gor *et al.*, 2005), monitoring of the pelagic, littoral, river mouths and near shore urban environments (Mwirigi *et al.*, 2005) for nutrients, physico-chemical parameters, phytoplankton, zooplankton, micro-invertebrates, primary productivity, hydrodynamics and thermals.

Phytoplankton biomass (chlorophyll-a) showed a reducing trend along the gulf into the main lake with the highest average value of 21.1 $\mu g\ l^{-1}$ and lowest value of 6.9 $\mu g\ l^{-1}$ Silsbe *et al.* (2006). Chlorophyll-a had a maximum value of 13.9 $\mu g\ l^{-1}$ at about 3m depth in the gulf but decreased to 5.4 $\mu g\ l^{-1}$ at 50m. The main lake had an average value of 18.18 $\mu g\ l^{-1}$. The maximum biomass in the gulf and the in the main lake was observed in August-October (29.6 $\mu g\ l^{-1}$ and 15.8$\mu g\ l^{-1}$ respectively). According to Silsbe *et al.* (2006) and Hecky *et al.* (2010), chlorophyll-a showed a negative variation with PO_4-P and Inorganic Nitrogen (IN) compounds, but varied randomly with SiO_2-Si. The pH increased with chlorophyll-a in the upper depth (0-5m) and Dissolved Oxygen (DO) decreased with depth from 5m depth (7.4 to 3.2 mg l^{-1}). Both the studies show that in both littoral and pelagic areas, Cyanobacteria was the most abundant, contributing between 45 and 65% of the total phytoplankton abundance and diatoms contributed between 20 and 40% of total abundance. The main lake had higher relative diatom abundance than the gulf. Littoral stations recorded the highest density of phytoplankton (320.6±86.6 individuals l^{-1}).

2.3 General Biology and Ecology of *R. argentea*

The food and feeding habits of *R. argentea* in Lake Victoria has been studied by Wanink (1989) and Wandera (1992). The species is reported to feed on zooplankton (mainly copepoda) during the day but peak feeding periods have been reported during the night. Differential feeding patterns between the juveniles and adults however still remains poorly investigated, which is partially attributed to vertical migration in response to diel vertical zooplankton migration. Similar observations have been made on the haplochromis species (Kudoja *et al*., 1992). Very few published studies are available on the food and feeding habits of *R. argentea* from 1994 onwards (Manyala, 1994).

Most of *R. argentea* caught in the lake range between 26 and 40 mm SL. The length at 50 % maturity is at 38 mm SL for females and 39 mm SL for males (Wanink, 1989; Wandera, 1992; Manyala *et al*., 1992; LVFO, 2005). Okedi (1971) analyzed 604 specimens from Winam Gulf, Mwanza Gulf, Bukoba and Musoma and found a female to male sex ratio of 1.6:1. Okedi (1971) also found that out of 2952 specimens examined in Ugandan waters 1027 were males and 1925 were females, giving a sex ratio of 1.8:1, while Wandera (1992) found an overall sex ratio of two females to one male (2:1).

The maximum length attained by *R. argentea* in various regions of the lake varies between 64 and 69 mm SL; Uganda waters with 69 mm SL (Wandera, 1990; 1992) and Kenya waters with 64 mm SL (Manyala *et al.*, 1992; 1993;) respectively. There is also variation of size at maturity with sex, with most studies indicating that females mature at bigger sizes than males. Studies also reveal that the size at 50% maturity amongst females decreases from inshore to offshore waters (Wanink, 1995; Wandera, 1999; LVFO, 2005).

Rastrineobola argentea produces floating eggs in the lake (Graham, 1929). Earlier works have reported that egg production (fecundity) of *R. argentea* increased with fish size (Okedi, 1971). Okedi (1971) estimated mean fecundity at 2292 ova (range 582 - 4771) while Wanink (1989) found that the fecundity of *R. argentea* was related to the total length (TL) according to the relationship:

$$\text{Fecundity (number of eggs)} = 0.005875 \bullet TL^{2.95} \qquad \textit{Equation 1}$$

Manyala *et al*. (1992) found egg production to vary from 170 to 1350 eggs for specimens of 4l-60 mm SL with the relationship expressed as:

$$\text{Fecundity (number of eggs)} = 3.3 \bullet 10^{-7} \bullet TL^{5.376} \qquad \textit{Equation 2}$$

According to Wanink (1989) the Apparent Fecundity (AF) based on Standard Length (SL mm) other than Total Length (TL mm) was expressed as:

$$AF = 0.0092 \bullet SL^{2.97} \qquad \textit{Equation 3}$$

Rastrineobola argentea nursery grounds are in the shallow sheltered areas of the lake. In the Kenya waters most of the larvae were found mostly in the sheltered bays such as

Kisumu, Homa and Asembo Bays and near river mouths (Manyala and Ojuok, 2007). Wanink (1989) reported that in Mwanza Gulf, juveniles of about 10 mm SL are present at depths between 0-30 m. After spending their larval stage in the shallow areas, juvenile *R. argentea* are thought to migrate away from the shore with highest densities of adults occurring at a distance of 2 km from shore (Wanink *et al*., 2002). Observations made in Uganda waters of the lake show that *R. argentea* do not move far offshore and different populations occur over short distances (Wandera, 1990; 1992; LVFO, 2005). Tumebwaze (2003) and Tumebwaze *et al*. (2007) however showed lakewide distribution of *R. argentea* using hydroacoustic assessment method, thereby shedding light into the constraints imposed by limited sampling in earlier studies.

Earlier studies by Okedi (1973) indicated that the species breeding season spreads from June to August on a lake-wide basis. Based on the condition factor (K_n) of *R. argentea*, Manyala *et al*. (1995b) reported maxima of breeding in April/May and December/January in the Winam Gulf. However, Wandera (1992) observed that *R. argentea* breeds throughout the year with peaks in April/May and August/September in The Machison Bay.

Several workers have reported on the population and growth parameters of *R. argentea* in Lake Victoria. Table 1 provides a summary of the available information on these aspects of the species biology which is highly variable between studies.

Table 1: Growth characteristics of *rastrineobola argentea* in Lake Victoria, where L_∞ (mm) is asymptotic length, K yr^{-1} is growth curvature, M yr^{-1} is natural mortality coefficient, F yr^{-1} is fishing mortality coefficient, Ø' is growth performance index and Z yr^{-1} is total mortality coefficient

L_∞	K	M	F	Ø'	Z	Region	Author(s)
67.8	0.58	0.88	1.98	2.86		Winam Gulf	Manyala *et al.*, 1995a
64.5	0.92	2.37	1.22	3.59		Uganda waters	Wandera, 1992
63.4	0.94				3.23	Winam Gulf	Manyala *et al.*, 1992
52.0	1.14					Mwanza Gulf	Wanink, 1989
59.0	0.74	1.12		1.89	3.47	Winam Gulf	Manyala *et al.*, 1995a
62.0	0.74	1.12	1.39	2.97		Winam Gulf	Manyala *et al.*, 1995a
58.0	0.68	1.07	1.80	3.38		Winam Gulf	Manyala *et al.*, 1995a
62.0	0.66	1.04	1.45	3.03		Mbita Area	Manyala *et al.*, 1995a
58.0	0.63	0.99	1.77	3.35		Open lake	Manyala *et al.*, 1995a
61.0	1.42					General Study	Wanink, 1989
53.0	3.00					Station G 1988	Wanink, 1989

2.4 Review of Classical Methods in Fisheries

This section reviews the existing knowledge, methods approaches and application of classical methods in the study of food and feeding habits in fishes, Stock-Recruitment (S-R) relationships and surplus production models including; Virtual Population Analysis (VPA), Cohort Analysis and Yield Per Recruit Model. The chapter further traces the development of fisheries models along the age-structured and size-structured approaches. The summary of the chapter observes that fisheries research has utilized a series of tools in the past, focusing on biological research instead of addressing technological advances and environmental interactions in the fishery sector.

2.4.1 Food Types and Food Selection

Food selection in fishes is often analyzed using the Ivlev's Index of Electivity (E) (Ivlev, 1961) or Forage Ratio. The Vanderploeg and Scavia's Electivity Index (Vanderploeg and Scavia, 1979) has been successfully used to examine spatio-temporal variability in planktivore predation by *Coregonus hoyi* and *Alosa pseudoharengus* in Lakes Michigan and Ontario respectively. The Vanderploeg and Scavia's Electivity Index is a weight based version of the the Chesson (1978) alpha selectivity coefficient. The most important food items in fish can also be determined by using the Index of Relative Importance (IRI) of Pinkas *et al*. (1971) and Chabot and Maly (1986) or the Jacob's Electivity Index (Jacobs, 1974) which is a version of the Ivlev's Electivity Index.

Several workers have noted some shortcomings of using classical methods for food analysis. For example, Strauss (1979) and Paloheimo (1979) have pointed out that Ivlev's Electivity Index and the Forage Ratio are significantly biased when the sizes of the prey samples from the gut of the predator and the habitat are unequal. The statistical reliability of each index was found to be a function of the absolute and relative sample sizes and the relative abundances of the prey species in the environment. Strauss (1979) has proposed a linear index of food selection which avoids most of the statistical and mathematical inadequacies of traditional electivity indices. Paloheimo (1979) also proposed an index of the preference that is independent of the prey abundance based on standardized forage ratios; standardized so that the forage ratios for the different prey species sum to one. According to Lecowicz (1982), quantification of feeding preferences is necessary for determining optimum foraging and for quantitative description of feeding ecology.

Despite the frequent use of all the food electivity indices, no comprehensive comparisons of their characteristics have been made.

2.4.2 Stock-Recruitment Relationship

Despite the earliest attempt to mathematically describe the stock-recruitment relationship (Thompson and Bell, 1934), it was Ricker (1954) who developed the classical stock-recruitment model while Paulik (1973) developed the generalized stock-recruitment relationships. Thereafter, many stock-recruitment models have had strong linkages to Baranov (1918) and Beverton and Holt (1956) age structured type of stock models. Stock–recruitment theory was first applied to the North American salmon (Ricker, 1954; Larkin and Ricker, 1964), where salmon runs could be readily censused and adult catches compared with smolt production in the progeny generation.

According to Ricker (1954), recruitment occurs when the fish reach a size where they can be caught by conventional gear. There is a close relationship between gear selectivity and recruitment. In the early models, the selection length equals the length at which 50% of the fish are caught. Age at recruitment (t_r) or length at recruitment (L_r), is normally equal to the lowest practical selection length. Ricker (1954) and Beverton and Holt (1957) state that the relation between parent stock size and recruitment is one of the most crucial factors in the regulation of a fishery but noted that it was still not possible to formulate a satisfactory model to predict recruitment.

2.4.3 Catch and Production Models

According to Caddy (1999), the origins of three schools of thought in fish stock assessment and some of their linkages date back to the first two decades of 1900s. The first school of thought is the global models based on whole populations under fishing pressure, derived from models of human demography or the analytical cohort model (Hjort and Petersen, 1905; Baranov, 1918). This school of thought linked recruitment success to parental stock size through the integrated theory of fishing and culminated to the yield per recruit model of Thompson and Bell (1934).

The second school of thought was based on the concept of whole population and was developed into the logistic model using the North Sea groundfish (Graham, 1935). The logistic model gave rise to the Schaefer (1954) surplus production model and the Gordon (1954) economic model. There are a number of further elaboration approaches of the Schaefer-Gordon models using variable geometry and error structure of production (Pella and Tomlinson, 1969), incorporation of mortality rates in production (Csirke and Caddy, 1983), delay of the impact of fishing effort on stocks (Deriso, 1980; Schnute, 1985; 1987) and observation error or time series fitting of production functions (Hillborn and Walters, 1992). Thus the second school of thought is associated with the surplus production models and relies on mortality and delay of the impact of fishing on the fish stocks.

The third school of thought focused on the analytical models, using size and age structure. This school of thought is associated with the von Bertalanffy (1938) growth model, the Virtual Population Analysis (VPA) (Fry, 1949) and the general analytical theory of population dynamics (Beverton and Holt, 1956). The catch-curve approach (Ricker, 1949) developed in parallel and at the same historical time period as the VPA. The VPA concept was further picked up and perfected by Gulland (1983), further developed into cohort analysis (Pope, 1972) and multispecies VPA (Pope and Knight, 1982).

According to Caddy (1999), the surplus production theory found its first management context in fisheries such as that for Pacific tunas, where age reading is impossible but catch and effort data are readily available. The surplus production model provided the only target reference point mentioned in the Law of the Sea Convention, the Maximum Sustainable Yield (MSY), which until the 1970s was regarded in most world areas as the appropriate target for management. The broad generality provided by this simple biomass model allowed early application of economic theory (Gordon, 1954) and led to the Maximum Economic Yield (MEY) as a 'target reference point' to the left of MSY on the fishing effort axis. The MSY/MEY approach gave rise to the concept of management by 'escapement' (Ricker, 1954; Larkin and Ricker, 1964).

The three schools of thought on age structured models, the surplus production models and the analytical models tend to be associated with corresponding management modes:

i) Production modelling is used to generate Total Allowable Catch (TAC) but more logically points to fishing effort as the control variable and hence advocates for constant exploitation strategies

ii) Yield Per Recruit (Y/R) analysis and VPA provide a theoretical basis for quota management;

iii) Stock–recruitment models lead directly to management by an optimal level of escapement. As such, stock–recruitment models form the basis for ‘minimum Spawning Biomass’ (SB) limits in many fisheries

There are still a number of challenges in using classical fisheries stock assessment methods. According to Frøysa *et al*. (2002), the classical `book-keeping' methods assume that the reported catch numbers at age are exact. They also utilize assumptions about natural mortality and about relationships between abundance indices and stock size. On the other hand, age-structured assessment models sometimes termed statistical “Catch at Age Analysis” (CAGEAN) (Fournier and Archibald, 1982; Deriso *et al*., 1985) fits a self-contained population model to the data. This is different from the commonly used VPA-based methods, where the stock abundance numbers and fishing mortalities are derived directly from catches-at-age. In particular, the reported catch numbers at age are not assumed to be exact.

The conversion of early age-based methods (Beverton and Holt, 1956) into size-based methods greatly developed due to computation power of modern computers. The size-based methods (Sims, 1985; Thiam, 1986; Sparre, 1987; Sparre *et al*., 1989) are therefore technology-driven. Many of the computer-based methods are still based on the ‘book keeping’ mechanical aids such as log paper (Bhattacharya, 1967). Despite the increased application of stock assessment methodology and tools, more attention is still placed on biological research instead of the rapid technological changes that are revolutionizing the fishing sector. The situation therefore calls for a new approach based on variables that are dependent on both technological advances and environmental information. Besides, the same technological advances would provide the required technological capabilities to deal with the complex fishery-environment interaction and the non-classical methods offer an alternative way of analyzing many types of fisheries data.

Fitzpatrick (1995) estimated the relative value of the "technology coefficient" calculated for 13 different types of fishing vessels ranging from super trawlers (of 120 meters) to pirogues (of 10 meters) in 1965, 1980, and 1995, taking the value of the coefficient in 1980 as a basis. On average, this coefficient increased from 0.54 and 0.26 in 1965 to 1.0 in 1980 (the basis) and 2.0 to 0.9 in 1995. Improved knowledge of fleet–stock interactions at the appropriate ecosystem scale is necessary to build and parameterize the integrated models required for integrated ecosystem assessment (Levin *et al*., 2009) and operating models in management strategy evaluation frameworks (Peterman, 2004) or to address more general questions such as the ecological impact of rising fuel costs (Sumaila *et al*., 2008). In a single-stock, single-fleet perspective, classical population dynamics models provide appropriate answers. But when it comes to multispecies, multi-fleet fisheries, fleets depend on several fish stocks (Daurès *et al*., 2009), and stocks are exploited by several competing fleets (Rijnsdorp *et al*., 2008). Fleet behaviour changes in response to various factors, including technological progress, management regulations, and resource availability (Baelde, 2001; Christensen and Raakjær, 2006).

2.5 New Approaches in Analysis of Fishery Data

In recent times, there are quite a number of new approaches in fisheries biology and ecology and many such approaches tend to revolve around artificial intelligence, probability theory and non-conventional distributions applicable to bio-physical systems.

2.5.1 Food Types and Selection

Many of the non-classical methods still require distributional assumptions that should be evaluated before analysis. Resource selection functions (Manly *et al*., 1972) provide a unifying theoretical framework for selection study techniques, including many of the methods in use. For a given study design and data type, functions are defined that yield estimates of the probability, or a value proportional to the probability that a resource unit will be selected. There have been recent attempts to obtain estimates of true probability of selection based on use-availability data, but these methods require evaluation of complex likelihoods using Monte Carlo methods (Johnson *et al*., 2008; Lele, 2007; Horne *et al*., 2008) or machine learning tools (Phillips *et al*., 2006).

The non-classical approaches to food habits and food selection include logistic regression (Manly *et al*., 1972), polytomous logistic regression (North and Reynolds, 1996), matched-pair logistic regression and Mahalanobis distances (Clark *et al*., 1993; Knick and Rotenberry, 1998), Discriminant Analysis (Dunn and Braun, 1986), Principal Components Analysis, animal telemetry locations by means of General Linear Model (GLM), Poisson or Negative Binomial distributions (Marzluff *et al*., 2004), discrete choice models (McDonald *et al*., 2006), and a linear model regressing the height of the estimated utilization distribution on habitat characteristics (Marzluff *et al*., 2004). These methods allow for exploratory analyses and provide information to researchers about which characteristics or resources are selected.

A comparison of top-down and bottom-up trophic food selection models has been attempted by using non-classical Bayesian Information Criterion (BIC) with weighs improvement (Claeskens and Hjort, 2008; Argeant *et al*., 2012). It was found that BIC provides more consistent model selection as model complexity increases as compared to the more common Akaike's Information Criterion (AIC). BIC applies a higher penalty than AIC for adding more parameters in the model estimation (Claeskens and Hjort, 2008; Burnham *et al*., 2011) and is more conservative than AIC in the presence of model uncertainty. Existing literature tends to promote the use of information-based decision methods for food selection studies using probability approaches such as model selection procedures (Johnson and Omland, 2004), Markov Chain Monte Carlo (MCMC) (Johnson *et al*., 2008), Bayesian Belief Network (BBN) (Kaedi and Ghasem-Aghaee, 2011) with case-based reasoning or logistic regression (North and Reynolds, 1996)

2.5.2 Stock-Recruitment Relationship

The application of non-normal distributions such as lognormal and gamma for fisheries data analysis has also been used mainly in recruitment (R) and spawners (S) relationships (Myers and Pepin, 1990). Myers *et al*. (1995) further outlined the interpretation of S-R relationship for the the classical Stock-Recruitment (S-R) models such as the Power, Ricker, Beverton and Holt and the Shepherd models (Bellows, 1981). All these studies proposed many other functional forms for the relationship between spawners and recruitment such as Lognormal, Gamma and Weibull, with 3 to 4 parameters. Since the alternatives usually have more parameters which are difficult to estimate, there is scarcity of their application in fisheries. However, with modern computational power of the modern digital computers and softaware (Caddy, 1999), it has become easier to apply and test many alternative fisheries models including S-R relationships (Myers *et al*., 1995). Specifically, a number of routines (Ahrens and Dieter, 1982; Choi, 1994; Limpert *et al*., 2001; Swamee, 2002) and computer software (Taylor, 2007) are now available for analysis of probability and statistical sistributions.

Among the many other functional forms of probability distribution, the following families of distributions have a potential of describing the spawners-recruitment function:

1. **Lognormal distribution:** In probability theory and statistics, the logistic distribution is a continuous probability distribution (Holgate, 1989; Swamee, 2002; Park *et al*., 2009). Its cumulative distribution function is the logistic function, which appears in logistic regression and feed-forward neural networks.
2. **Weibull distribution:** The Weibull distribution (named after Waloddi Weibull) (Weibull, 1951) is a continuous probability distribution. The Weibull distribution is often used in the field of life data analysis due to its flexibility; it can mimic the behavior of other statistical distributions such as the normal and the exponential.
3. **Gamma distribution:** The variance-gamma distribution is a continuous probability distribution that is defined as the normal variance-mean mixture where the mixing density is the gamma distribution. The tails of the distribution decrease more slowly

than the normal distribution. It is therefore suitable to model phenomena where numerically large values are more probable than is the case for the normal distribution. The distribution was introduced in literature by Madan and Seneta (1990). The variance-gamma distributions form a sub-class of the generalized hyperbolic distributions. The advantage of Gamma distribution is that there is a simple expression for the moment generating function which implies that simple expressions for all moments are available.

2.5.3 Catch and Production

Raymond *et al*. (1999) showed that Artificial Neural Network (ANN) could be used to predict fish yields in 59 African lakes using a three-layered feed-forward a ANN. The empirical approach for the selection of the network consisted of a test for the number of different possible configurations and the selection of the one that provided the best compromise between bias and variance according to Geman *et al*. (1992) and Kohavi (1995). Results from the study were compared to multiple linear regression using a leave one out procedure (Efron and Gong, 1983; Jain *et al*., 1987) and showed more consistent prediction. The leave one out procedure is similar to the Jacknife Sampling (Sokal and Rohlf, 1995).

The true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modelled (Cybenko, 1989). Traditional linear models are also inadequate when it comes to modeling data that contains non-linear characteristics as compared to neural network according to (Hornik *et al.*, 1989).

The most common neural network model is the Multi-Layer Perceptron (MLP) (Callan, 1999; Luger, 2005). This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown (Luger, 2005).

CHAPTER THREE

This chapter describes describes the physical characteristics of Lake Victoria and summarizes the economic importance of its fisheries. The non-classical methods used for analysis of food and feeding habits such as multinomial logistic regression and BBN are described. The chapter also describes how stock-recruitment relationship was analyzed through simulation and concludes by describing the use and procedures in ANN for modeling production based on environmental and catch data.

3.1 Physical Characteristics of the Lake Victoria and Fish Production

Lake Victoria is the largest of the African Great Lakes, with a surface area of 68,800 km^2 and contains about 2,750 km^3 of water. The lake receives its water primarily from direct precipitation and thousands of small streams. The largest stream flowing into this lake is the Kagera River, the mouth of which lies on the lake's western shore (vanden Bossche and Bernacsek, 1990). Two rivers leave the lake, the White Nile (known as the "Victoria Nile" as it leaves the lake), flows out at Jinja, Uganda on the lake's north shore, and the Katonga River flows out at Lukaya on the western shore connecting the lake to Lake George. Lake Victoria occupies a shallow depression in Africa and has a maximum depth of 84 m and an average depth of 40 m. Its catchment area covers 184,000 km^2. The lake has a shoreline of 4,828 km, with islands constituting 3.7% of this length, and is divided among three countries: Kenya (6% or 4,100 km^2), Uganda (45% or 31,000 km^2) and Tanzania (49% or 33,700 km^2) according to Prado *et al*. (1991). This study is focused on the Kenya portion of Lake Victoria.

Lake Victoria is the most productive freshwater fishery in Africa. Fishery yield from the lake is of the order of magnitude of 800,000 – 1,000,000 mt valued at US $ 350 – 400 million at the beach, with export earnings estimated at US $ 250 million. The fishery is supported by three main important fish stocks, the Nile perch (*L. niloticus*), *R. argentea* and Nile Tilapia (*Oreochromis niloticus*). Over 75% of the Nile perch is directly to the fish processing factories for export while *R. argentea* and tilapia are serving the regional and local markets (GoK, 2010).

3.2 Food and Feeding Habits

3.2.1 Nominal Logistic Regression of Food Type by Fish Size

In the present study, ontogenic changes in food habits of *R. argentea* in Lake Victoria were analyzed based on numerical counts of zooplankton species in the stomach of different sizes of *R. argentea* (Table 2).

Table 2: Food selection data in numbers used in logistic regression (Source: Manyala, 1994)

Food item	Fish Size range (mm)				
	10 -20	21 - 30	31 - 40	41 - 50	>50
Thermocyclops emini	0	39	308	28	0
T. oblingatus	63	259	636	1860	2359
D. lumhortzii	11	8	0	3	0
Ceriodaphnia riguadi	0	0	0	20	0
Moina macrourus	0	8	380	1641	2626
Bosmina longirostris	3	8	162	74	34
Brachionus caudatus	28	105	28	31	2
B. falcatus	1	20	2	0	0
B. calyciflorus	0	2	12	0	0
B. angularis	0	30	4	12	0
Keratella serrulata	0	8	7	12	0
Epiphanes sp.	0	145	4	7	4

The regression was performed using multinomial logistic regression equation (Minitab, 1997; Brown 1982; Agresti, 1984; 1990) given as:

$$g(\chi_j) = \theta_i + x_j' \beta, i = 1,...,k-1$$ *Equation 4*

where:

k	=	the number of distinct values of the response or the number of possible events
χ_j	=	the cumulative probability up to and including event i for the jth factor/covariate pattern
$g(\chi_j)$	=	the link function (described below)
θ_i	=	the constant associated with the ith distinct response value
x_j'	=	a vector of predictor variables associated with the jth factor/covariate pattern
β	=	a vector of coefficients associated with the predictors

The ordinal logistic regression was used with the logit link function given as:

$$g(\chi_j) = Log_e(\chi_j / (1-\chi_j))$$ *Equation 5*

Application of logistic regression on numerical count data of various prey shown in Table 2 (Manyala, 1994) resulted in response information for each prey (Table 3). Response information displays the number of observations that fall into each of the response categories (zooplankton food items). The response value that has been designated as the reference event is the first entry under value for prey items, defined as *Epiphanes sp*.

Table 3: Response information for the ordinal logistic regression of food items versus length size (Source of raw data: Manyala, 1994)

Response variable value	Count	
Epiphanes sp.	160	(Reference Event)
Keratella serrulata	27	
Brachionus angularis	46	
Brachionus calyciflorus	14	
Brachionus falcatus	23	
Brachionus caudatus	194	
Bosmina longirostris	281	
Moina macrourus	4,655	
Ceriodaphnia riguadi	20	
Daphnia lumholtzi	22	
Thermocyclops oblingatus	5,177	
Thermocyclops emini	375	
Total	**10,994**	

Among the prey items, the copepod *Thermocyclops oblingatus* (5,177) and the cladopceran *Moina macrourus* (4,655) were the most abundant in the stomach. The copepod *Themocyclops emini* (375) and the cladocera *Bosmina longirostris* (281) were found in moderate numbers.

The only rotifers found in moderate quantities were *Brachionus caudatus* (194) and *Epiphanes sp.* (160). The other two cladocera, *Ceriodaphnia riguadi* and *Daphnia* *lumholtzi* were found in relatively low numbers in the stomach. Factor Information in the model was considered to be the length-classes 10 mm, 20 mm 30 mm, 40 mm and 50 mm resulting into 5 levels for the factor (Table 4). The factor level that has been designated as the reference even is the first entry under values which is 10 mm. Here, the default coding scheme defines the reference level as 10 mm using alpha-numeric order. The length-classes was based on life cycle strategies to examines the food of immature specimen (<35 mm) and mature specimen (≥35 mm) TL.

Table 4: Factor information for the ordinal logistic regression of food items versus length size (Source: Author)

Factor information	Value	Count	
Group	10 mm	106	(Reference event)
	20 mm	632	
	30 mm	1,543	
	40 mm	3,688	
	50 mm	5,025	
	Total	**10,994**	

3.2.2 Bayesian Belief Network (BBN)

The mean numerical counts in both the environment within the Winam Gulf of Lake Victoria and in the stomach of *R. argentea* were converted to proportions (Table 5) in order to satisfy the requirements for application of Bayesian Belief Networks (BBN) (Ellison, 2004; Dickson and Ellison, 1996; Olsson and Holmgren, 1999).

This analysis provided the posterior probabilities of food selection given the prior and conditional probabilities of the occurrence of different species in both the environment and the stomach of *R. argentea*. The BBN provided a dynamic model that was capable of recalculating new probabilities wherever there is any change in one or more food items in the environment. Both datasets used in the analysis came from Manyala (1994) as shown in Table 2 and 5.

Table 5: The average numerical abundance of food items in the stomach of *R. argentea* and the environment and the respective calculated proportions R_i and P_i (Source: Manyala, 1994)

Food items	Stomach	Environmnt	R_i	P_i
Thermocyclops emini	2.65	29.31	0.00788	0.64779
Thermocyclops oblingatus	173.56	8.53	0.51585	0.18843
Daphnia lumhortzii	0.08	5.20	0.00024	0.11489
Moina macrourus	151.17	0.00	0.44930	0.00000
Bosmina longirostris	4.45	0.00	0.01323	0.00000
Brachionus caudatus	1.30	1.74	0.00385	0.03836
Brachionus falcatus	0.12	0.00	0.00034	0.00000
Brachionus calyciflorus	0.12	0.41	0.00034	0.00907
Brachionus. angularis	0.53	0.00	0.00158	0.00000
Keratella serrulata	0.37	0.00	0.00111	0.00000
K. quadrata	0.00	0.07	0.00000	0.00146
Trichocerca sp.	0.11	0.00	0.00032	0.00000
Epiphanes sp.	1.99	0.00	0.00591	0.00000
K. cochlearis	0.02	0.00	0.00005	0.00000

A Bayesian feeding model was constructed by creating nodes that represented each of the food items and two additional nodes that represented the proportion of food items in the environment and in the stomach thereby representing all the variables or food items. An arcs represent statistical dependence relations among the food items and local probability distributions for each food item, given values of its parents was then created between each food item, its proportion in the environment and also the proportion in the stomach and all the nodes representing each food item.

The created Bayesian Belief Network (BBN) consequently represented the dependence of each of the food items in the stomach and its dependence on the food items in the environment with a given prior probability and on condition that the food item occurs in the stomach with a given prior probability, thereby providing a conditional/joint probability for that food item. Statistically, for each food item X_i, i= 1 to n, and a set of parent variables denoted by parents (X_i), the conditional probability distribution of the variables is product of the local distributions:

$$Pr(X_1, \ldots, X_n) = \prod_{i=1}^{n} Pr(X_i \mid \text{parents}(X_i))$$

Equation 6

3.3 Modelling Fecundity and Recruitment of *R. argentea*

Historical data (Table 6) on sex ratio (Okedi, 1973; Wandera, 1992), fecundity relationship (Manyala *et al.*, 1992) and length-weight relationship (Manyala, 2005b) were used to determine female parental breeding biomass from sample data. The sample length-frequency sample data of *R. argentea* in Lake Victoria were obtained from the LVFO (2005) lakewide survey on the gear selectivity, maturity and catch rates.

Appendix I give details of the sample length-frequency data and how the processing was carried out before modeling of stock-recruitment relationship. All samples were grouped into 1 mm size classes according to recommendations of LVFO (2005).

Table 6: Biological parametes used in stock-recruitment analyses and their sources (Source: Author)

Parameter	Equation	Location	Author
Sex ratio:	1.6:1	Tanzania	Okedi (1973)
	1.8:1	Uganda	Okedi (1973)
	2.0:1	Uganda	Wandera (1992)
Fecundity:	2292 (582-4771)	Tanzania	Okedi (1971)
	860 (170 - 1350	Kenya	Manyala *et al*. (1992)
	$F=5.875 \bullet 10^{-3} \bullet TL^{2.95}$	Tanzania	Wanink (1989)
	$F = 3.3 \bullet 10^{-7} \bullet TL^{5.376}$	Kenya	Manyala *et al*. (1992)
L-L relationship:	$TL=1.74+1.11 \bullet SL$	Kenya	Manyala *et al*. (1995a)
L-W relationship:	$W=0.0000025 \bullet TL^{3.4}$	Kenya	Manyala *et al*. (1995a)

Since the data were collected using Standard Length (SL mm), the entire individual SL measurements were converted to Total Length (TL mm) according to Manyala *et al.* (1995b). The weight (g) for each TL (mm) was then computed from the relationship $W=0.0000025 \bullet TL^{3.4}$ (Manyala *et al*., 1995b).

The frequency distribution was split into equal parts for males and female based on a sex ratio of approximately 2:1 (Okedi, 1973; Wandera, 1992; Manyala *et al*., 1992). The number of mature females was calculated from the original LVFO (2005) survey data and used to calculate the percentage of mature fish for each size class in the sample. The total biomass for each size class was calculated by multiplying the number of all female fish in that size class by the individual weight of fish in the same size class. The mature biomass was obtained by calculating the percentage of mature biomass as a proportion of the total biomass for each size class. The values under mature biomass column in Table 5 provide the independent variable in the Stock-Recruitment relationships of *R. argentea*. Relationship between size class and Spawning Stock Biomass (SSB) and between length size and egg production were compared graphically. The resulting relationship between spawning biomass and egg production provided a basis for further analysis of the SSB (stock) and recruitment (egg production). The SSB was considered to be based on the cumulative biomass from size TL_i to TL_n and was calculated for every size class in the breeding category as:

$$\sum_{TL=i}^{TL=n} W_i N_i \qquad \textit{Equation 7}$$

where,

W_i = Weight of fish of length TL_i

TL = Total length of the fish

N_i = Number of fish in length TL_i

The resulting potential stock-recruitment data in terms of egg production and cumulative SB was used to determine the best distribution of recruitment potential and to generate the parameters for Gamma (Banneheka and Ekanayake, 2009), Weibull (Weibull, 1951) or Lognormal (Holgate, 1989; Limpert *et al.*, 2001; Swamee, 2002) family types of distribution. The preliminary distribution parameters of egg production frequency were used to simulate and generate a large amount of egg production frequencies (10,000) for recruitment analysis and modeling. This approach allowed the determination of three parameter model for recruitment to describe the magnitude, elasticity and biological reference point for maximum recruitment, on the basis of SSB and egg production.

Based on the descriptive statistical summary of the frequency distribution of egg production and on a sample data, 10,000 new datasets were generated based on

Lognormal, Weibull and Gama family of distribution, their parameters (Location, Shape and Scale) determined and tested for consistency using the Skewness-Kurtosis all Tests (Taylor, 2007). Using the mean, standard deviation, skewness and kurtosis, the 10,000 new data set were generated according the the following procedures:

The generalized three parameter gamma: where ε>0 is the location parameter, λ>0 is the scale parameter and η>0, the shape parameter were used for parameter estimation according to Marsaglia and Tsang (2000). This method applied the probability density function:

$$f(x;\eta,\lambda,\varepsilon)=\frac{(\varepsilon/\eta^{\lambda})x^{\lambda-1}e^{-(x/\eta)^{\varepsilon}}}{\Gamma(\lambda/\varepsilon)},\quad x\geq 0 \qquad \textit{Equation 8}$$

to first generate partial sums Z~N(0,1) and U~U(0,1) independently. If the partial sum $Z > -1/c$ and the logarithm of the uniform variate $\log U < \frac{1}{2}Z^2 + d - dV + d\times lnV$, then the random variable wa generated as X=$d\times V$. V is an independent variate describing the times the random variable is generated. In this procedure:

$$V=(1+cZ)^3) \qquad \textit{Equation 9}$$

$$d=\eta-\frac{1}{3} \qquad \textit{Equation 10}$$

$$c=\frac{1}{\sqrt{9d}} \qquad \textit{Equation 11}$$

The probability density function of the generalized 3-parameter Weibull distribution was calculated using the relationship:

$$f(x;\eta,\sigma,\varepsilon)=\frac{\eta}{\sigma}\left(\frac{x-\varepsilon}{\sigma}\right)^{\eta-1}e^{\left(\frac{x-\varepsilon}{\sigma}\right)^{\eta}} \qquad \textit{Equation 12}$$

according to Weibull (1951) for $x \geq \varepsilon$ and $f(x;\eta,\sigma,\varepsilon) = 0$ and for $x < \varepsilon$, where $\eta > 0$ is the shape parameter, $\sigma > 0$ is the scale parameter and ε is the location parameter of the distribution.

Given a random variate U drawn from the uniform distribution in the interval (0, 1), provided a direct variate according to the following relationship:

$$X = \sigma(-\ln(U))^{1/\eta} \qquad \textit{Equation 13}$$

which had a Weibull distribution with parameters η and σ. In generating random numbers belonging to (0,1), zero values were excluded to avoid the undefined natural log of zero.

3.4 Catch and Production

Based on the number of major commercial species, a self propagating feedforward Artificial Neural Network (ANN) based on the outline of Raymond *et al.* (1999), was used to determine the output production layer over a period of time for *R. argentea,* predict future yields and compare these with actual data from Catch Assessment Survey (CAS) data (GoK, 2008) with catch data from 1968 to 2007 (Appendix II). Environmental variables such as average annual temperature, rainfall, river discharge and lake level (Mwirigi *et al.*, 2005) for 1950 to 2005 were obtained but only matching time series data from 1968 to 2005 (with projections for 2006 and 2007) were used as explanatory variables for developing environmental based networks (Appendix II). The following sections provide description of the procedures followed in the analyses.

3.4.1 Analysis

The methods of analysis of catch and environmental data followed the outline in the NeuroDimension (Lafebre *et al*., 2005) and Alyuda Neurointelligence (Alyuda Research, 2005). The datasets that were accepted for the network were partitioned into three sets: the Training set (68.6%), the Validation set (15.7%) and the Test set (15.7%) based on the recommendation of automatic partitioning (Alyuda Research, 2005). This partitioning method was based on the concept that at least 70% of the data provided enough representation for training to identify the specific patterns in the datasets and allowed generalization of the ANN results:

i) The Training set was part of the input dataset used eventually for neural network training, i.e. for adjustment of network weights for maximizing predictive ability and minimizing forecasting error according to Williams and Zipser (1989).

ii) The Validation set was part of the data used to tune network topology or network parameters other than weights. The Validation set was used to calculate generalization loss and retain the best network (the network with the lowest error on Validation set) according to Lafebre and Principe (1992).

iii) The Test set was part of the input data set used only to test how well the neural network would perform on new data. The Test set was used after the network was already trained, to test what errors occurred during the training and that would occur during future network application. This set was not used during training and thus was considered as consisting of new data for the neural network application (Lafebre and Principe, 1992).

3.4.2 Preprocessing and Post-processing

Data preprocessing involved the modification of the original data before input to the Artificial Neural Network. Preprocessing transformed the data by scaling to values between -1 and +1 to make it suitable for neural network. Post-processing means modifying the neural network output to make it understandable by user and/or suitable for computation (Jacobs, 1988).

Numeric columns were automatically scaled during data preprocessing. The numeric values were scaled using the following formulae:

i) $SF = (SR_{max}-SR_{min})/(X_{max}-X_{min})$ *Equation 14*

ii) $Xp = SR_{min} + (X-X_{min})* SF$ *Equation 15*

where:

X = actual value of a numeric column

X_{min} = minimum actual value of the column

X_{max} = maximum actual value of the column

SR_{min} = lower scaling range limit

SR_{max} = upper scaling range limit

SF = scaling factor

Xp = preprocessed value

For both the input and the target columns, scaling range was -1 to +1, based on the Hyperbolic Tangent activation function with a sigmoid curve for catch data (Appendix III) and environmental data (Appendix IV). Data post-processing report briefly produced results about the number of columns and records analyzed as well as about encoded columns. It included columns number before and after preprocessing, column type, scaling range and factors for numeric columns and number of categories for categorical columns.

3.4.3 Network Design

For the network design, Alyuda NeuroIntelligence Ver. 2.2 Software was used, with a built-in search method to determine the network architecture (number of hidden layers and units in each layer) and network properties. This procedure allowed the creation of a feed-forward fully-connected Artificial Neural Network (Multi-Layer Perceptron - MLP). A Heuristic Search (large search range) was used since the problem complexity was not known, according to the procedure of Williams and Zipser (1989) and Werbos (1990). The Artificial Neural Network architecture was then subjected to testing using the Absolute Error (AE) and the best Artificial Neuraal Network selected based on the minimum AE between the training set and the testing set for further training and testing. Absolute Error (Zar, 1984) indicated the quality of the Artificial Neural Network and was calculated by subtracting the observed values (current output) with the predicted values (network output). The Absolute Error was applied as the sum of the squared differences between the actual value (target column value) and neural network output to avoid the errors cancelling out to zero as a standard statistical procedure (Zar, 1984). The number of hidden units found by the Heuristic search was used as a medium point to determine a new and narrower search range, also known as exhaustive search (Alyuda Research, 2005).

3.4.4 Training Networks

The network training (Anderson and Rosenfed, 1990) was monitored and assessed through progress bar, training graphs and training parameters according to the procedures in Alyuda NeuroIntelligence and NeuroDimension (Lafebre *et al*., 2005). The real-time dataset error, correlation and r^2 were used to determine when the optimum network training was achieved. The second monitoring tool involved a plot of the network error or network error improvement against the number of iteration on the training set and involved checking on the:

i) Graphs that plotted both the correlation and r^2 for training and/or validation set.
ii) Dataset errors graph used to plots the average absolute dataset error against the number of iteration on training and/or validation set.

3.4.5 Testing Networks

Network testing was carried out after training completion in which the Actual versus Output Table was produced and a testing summary report. The Actual versus Output Table showed error values for each record from the input dataset, whether for training, testing or validation. The Absolute Error (AE) and Absolute Relative Error (ARE) represented the difference between the actual value of the target column and the corresponding network output. The difference was computed as absolute values and in percentage terms.

3.4.6 Querying and Applying Networks

The trained network was queried with new data using the query mode. New data were entered in columns with names taken from the input data file and used as captions for the query entry cells. The new data was based on weight of input columns determined during the network training. For numerical inputs, minimum and maximum values were used to determine the minimum and maximum values that were presented to the neural networks during training and guided the entry of new values for the queries.

3.4.7 Network Validation and Testing

In order to avoid overfitting, it was necessary monitor the progress of the training and stop training early enough when the minimum AE, RAE and high R^2 were obtaine and when further training was not resulting in better generalization. Finally, the model was validated by testing its ability to to generalize by evaluating its performance on a set of data not used for training, which is assumed to approximate the typical unseen data that a model will encounter (the validation set) and by querrying the network with a set of new data in the range and outside the range of the original datasets. All the queries were tagged to the class benchmark to produce a new table of simulated output independent of the original environment and catch data. The results of the network query were then compared between the environment and catch data for validation.

CHAPTER FOUR

4.0 RESULTS

4.1 Food Type and Food Selection

The Logistic Regression results shows the estimated coefficients (parameter estimates), standard error of the coefficients, z-values, and p-values (Table 7). The positive coefficients, odds ratios higher than 1.000 and $P<0.0005$ indicated for *Thermocyclops oblingatus*, *T. emini* and *Moina macrourus* shows that *R. argentea* is likely to select these species at a higher rate as compared to *Epiphanes* as the size increases. For *T. oblingatus*, the odds ratio is 10 times for 50 mm as compared to 10 mm fish whereas for *M. macrourus*, the odds ratio is 11 times for 50 mm as compared to fish of 10 mm. The 95% confidence interval for the odds ratio provided the level of uncertainty that could be attached to the odds ratio in 95 percent of the times based on the data used in the analysis.

The coefficient associated with the length sizes (predictors) is the estimated change in the logit with a one unit change in the predictor (length size), assuming that all other factors and covariates are constant. Only logit 3, 8 and 11 showed significant changes in the odds ratio.

Table 7: **Logistic regression table showing logits with p-values less than 0.05 marked with asterix for groups of length classes (Source: Author)**

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Logit 1: (*Keratella serrulata* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.154151	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	<0.0000005	0.0922740	<0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.1126110	<0.05	1.000	1.00	0.80	1.25
Logit 2: (*B. angularis* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.121361	0.0859610	-1.41	0.158	0.89	0.75	1.05
30 mm	<-0.000005	0.0922740	<-0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.1126110	<0.05	1.000	1.00	0.80	1.25
Logit 3: (*T. oblingatus* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	0.0555699	0.0844260	0.66	0.510	1.06	0.90	1.25
30 mm*	0.628609	0.0868462	7.24	<0.005	1.88	1.58	2.22
40 mm*	1.33318	0.0807868	16.50	<0.005	3.79	3.24	4.44
50 mm*	2.32728	0.0944510	24.64	<0.005	10.25	8.52	12.33
Logit 4: (*B. calyciflorus* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.154151	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	<0.0000005	0.0922740	<0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.1126110	<0.05	1.000	1.00	0.80	1.25
Logit 5: (*B. falcatus* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.154151	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	<-0.000005	0.0922740	<-0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.112611	<0.05	1.000	1.00	0.80	1.25
Logit 6: (*Brachionus caudatus* / *Epiphanes sp*)							
Constant	-0.0000000	0.0659380	<-0.05	1.000			
GROUP							
20 mm	-0.154151	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	0.0408220	0.0918213	0.44	0.657	1.04	0.87	1.25
40 mm	<0.0000005	0.112611	<0.05	1.000	1.00	0.80	1.19
50 mm	<0.0000005	0.112611	<0.05	1.000	1.00	0.80	1.25

Table 7 (contd): Logistic regression table showing significant logits having p-values less than 0.05 marked with asterix for groups of size classes (Source: Author)

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Logit 7: (*Bosmina longirostris* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.1541510	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	0.1177830	0.0910111	1.29	0.196	1.13	0.94	1.34
40 mm	0.0339016	0.0879695	0.39	0.700	1.03	0.87	1.23
50 mm	<-0.000005	0.1126110	<-0.05	1.000	1.00	0.80	1.25
Logit 8: (*Moina macrourus* / *Epiphanes sp*)							
Constant	<0.0000005	0.0659380	<0.05	1.000			
Group:							
20 mm	-0.154151	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm*	0.377294	0.0886551	4.26	<0.005	1.46	1.23	1.74
40 mm*	1.257680	0.0810072	15.53	<0.005	3.52	3.00	4.12
50 mm*	2.435070	0.0942309	25.84	<0.005	11.42	9.49	13.73
Logit 9: (*Ceriodaphnia riguadi* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
GROUP							
20 mm	-0.1541510	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	<0.0000005	0.0922740	<0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.1126110	<0.05	1.000	1.00	0.80	1.25
Logit 10: (*D. lumholtzi* / *Epiphanes sp*)							
Constant	<-0.000005	0.0659380	<-0.05	1.000			
Group:							
20 mm	-0.1541510	0.0862732	-1.79	0.074	0.86	0.72	1.02
30 mm	<0.0000005	0.0922740	<0.05	1.000	1.00	0.83	1.20
40 mm	<0.0000005	0.0882955	<0.05	1.000	1.00	0.84	1.19
50 mm	<0.0000005	0.112611	<0.05	1.000	1.00	0.80	1.25
Logit 11: (*Thermocyclops emini* / *Epiphanes sp*)							
Constant	<0.0000005	0.0659380	<0.05	1.000			
Group:							
20 mm	-0.1213610	0.0859610	-1.41	0.158	0.89	0.75	1.05
30 mm*	0.3772940	0.0886551	4.26	<0.005	1.46	1.23	1.74
40 mm	0.0339016	0.0879695	0.39	0.700	1.03	0.87	1.23
50 mm	<-0.0000050	0.1126110	<-0.05	1.000	1.00	0.80	1.25

Logit 3 between *Thermocyclops oblingatus* and *Epiphanes sp*:

i) Table 7 shows a p-values of <0.0005 for length-class of 30 mm, indicating that there is sufficient evidence to conclude that a change in length-class from 10mm to 30mm affected the quantity of *T. oblingatus* as compared to *Epiphanes* in the stomach contents of *R. argentea*. The positive coefficient of 0.629 and an odds ratio of 1.88 indicates that *T. oblingatus* is likely to occur 1.88 times more than *Epiphanes* in in the stomach contents as the size-class changes from 10 mm to 30 mm.

ii) Table 7 also shows a p-values of <0.0005 for length-class of 40 mm, indicating that there is sufficient evidence to conclude that a change in length-class from 10 mm to 40 mm affected the quantity of *T. oblingatus* as compared to *Epiphanes* in the stomach contents of *R. argentea*. The positive coefficient of 1.333 and an odds ratio of 3.79 indicates that *T. oblingatus* is likely to occur 3.79 times more than *Epiphanes* in the stomach contents as the size-class changes from 10 mm to 40mm.

For Logit 8 between *Moina macrourus* and *Epiphanes sp*:

i) The p-values of <0.0005 for length-class of 30 mm (Table 7), indicate that there is sufficient evidence to conclude that a change in length-class from 10 mm to 30 mm affected the quantity of *M.macrourus* as compared to *Epiphanes* in the stomach contents of *R. argentea*. The positive coefficient of 0.377 and an odds ratio of 1.46 indicates that *M. macrourus* is likely to occur 1.46 times more than *Epiphanes* in the stomach contents as the size-class changes from 10 mm to 30 mm.

ii) Table 7 shows a p-values of <0.0005 for length-class of 40 mm, indicating that there is sufficient evidence to conclude that a change in length-class from 10 mm to 40 mm affected the quantity of *M. macrourus* as compared to *Epiphanes* in the stomach contents of *R. argentea*. The positive coefficient of 1.258 and an odds ratio of 3.52 indicate that *M. macrourus* is likely to occur 3.52 times more than *Epiphanes* in the stomach contents as the size-class changes from 10 mm to 40 mm.

iii) The p-values of <0.0005 for length-class of 50 mm (Table 7), further indicate that there is sufficient evidence to conclude that a change in length-class from 10 mm 30 mm, 10mm to 20 mm and 10 mm to 50 mm affected the quantity of *M. macrourus* as compared to *Epiphanes* in in the stomach contents of *R. argentea*. The positive coefficient of 2.435 and an odds ratio of 11.42 indicate that *M. macrourus* is likely to occur 11.42 times more than *Epiphanes* in in the stomach contents as the size-class changes from 10 mm to 50 mm.

For Logit 8 between *T. emini* and *Epiphanes sp*:

i) The p-values of <0.0005 for length-class of 30 mm (Table 7), indicate that there is sufficient evidence to conclude that a change in length-class from 10 mm to 30 mm affected the quantity of *T. emini* as compared to *Epiphanes* in the stomach contents of *R. argentea*. The positive coefficient of 0.377 and an odds ratio of 1.46 indicates that *M. macrourus* is likely to occur 1.46 times more than *Epiphanes* in the stomach contents as the size-class changes from 10 mm to 30 mm.

The Chi-square analysis between pairs of the zooplankton species (Table 8) showed significant differences in the stomach contents among pairs of zooplankton for logit 3, 7, 8 and 11 with *Epiphanes* as the reference. The quantity of zooplankton species observed in the stomach of *R. argentea* was heterogeneous with respect to fish size for *T oblingatus* ($\chi^2_{0.05,4}$=992.89; p<0.0005), *B. longirostris* ($\chi^2_{0.05,4}$=11.46; p=0.022), *M. Macrourus* ($\chi^2_{0.05,4}$=1175.41; p<0.0005) and *T. emini* (χ^2=40.94; p<0.0005) as compared to *Epiphanes*.

Table 8: Chi-square (χ^2) tests for terms for all the Logits with more than 1 degree of freedom and showing the significant pairs with asterix (Source: Author)

Term	(χ^2)	DF	p-value
Logit 1: (*Keratella serrulata* / *Epiphanes sp*)	5.62	4	0.230
Logit 2: (*B. angularis* / *Epiphanes sp*)	3.53	4	0.474
Logit 3: (*T. oblingatus* / *Epiphanes sp*)*	992.89	4	<0.005
Logit 4: (*B. calyciflorus* / *Epiphanes sp*)	5.62	4	0.230
Logit 5: (*B. falcatus* / *Epiphanes sp*)	5.62	4	0.230
Logit 6: (*Brachionus caudatus* / *Epiphanes sp*)	6.77	4	0.148
Logit 7: (*Bosmina longirostris* / *Epiphanes sp*)*	11.46	4	0.022
Logit 8: (*Moina macrourus* / *Epiphanes sp*)*	1175.41	4	<0.005
Logit 9: (*Ceriodaphnia riguadi* / *Epiphanes sp*)	5.62	4	0.230
Logit 10: (*D. lumholtzi* / *Epiphanes sp*)	5.62	4	0.230
Logit 11: (*Thermocyclops emini* / *Epiphanes sp*)*	40.94	4	<0.005

The log-likelihood ratio, calculated as G-statistic was 6220.024 (DF = 44; p-value <0.0005), indicating that all the slopes are not zero. The G-statistic was used to test the null hypothesis that all the coefficients associated with predictors equal zero versus them not all being zero where predictors are the different size classes used in the logistic regression. The log-likelihood ratio is applicable to goodness-of-fit analysis in circumstances having data for which chi-square may be employed. The log-likelihood ratio, considers the ratio between likelihoods or probabilities of two food items. Twice the log-likelihood is the G-statistic.

The calculated probabilities of food items in the environment (R_{i2}) and in the stomach of *R. argentea* were highest for *P. oblingatus* and *M. macrourus* by a magnitude of 10s to 1000s as compared to all other species. The proportion of the food items in the enviroment (P_i) were higher for *T. emini* (0.6478), *T. oblingatus* (0.1884) and *Daphnia lumholtzii* (0.1149) in the order of 100s to >1000s as comparededed to others species (Table 9).

Table 9: Semi processed data for BBN based on stomach content species (R_{i2}) and the environment (P_i) (Source: Author)

Food items	Stomach	Environment	R_{i2}	P_{i2}
Thermocyclops emini	2.64	29.31	0.0079	0.6478
T. oblingatus	173.56	8.53	0.5158	0.1884
Daphnia lumholtzi	0.08	5.20	0.0002	0.1149
Moina macrourus	151.17	0.00	0.4493	0.0000
Bosmina longirostris	4.45	0.00	0.0132	0.0000
Brachionus caudatus	1.30	1.74	0.0039	0.0384
B. falcatus	0.12	0.00	0.0003	0.0000
B. calyciflorus	0.12	0.41	0.0003	0.0091
B. angularis	0.53	0.00	0.0016	0.0000
Keratella serrulata	0.37	0.00	0.0011	0.0000
K. quadrata	0.00	0.07	0.0000	0.0015
Trichocerca sp.	0.11	0.00	0.0003	0.0000
Epiphanes sp.	1.99	0.00	0.0059	0.0000
K. cochlearis	0.02	0.00	0.0001	0.0000

The BBN developed from the prior probalility of the food items in the stomach and in the environment provided the posterior probabilities of each food in the stomach (Fig. 1) based on joint/marginal probabilities of the food items in both the stomach and the environment.

Figure 1 provided the posterior probabilities in percentage of getting each food item in the stomach of *R. argentea* such as *Bosmina longirostris* (67%) in the stomach on condition that it occurs in the environment 80% of the times and also in the stomach 90% of the time (joint probability) but also on condition that some other food items also occur in both the environment and the stomach with given probabilities. The relationship is represented by:

$$P(H|E,c) = \frac{P(H|c) \cdot P(E|H,c)}{P(E|c)}$$

Equation 16

where we can update our belief in hypothesis H; probability of *B. longirostris* in the stomach) given its prior probability in the environment (evidence E) and the background information c (its proportion among the cladocera in the environment). The left-hand term, P(H|E,c) is the "posterior probability," or the probability of H after considering the effect of E given c. The term P(H|c) is called the "prior probability" of H given c alone. The term P(E|H,c) is the "likelihood" and gives the probability of the evidence assuming the hypothesis H and the background information c is true. Finally, the last term P(E|c) is called the "expectedness", or how expected the evidence is given only c. It is independent of H and can be regarded as a marginalizing or scaling factor. All these probabilities are conditional and they specify the degree of belief in propositions based on the assumption that some other propositions are true. This procedure applies to each and every food item in the network.

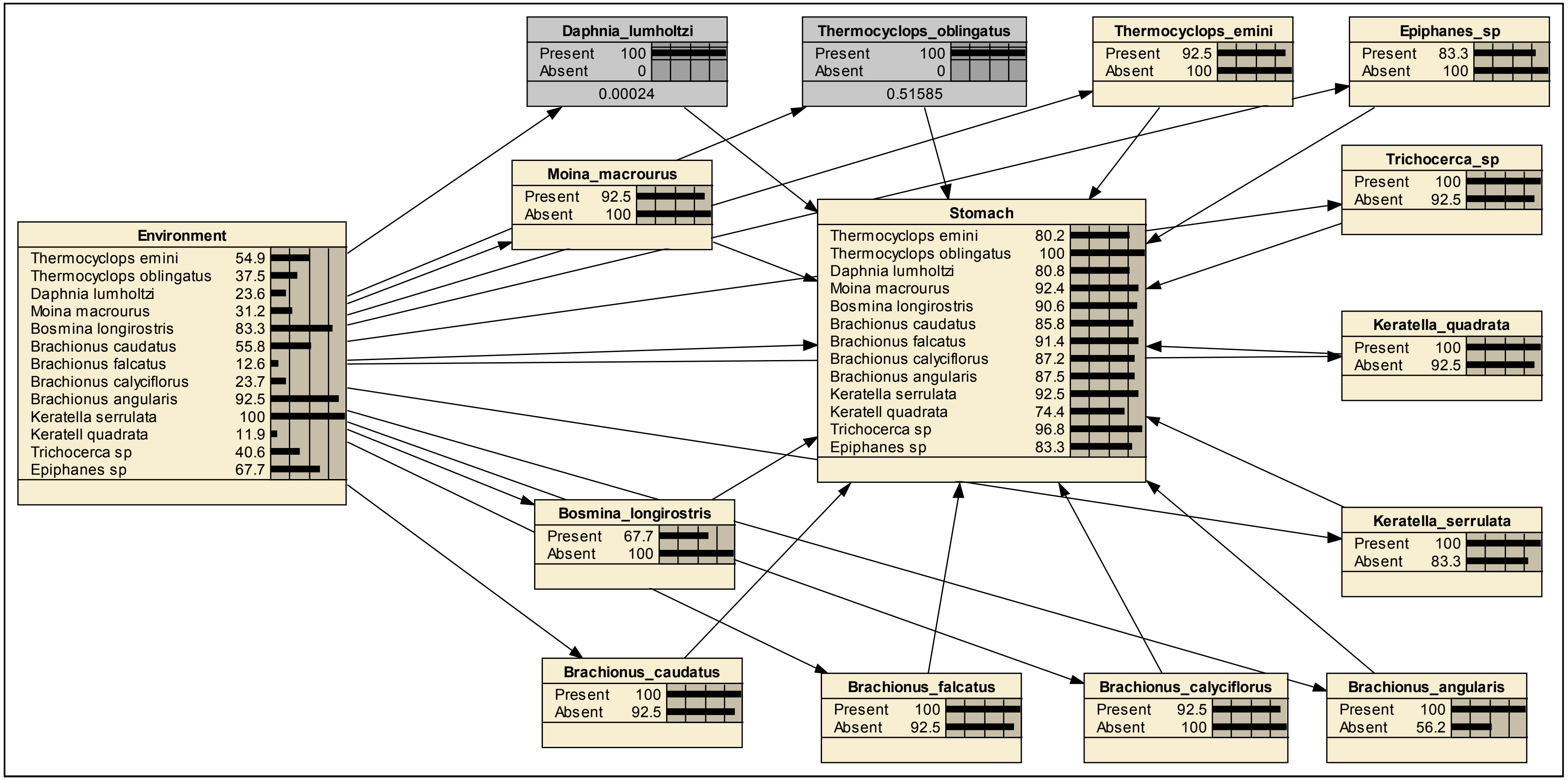

Figure 1: **Bayesian Belief Network (BBN) for food items in the environment and in the stomach (Source: Author)**

Network sensitivity of the findings was analyzed to determine how much the beliefs and the mean value of the target node could be influenced by a single finding at each of the other nodes in the network (each called a "findings node"). The results show how much the findings node can affect the target node using different sensitivity measures (Table 10) for stomach content and the prior probabilities. The Minimum and maximum real are the lowest and highest probabilities of the food item at a given node (0.0759 – 0.0791), thereby indicating minimal changes in the model

For the BBN, the percentage variance reduction of 3.14×10^{-3}, is the expected reduction in variance of the probability of any food item due to finding in other node. This is the square of RMS change, a small measure of variance in the belief that the feeding model is stable at all nodes.

The belief variance of 4.9×10^{-4}, is the probability obtained under the assumption that a posterior distribution on the uncertainty of any food item in the stomach can be approximated by its prior distribution. The belief variance percentage is only 0.0491% thereby strengthening the belief that the network is stable.

The percentage entropy reduction of 1.38% is the sum of the products of the posterior probabilities and the logarithm of the error value on each expected food item in the stomach. High entropy indicate disorder (unstable) feeding model while low entropy indicate a stable feeding model.

Table 10: Sensitivity of environment to findings in stomach contents (Source: Author)

Species	Min	Current	Max	RMS Change
Thermocyclops emini	0.07233	0.07633	0.08421	0.00233
Thermocyclops oblingatus	0.07470	0.07704	0.08273	0.00165
Daphnia lumholtzi	0.07277	0.07662	0.08069	0.00258
Moina macrourus	0.07369	0.07783	0.08078	0.00157
Bosmina longirostris	0.07079	0.07723	0.08145	0.00217
Brachionus caudatus	0.07075	0.07606	0.08148	0.00221
Brachionus falcatus	0.07444	0.07703	0.08289	0.00217
Brachionus calyciflorus	0.07164	0.07734	0.07923	0.00142
Brachionus angularis	0.07512	0.07713	0.08107	0.00155
Keratella serrulata	0.07161	0.07707	0.08314	0.00268
Keratell quadrata	0.07257	0.07652	0.07979	0.00174
Trichocerca spp.	0.07363	0.07681	0.08076	0.00202
Epiphanes spp.	0.07219	0.07700	0.08148	0.00207
Mean of Real Value:	**0.07590000**	**0.07738000**	**0.07907000**	**0.00097520**
Variance reduction% =	**0.00000095**	**0.00314000**		
Entropy reduction % =	**0.00051230**	**0.01380000**		
Belief Variance % =	**0.00000419**	**0.00049100**		

The highest Root Mean Square (RMS) (0.00268 and 0.00258) was recorded for *Keratella serrulata* and *D. lumholtzii* respectively. The RMS, also known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity from the base level. The base level in the network is considered to be no change in prey abundance in the stomach regardless of the density of prey in the environment. The RMS results indicate that these two species were more sensitive to changes of abundance in the environment. Lower RMS changes (0.00142, 0.00155 and 0.00157) were recorded for *Brachionus calyciflorus, B. angularis* and *M. macrourus* thereby indicating that these species were less sensitive to the changes in abundance in the environment. The higher the RMS, the more sensitive is the prey species found in the stomach to the density in the environment.

4.2 Fecundity and Recruitment of *R. argentea*

The estimate of spawning stock biomass and egg production (Fig. 2) show that both have a maximum at a length of 46 mm with data from different source (length frequency distribution (LVFO, 2005), sex ratio of 2 females: 1 male (Okedi, 1973; Wandera, 1992), fecundity (Manyala *et al*., 1992) and length-weight relationship (Manyala *et al*., 1995b).

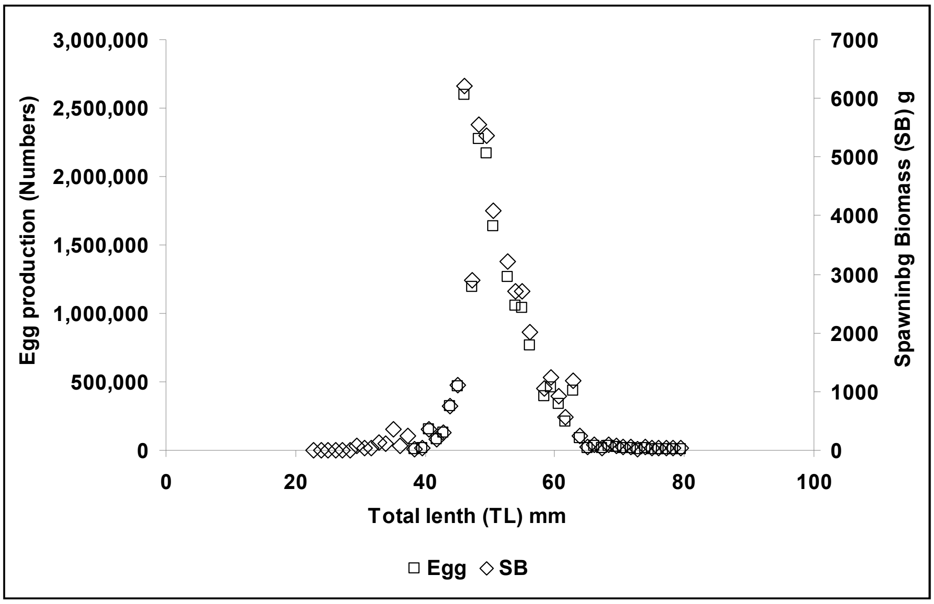

Figure 2: Spawning stock biomass (SB) and egg production as a function of sizes of *R. argentea* in Lake Victoria based on length frequency distribution (LVFO, 2005), sex ratio of 2 females: 1 male (Okedi, 1973; Wandera, 1992), fecundity (Manyala *et al*., 1992) and length-weight relationship (Manyala *et al*., 1995b)

Both the SSB and egg production start increasing just below 40 mm. The maximum egg production SSB is realized between 40 mm and 60 mm. This size range is also where most of the fish are found (Appendix I and Fig. 2). Due to this heavy pressure, local overfishing has been reported in some parts of Winam Gulf (Manyala *et al*., 1995; Manyala and Ojuok, 2007) but the species has survived probably due to its r-selecion strategy, high P/B ratio and high reproductive potential.

The cumulative spawning biomass provided "biomass indices" for modeling the actual egg production by size against the cumulative spawning of all sizes reaching sexual maturity. Based on these indices, the relationship between the spawning biomass and the egg production (Fig. 3) showed a polynomial distribution of order 4:

$$\text{Egg Production}=2.2^{-11}\text{SSB}^4 - 1.1^{-6}\text{SSB}^3 + 0.02\text{SSB}^2 + 35.7\text{SSB}\ (R^2 = 0.959).$$

Equation 17

Equation 16 has a zero intercept, indicating that if there is no SSB, then there is no recruitment. The error term is also not included since the data do not represent a population. A polynomial distribution in the context of stock-recruitment implies that there are a number of approximately normal distribution and rescaling the polynomial distribution provides the exact distribution of the random variates being generated. Combining the descriptive statistics and the polynomial stock-recruitment relationship resulted in the determination of the scaling parameters (skewness and kurtosis).

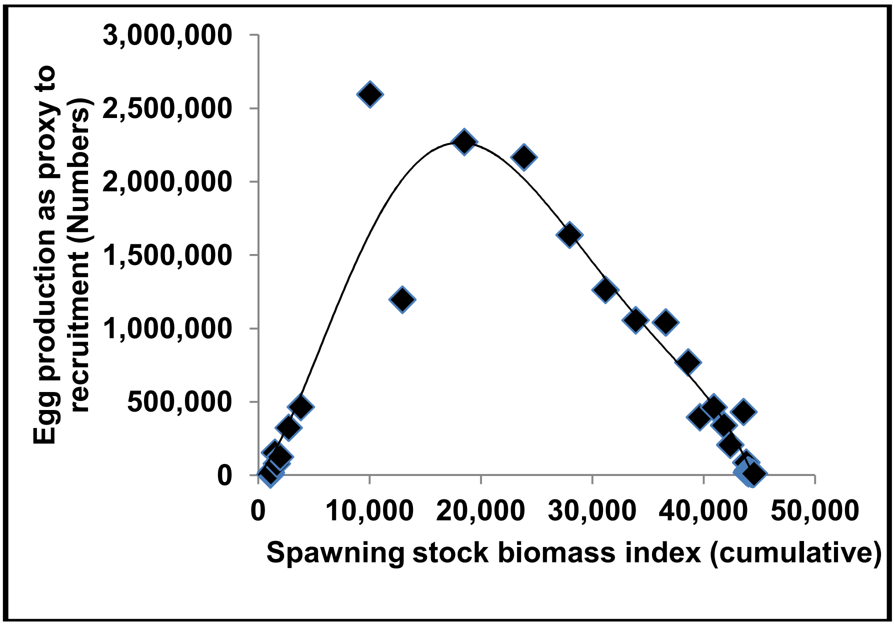

Figure 3: Relationship between spawning biomass and egg production of *R. argentea* in Lake Victoria (Source: Author)

Descriptive statistics of the egg production (Table 11) showed that the egg production data was not normally distributed, hence demanding for further analysis of egg production as a proxy to recruitment. The results indicate a skewness of 1.2 and a kurtosis of 5.6.

Table 11: Descriptive statistics of egg frequency distribution of *R. argentea* (Source: Author)

Variable	Eggs
n	2,419
Mean	992,139
SE Mean	17,070
StDev	839,551
Skewness	1.20
Kurtosis	5.6

The estimated minimum egg production generated was a minimum of 179,239 and maximum of 3,590,778 for Gamma distribution and a minimum of 198,030 and maximum of 3,966,603 for the Weibull distribution (Fig. 4) following the procedures outlined in the methodology according to Marsaglia and Tsang (2000) and Weibull (1951) respectively.

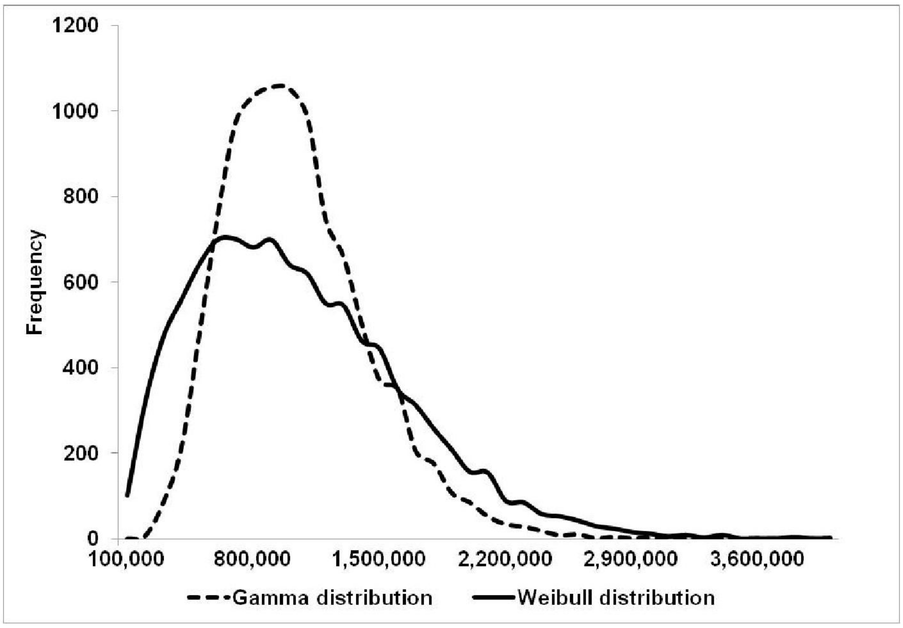

Figure 4: Frequency of egg frequency based simulation of 10,000 data points for Gamma and Weibull distributions (Source: Author)

The random variables generated showed a positive skewness for both Gamma and the Weibull distribution with heavy tails at the beginning and thin tails at the end (Table 12). The maximum frequency observed was between 700,000 and 1,100,000 million eggs for the Gamma distribution and between 600,000 and 900,000 for the Weibull distribution. Whereas the Gamma distribution was terminating at about 3,600,000 million eggs, the Weibull terminated at 4,000,000 egss occurring at extremely low frequencies.

Table 12: Random variables generated for egg production for Gamma and Weibull distributions (Source: Author)

Bin	Gamma distribution	Weibull distribution
100,000	0	101
200,000	7	320
300,000	85	472
400,000	213	559
500,000	478	642
600,000	751	699
700,000	973	701
800,000	1035	681
900,000	1056	699
1,000,000	1052	641
1,100,000	985	619
1,200,000	748	550
1,300,000	659	545
1,400,000	504	463
1,500,000	370	444
1,600,000	349	349
1,700,000	205	312
1,800,000	176	257
1,900,000	107	208
2,000,000	85	156
2,100,000	52	154
2,200,000	34	88
2,300,000	28	84
2,400,000	18	57
2,500,000	8	52
2,600,000	10	42
2,700,000	1	29
2,800,000	4	23
2,900,000	1	15
3,000,000	3	11
3,100,000	1	5
3,200,000	1	8
3,300,000	0	2
3,400,000	0	8
3,500,000	0	0
3,600,000	1	0
3,700,000	0	0
3,800,000	0	3
3,900,000	0	0
4,000,000	0	1

Based on the range of estimates of the maximum egg production and maximum frequency, both the Weibull and Gamma distributions adequately fitted the data and the variations reflect difference mostly in the shape parameter. The total egg production estimates were 10,381,900,000 and 10,503,200,000 for Gamma and Weibull respectively, thereby providing a good fit both the distributions.

The random variables so generated were subject to the Skewness-Kurtosis All Test to determine their consistency and generate the three distribution parameters that would explain recruitment (Table 13). The Skewness-Kurtosis all Test yielded p-values of 0.0001 for both the Gamma and Weibull distribution, indicating that they were significantly different from normal distribution.

Table 13: Egg data distribution analysis based on 10,000 generated datasets for each distribution and the location, scale and shape parameters (Source: Author)

Raw Data	Gamma	Weibull
Mean	992,139	992,139
SD	591,000	713,500
Skewness	1.2	1.2
Excess Kurtosis	2.16	1.85
Location parameter (ε)	7,139	7,057
Scale parameter (λ,σ)	354,600	1,080,678
Shape parameter (η,η)	2.78	1.4
Test Results		
Mean	986,272	997,367
SD	584,605	688,364
Skewness	1.14	1.23
Excess Kurtosis	1.78	20.8
p-value (Skewness-Kurtosis all Test)	0.0001	0.0001

4.3 Catch and Production

4.3.1 Network Data Processing

The network data processing for catch data and class benchmark categories produced 13 columns. The 13 columns represent the 8 species (including *R. argentea*) used in the analysis and the 5 categories (pristine, growth, dominance, decline and collapse) used in the class benchmarks. All the 13 columns were scaled during the processing and the outputs for each record of class benchmaks was either 1 for positive category or -1 for negative category. The 8 columns were also scaled during processing and all the values produced were between -1 and 1 using the Hyperbolic Tangent function (Appendix IV)

The network data processing for environmental data and class benchmark categories produced 11 columns. The 11 columns represent the 5 environmental variables (discharge (cusecs), rainfall (mm), evaporation (mm), outflow (cusecs) and lake level (m.a.s.l.)) and *R. argentea* as the target used in the analysis and the 5 categories (pristine, growth, dominance, decline and collapse) used in the class benchmarks. All the 11 columns were scaled during the processing and the outputs for each record of class benchmaks was either 1 for positive category or -1 for negative category. The 6 environmental columns and the target species column (*R. argentea*) were also scaled during processing and all the values produced were between -1 and 1 using the Hyperbolic Tangent function (Appendix IV).

4.3.2 Artificial Neural Network Design

The best Neural Network Architecture for the catches of *R. argentea* was 10-9-1 based on environmental variable and using Heuristic search method with Hyperbolic Tangent activation function for both input and output (Fig. 5). The output parameters were catches of *R. argentea* with sum-of squares error function. Fitness criteria used to determine the best network was established for nine hidden layers [10-9-1] to be 0.000194.

The neural network input layer consisted of 5 categorical variables (pristine, growth, dominance, decline and collapse) and 5 numerical variables (discharge, rainfall, evaporation, outflow, level). There were 9 hidden layers and 1 output layer which represent the numerical target catch of *R. argentea* (Fig. 5). This combination of neurons produced the 10-9-1 architecture design.

The best Neural Network Architecture for the catches of *R. argentea* was 12-6-1 based on fish catch data and using Heuristic search method with Hyperbolic Tangent activation function for both input and output (Fig. 6). The output parameters were catches of *R. argentea* with sum-of squares error function. Fitness criteria used to determine the best network for six hidden layers [12-6-1] was 0.000214.

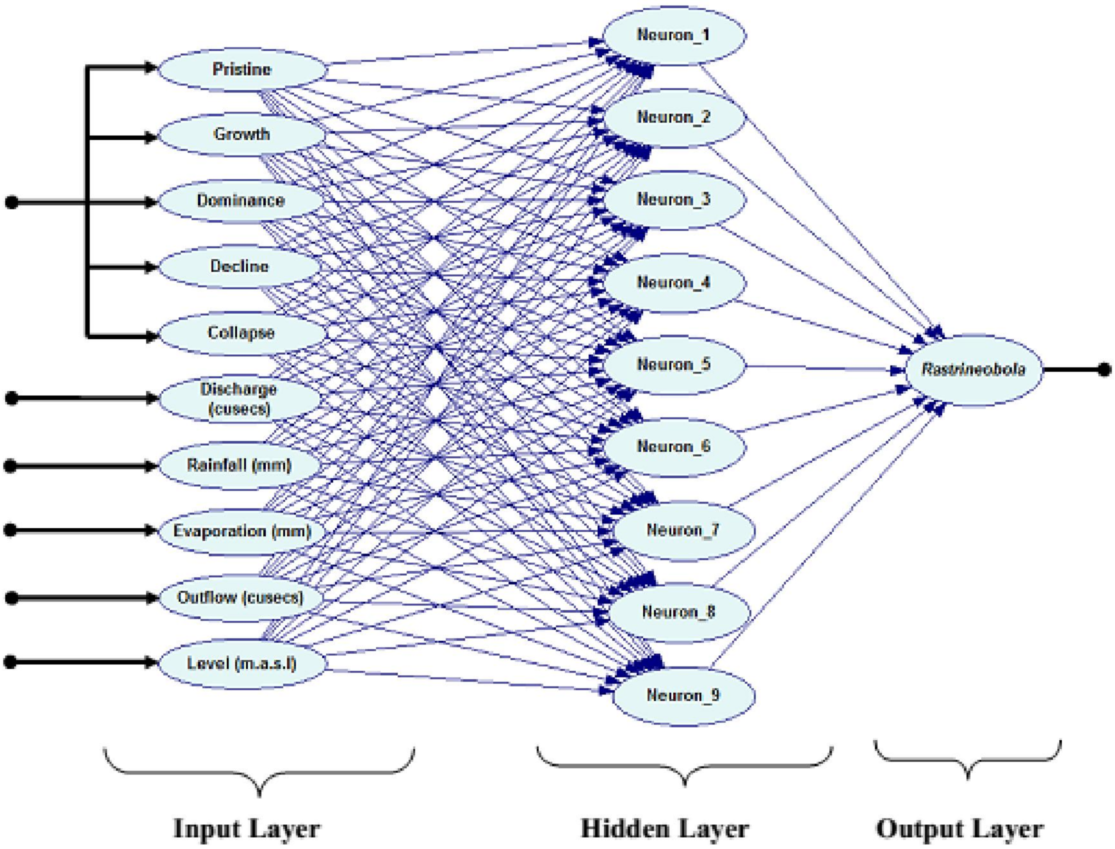

Figure 5: **A 10-9-1 Artificial Neural Network architecture for production of *R. argentea* based on environmental variables (Source: Author)**

The neural network input layer consisted of 5 categorical variables (pristine, growth, dominance, decline and collapse) and 7 numerical variables (*Bagrus, Clarias, Haplochromis, Lates, Mormyrus, Protopterus, Oreochromis*) . There were 6 hidden layers and 1 output layer which represented numerical target catch of *R. argentea* (Fig. 6). This combination of neurons produced the 12-6-1 architecture design.

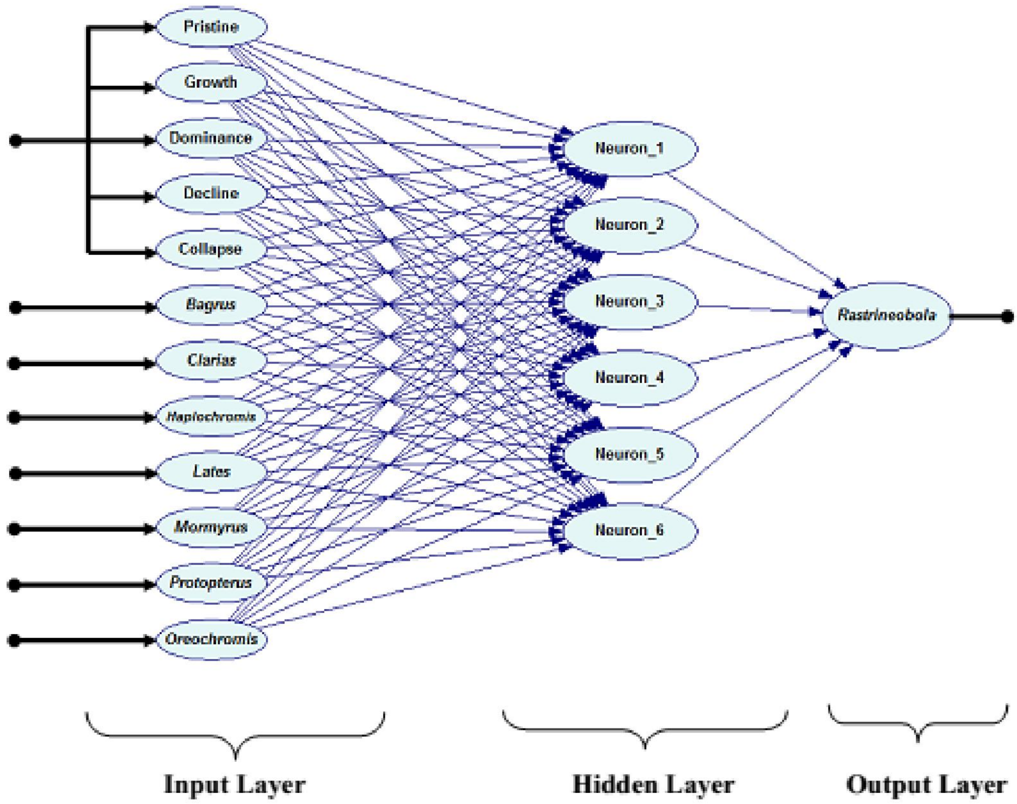

Figure 6: **A 12-6-1 Artificial Neural Network architecture for production of *R. argentea* based on fisheries variables (Source: Author)**

4.3.3 Artificial Neural Network Training

A maximum of 50 iterations were carried out on the best 10-9-1 network architecture based on environmental variables and using the Quasi-Newton Training algorithm. The results indicate that the maximum r-squared error was obtained after 5 iterations for the training set error and the validation set error. Only 5 iterations gave the least error difference (0.05) between the training set (0.9) and the validation set (0.85) (Fig. 7). The high r-squared indicated that the training set can expain 90% of the observed target and 85% of the validation can also explain the observed target catches of *R. argentea* in Lake Victoria using environmental data (Fig. 8). The minimum dataset error was also obtained after 5 iterations for the training set (5,000) and the validation set (5,000). The 5 iterations gave the least error difference (zero) between the training set and the validation set (Fig. 8). The training set, testing set and validation sets were all part of the original data set in order to avoid poor predictions.

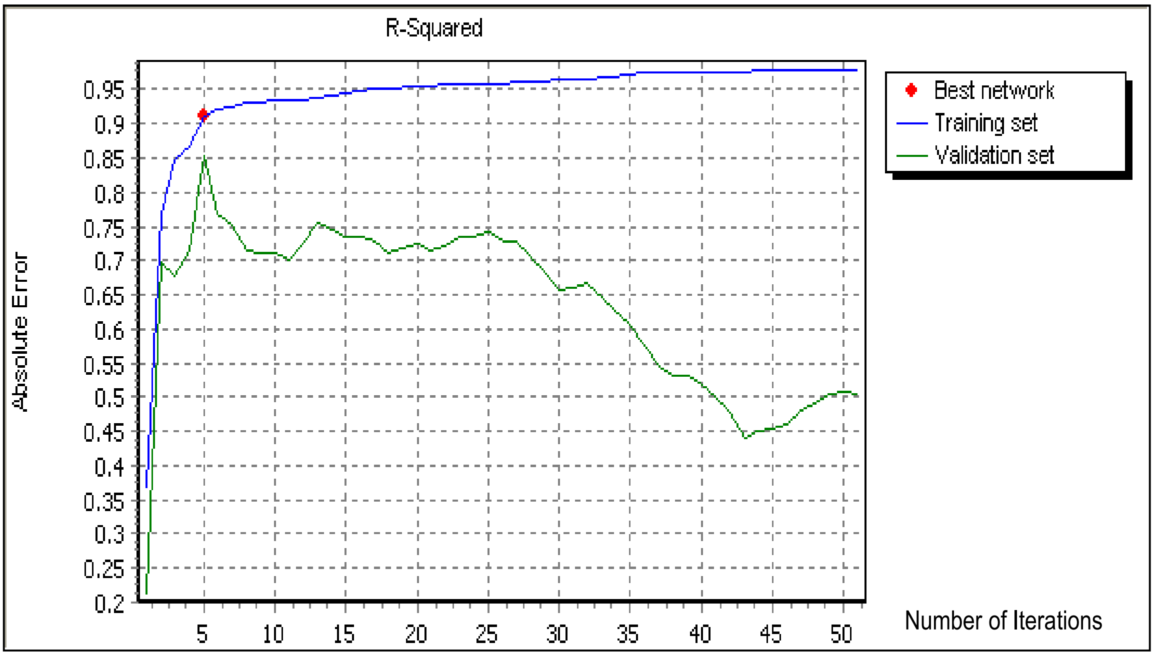

Figure 7: R-squared errors analysis for training and validation sets based on environmental variables after the network training (Source: Author)

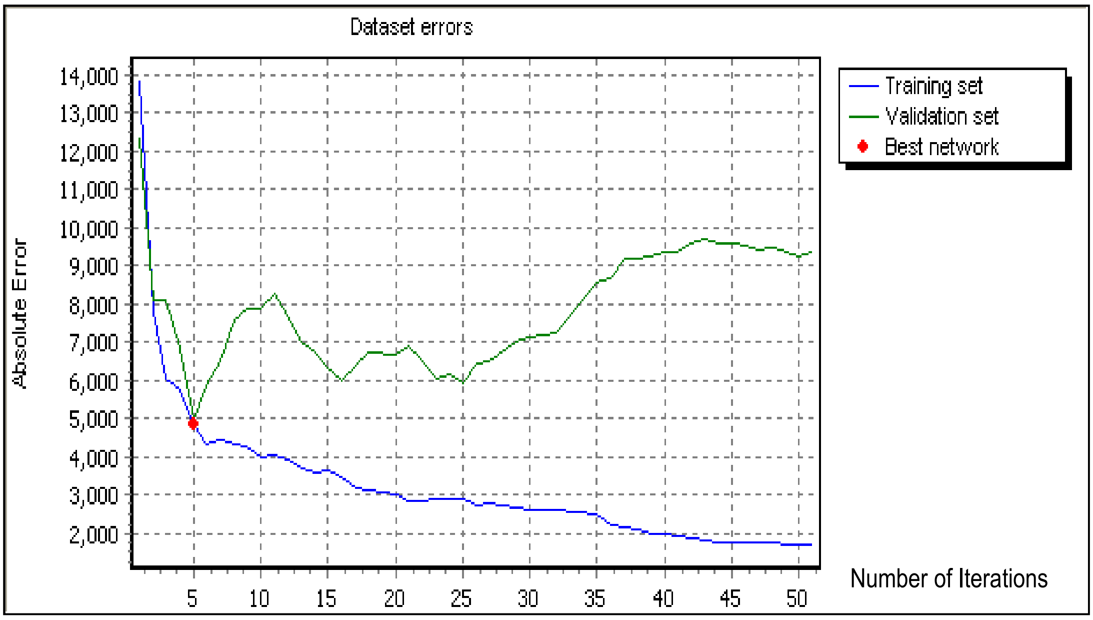

Figure 8: **Dataset Errors analysis for training and validation sets, based on environmental variables after the network training (Source: Author)**

The jackknife procedure for the input importance showed that classification of the fisheries development (39.7%), lake level (39.4%) and evaporation (11.7%) constituted the most important environmental variables in determining catches of *R. argentea* in Lake Victoria (Fig. 9). The input importance indicates that the class benchmark, lake level and evaporation account for 90.8% of the variations observed in the catches of *R. argentea* in Lake Victoria. In a typical jackknife applications, an empirical sampling distribution is generated by deleting a single data point (Efron and Gong, 1983), that is, by sampling n - 1 of the original observations. In general, sampling subsets that leave out one, two, or a whole group of observations, and then defining a distribution across such deletions, provides an empirical distribution based on the jackknife approach.

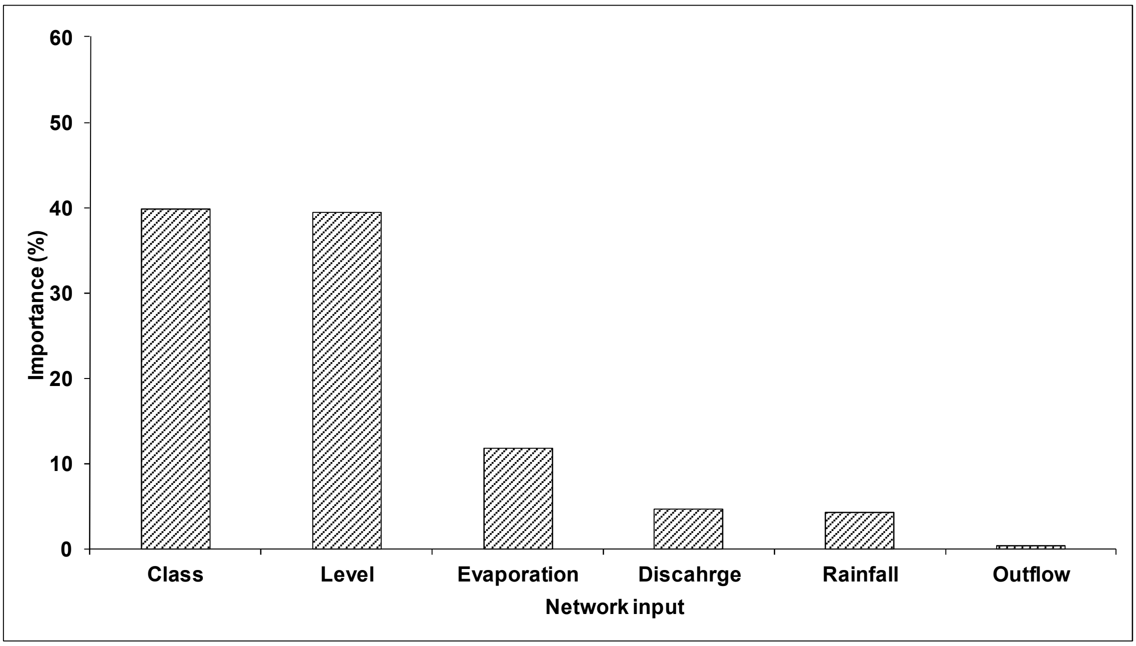

Figure 9: Importance of environmental variable on the network architecture for *R. argentea* catches in Lake Victoria (Source: Author)

A maximum of 50 iterations were carried out on the best 12-6-1 network architecture based on fisheries catch data and using the Quasi-Newton Training algorithm and the results indicate that the maximum r-squared error was obtained after 6 iterations for the training set and the validation set. Only 6 iterations gave the least error difference (0.05) between the training set (0.93) and the validation set (0.98) (Fig. 10). The high r-squared has indicated that the training set can expain 93% of the observed target and 98% of the validation can also explin the observed target catches of *R. argentea* in Lake Victoria using fisheries catch data. The minimum dataset error was obtained after 6 iterations for the training set (4,000) and the validation set (2,500) for the best network (Fig. 11).

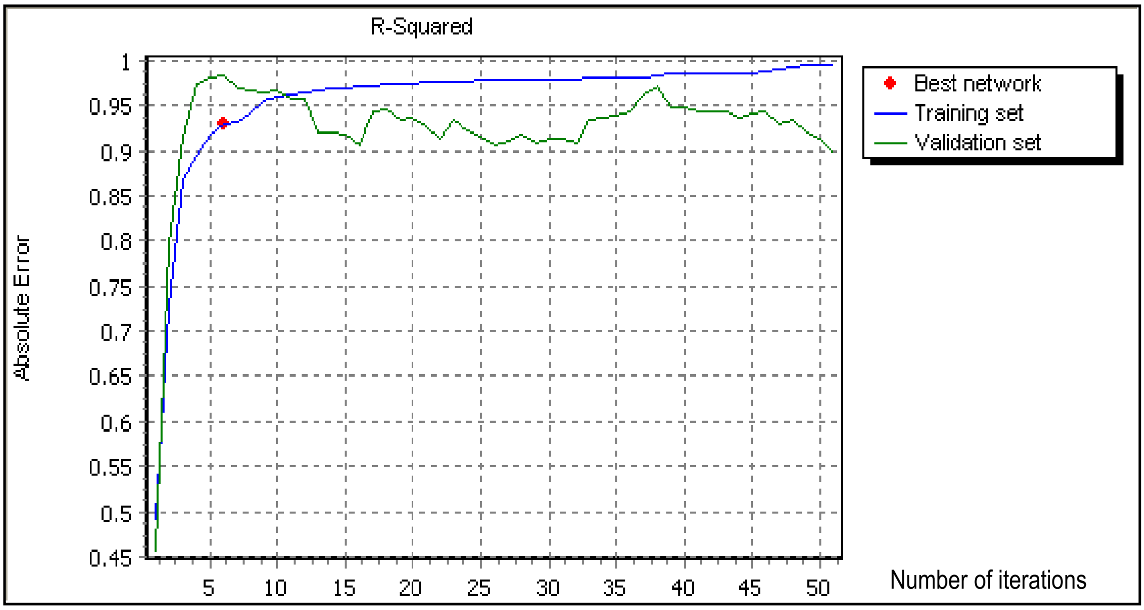

Figure 10: R-squared errors analysis for training and validation sets based on fisheries data after the network training (Source: Author)

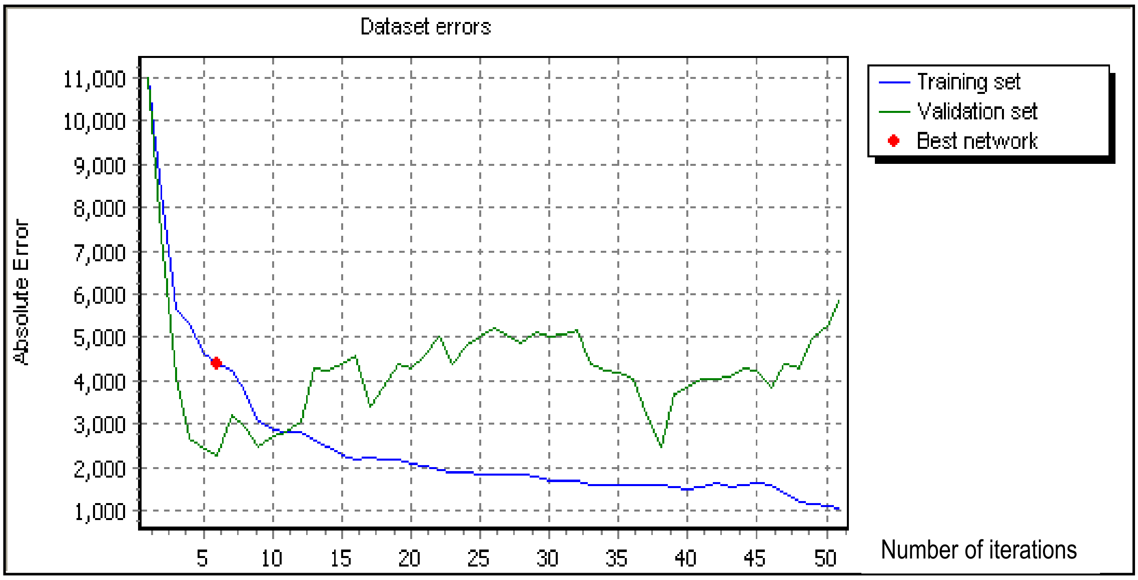

Figure 11: Dataset Errors analysis for training and validation sets, based on fisheries data after the network training (Source: Author)

The input importance analysis for fish catch statistics (Fig. 12) showed that *Lates niloticus* contributed (23%), classification of the fishery development (50%) and *Haplochromis* (17%) constituted the most important fisheries variables in determining catches of *R. argentea* in Lake Victoria. The class benchmark, *L. niloticus* and *Haplochromis* can explain 90% of the variation in catches of *R. argentea* in Lake Victoria.

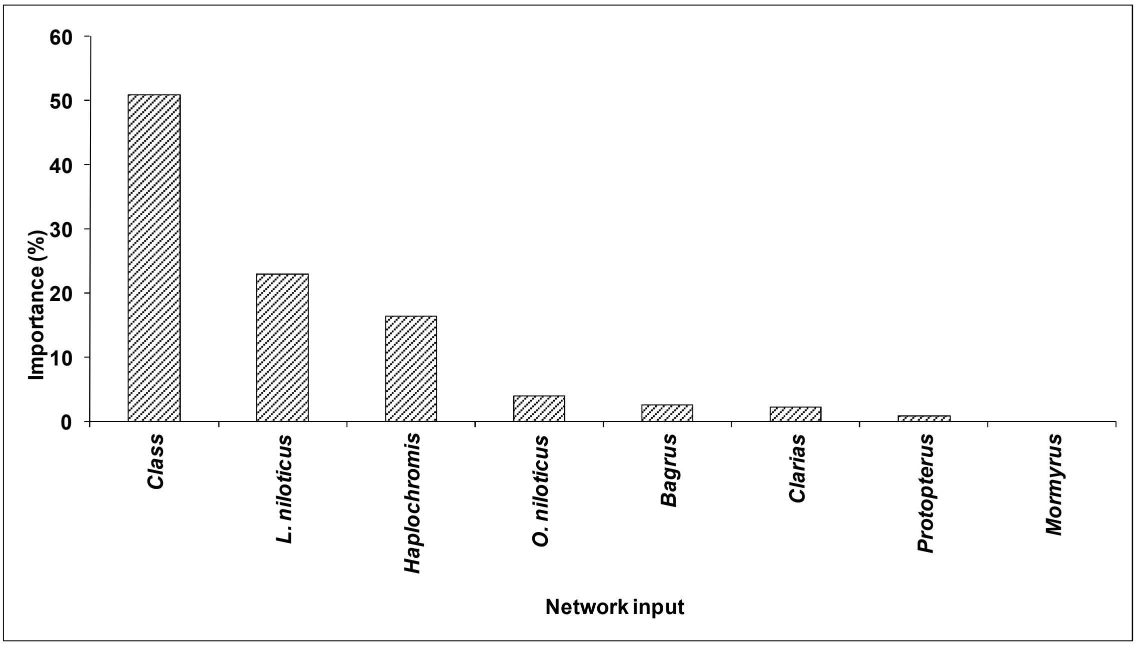

Figure 12: Importance of fisheries data on the network architecture for *R. argentea* catches in Lake Victoria (Source: Author)

4.3.4 Artificial Neural Network Testing

The training results produced an actual versus output table (Table 14), containing the following features:

i) Input column values selected for the x-axis in the actual versus output graphs.

ii) Input Columns from the dataset.

iii) Target - the target value that is taken from the input data file.

iv) Output - the output produced by the network for each record, i.e. the target value produced by the network.

v) AE and ARE - absolute error and absolute relative error %. Difference between the actual value of the target column and the corresponding network output.

The absolute relative error (ARE) between the target and the output was from 0.9% to 55% for the high catch values (9,321 to 69,134 tonnes) and 117% for the lower value (1,768 tonnes) for the testing set. For the training set, the ARE was between 0.4% to 77.1% with one high value of 246.7% while for the validation set, the ARE was 0.6% and 59.9%.

The actual versus output graphs were plotted on real time scale using the Serial Numbers (SN) as time index for environmental data. The graph so produced (Fig. 13) shows the target (observed) catch and the output (predicted) catch based on the network class benchmark and environmental variables; lake level, evaporation, discharge, rainfall and outflow.

Table 14: Real time actual versus output table for *R. argentea* catches in Lake Victoria (Kenya), showing the absolute error (AR) and the absolute relative error (ARE) for each estimate based on environmental data (Source: Author)

SN	Set	Class	Target	Output	AE	ARE
1	Testing	Growth	1,768	3844	2076	117.4%
2	Testing	Growth	9,321	4219	5102	54.7%
3	Testing	Decline	69,134	45674	23460	33.9%
4	Testing	Decline	38,968	36185	2783	7.1%
5	Testing	Decline	35,414	35724	310	0.9%
6	Training	Growth	1,255	4351	3096	246.7%
7	Training	Growth	9,443	2161	7282	77.1%
8	Training	Dominance	7,635	13250	5615	73.5%
9	Training	Collapse	35,455	45433	9978	28.1%
10	Training	Growth	5,652	4200	1452	25.7%
11	Training	Pristine	731	553	178	24.4%
12	Training	Collapse	54,019	45433	8586	15.9%
13	Training	Decline	40,168	46526	6358	15.8%
14	Training	Collapse	57,929	50614	7315	12.6%
15	Training	Dominance	40,861	45949	5088	12.5%
16	Training	Decline	58,098	51109	6989	12.0%
17	Training	Growth	8,710	9594	884	10.1%
18	Training	Dominance	16,444	14953	1491	9.1%
19	Training	Dominance	25,866	23662	2204	8.5%
20	Training	Collapse	49,472	53241	3769	7.6%
21	Training	Growth	5,448	5034	414	7.6%
22	Training	Collapse	49,165	52666	3501	7.1%
23	Training	Decline	42,505	44003	1498	3.5%
24	Training	Pristine	520	537	17	3.3%
25	Training	Growth	6,704	6917	213	3.2%
26	Training	Decline	40,318	41405	1087	2.7%
27	Training	Dominance	45,464	44641	823	1.8%
28	Training	Pristine	524	533	9	1.7%
29	Training	Dominance	33,145	33699	554	1.7%
30	Training	Decline	42,336	42524	188	0.4%
31	Training	Decline	49,670	49684	14	0.0%
32	Validation	Collapse	31,659	50614	18955	59.9%
33	Validation	Growth	3,742	4187	445	11.9%
34	Validation	Decline	56,827	50569	6258	11.0%
35	Validation	Dominance	34,518	34282	236	0.7%
36	Validation	Dominance	19,437	19328	109	0.6%

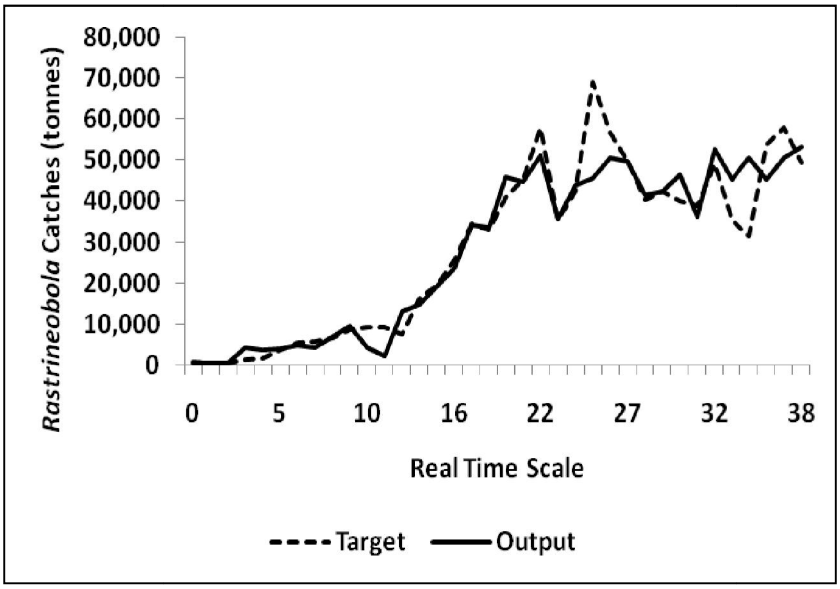

Figure 13: Actual versus output plot produced by network testing plotted on a real time scale using observation number of environmental data (Source: Author)

All the environmental variables gave the same trend in actual versus output graphs when plotted in real time. The use of rainfall data in testing the network gave close similarity between the observed catch (target) and the ANN predicted or output catch (Fig. 14). Results of the testing of rainfall time series in the range of 1,320 to 2,226 mm per annum resulted in catches ranging from 535 to 53,241 metric tones per annum as compared to observed catches ranging from 731 to 69,134 metric tones per annum.

When the target and output columns were plotted against detrended rainfall data, there was a cyclic trend of catches with increasing rainfall (Fig. 15). The low or high catch values shown in the plot were not necessarity related to low or high rainfall but were the results of the network training. The recognized pattern of catch at different annual rainfall values resulted from a comination of all the other environmental variables in addition to rainfall data.

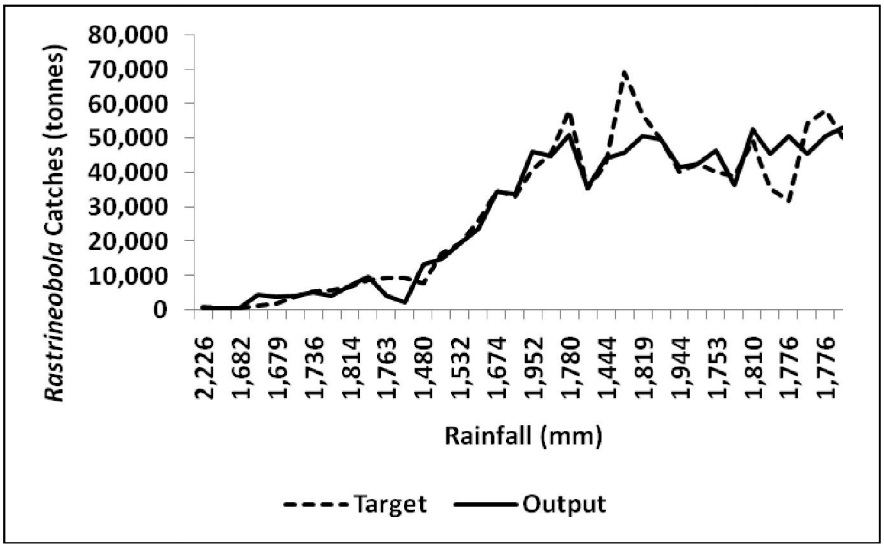

Figure 14: Actual versus output plot produced by network testing plotted on a real time scale using rainfall data (Source: Author)

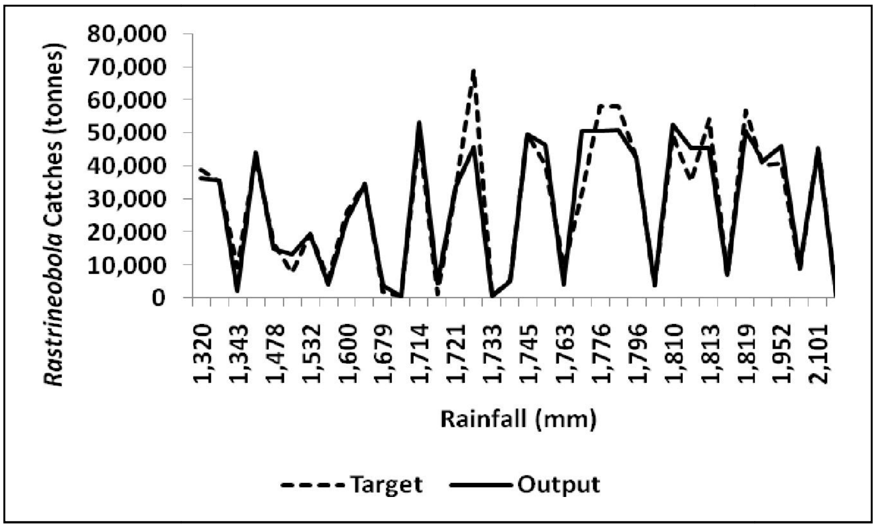

Figure 15: Actual versus output plot produced by network testing plotted on a non real time scale using the rainfall data (Source: Author)

Table 15 is an actual versus output table based on catch data estimated ARE between 6.2% and 43% for high values (8,710 to 69,134 tonnes) and upto 150.5% for one low value (1,255 tonnes). ARE for training set was between 0.2% and 92.5% while that for the validation set was between 9.6% and 56.3%.

Table 15: Real time actual versus output table for *R. argentea* catches in Lake Victoria (Kenya), showing the absolute error (AR) and the absolute relative error (ARE) for each estimate based on catch data (Source: Author)

SN	Set	Class	Target	Output	AE	ARE
1	Test	Growth	1255	3144	1889	150.5%
2	Test	Growth	8710	12452	3742	43.0%
3	Test	Decline	69134	54876	14258	20.6%
4	Test	Dominance	34518	37918	3400	9.8%
5	Test	Collapse	57929	61525	3596	6.2%
6	Training	Growth	1768	3404	1636	92.5%
7	Training	Dominance	7635	10115	2480	32.5%
8	Training	Growth	3742	2753	989	26.4%
9	Training	Pristine	520	634	114	21.9%
10	Training	Pristine	524	633	109	20.8%
11	Training	Growth	5652	6443	791	14.0%
12	Training	Growth	9443	8209	1234	13.1%
13	Training	Pristine	731	643	88	12.0%
14	Training	Dominance	16444	14768	1676	10.2%
15	Training	Dominance	25866	23335	2531	9.8%
16	Training	Dominance	19437	20596	1159	6.0%
17	Training	Decline	49670	51520	1850	3.7%
18	Training	Decline	56827	55175	1652	2.9%
19	Training	Decline	42336	43537	1201	2.8%
20	Training	Decline	40318	39319	999	2.5%
21	Training	Dominance	33145	33897	752	2.3%
22	Training	Dominance	40861	41428	567	1.4%
23	Training	Decline	38968	38462	506	1.3%
24	Training	Collapse	35455	35860	405	1.1%
25	Training	Decline	35414	35013	401	1.1%
26	Training	Growth	9321	9421	100	1.1%
27	Training	Collapse	54019	53604	415	0.8%
28	Training	Collapse	49472	49214	258	0.5%
29	Training	Collapse	31659	31497	162	0.5%
30	Training	Decline	40168	40342	174	0.4%
31	Training	Dominance	45464	45535	71	0.2%
32	Validation	Growth	5448	2381	3067	56.3%
33	Validation	Growth	6704	4165	2539	37.9%
34	Validation	Decline	42505	49862	7357	17.3%
35	Validation	Decline	58098	50398	7700	13.3%
36	Validation	Collapse	49165	44432	4733	9.6%

The actual versus output graphs were plotted on real time scale using the Serial Numbers (SN) as time index for fisheries based data. The graph so produced (Fig. 16) shows the target (observed) catch and the output (predicted) catch based on the network and fisheries catch data on the major species in the fishery since 1970 such as *Bagrus, Clarias, Haplochromis, Lates, Mormyrus, Protopterus* and *Oreochromis*. The fisheries catch data showed the same trend in actual versus output graphs when plotted in real time (Fig. 17). The network produced similar output regardless of the species used in the plot after training.

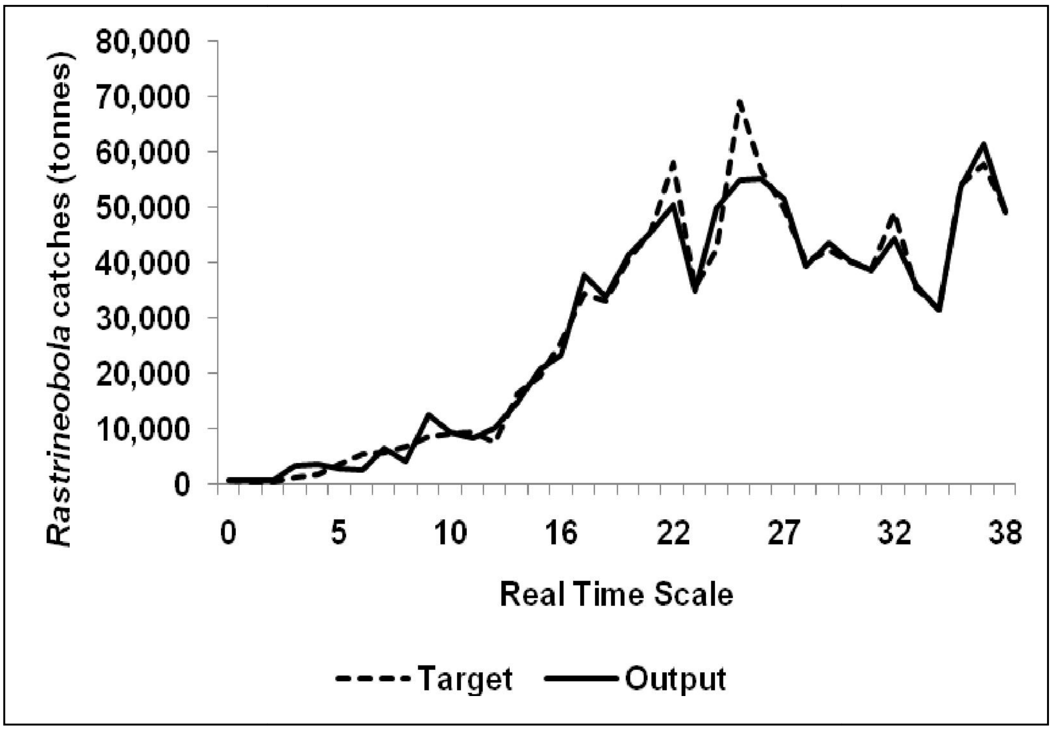

Figure 16: **Actual versus output plot produced by network testing plotted on a real time scale using the observation number based of fisheries data (Source: Author)**

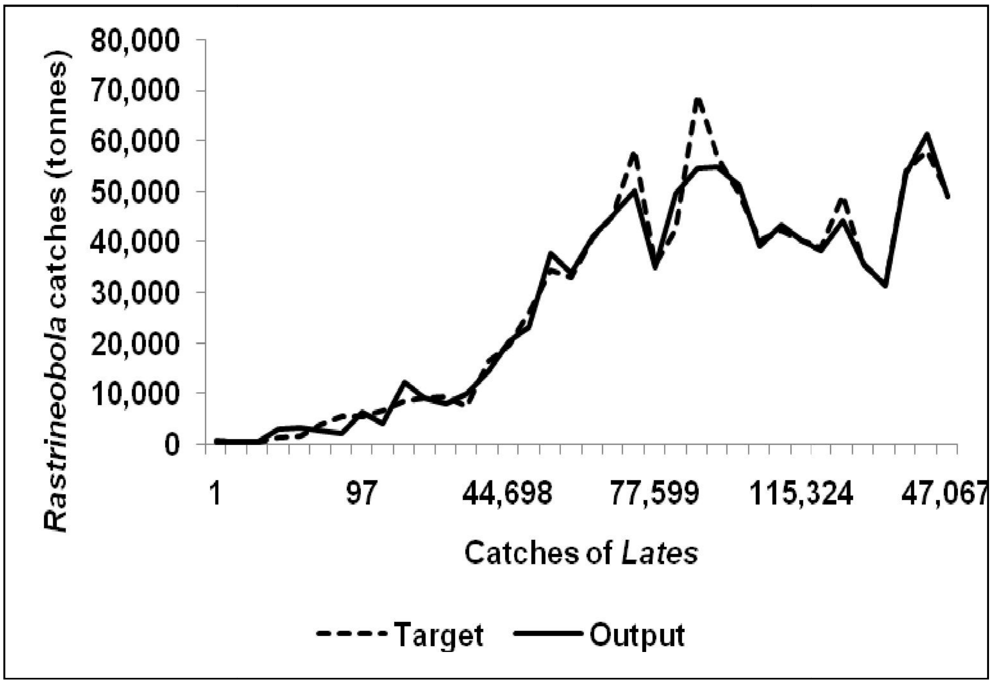

Figure 17: Actual versus output plot produced by network testing plotted on a real time scale using *Lates* data (Source: Author)

When the target and output columns were plotted against detrended catches of *L. niloticus*, there was a generally increasing trend of catches with increasing catches of *L. niloticus* (Fig. 18). The plot showed a possible linear relationship with increasing slope but instability between 30,000 and 58,000 catches of *L. niloticus*.

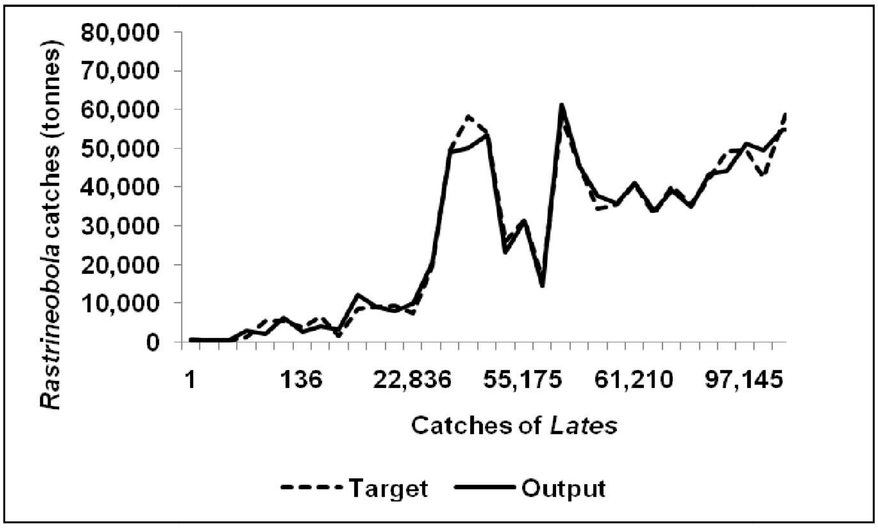

Figure 18: Actual versus output plot produced by network testing plotted on a non real time scale using the *Lates* data (Source: Author)

4.3.5 Querying and Applying the Network

Querying the trained network was carried out using new data to produce outputs based on the best network architecture and for the classification variables: pristine, growth, dominance, decline and collapse.

The network query was based on a record of values for all input columns. For the environmental data; class, rainfall and lake levels were used since they had the most significant influence on the network. The results of the network query for both environmental variables (Table 17) and fisheries data (Table 18) can be recalculated for any value combination of input columns.

For the two queries (Table 17 and 18), the datasets used were initially based on the test dataset and the columns with highest contribution to the networks, which were varied to generate the prediction of output indicated in the last column of both the tables. In both the queries, the catches of *R. argentea* were minimal during the pristine period and rose to maximum values when the *Lates* fishery was collapsing. For the environmental variable, the network predicts a catch of 1,196 tonnes to 2,583 tonnes in the pre *Lates* period and from 47,674 tonnes to 51,279 tonnes when *Lates* collapsed. For the fisheries based network, the output during pre *Lates* period was 805 tonnes to 809 tonnes while during the *Lates* collapse, the predicted catches were 65,249 tonnes to 65,783 tonnes.

Table 16: Network query output table for *R. argentea* based on new simulated data and on environmental variables for the various classifications (Source: Author)

Class	Level	Evaporation	Discahrge	Rainfall	Outflow	*Rastrineobola*
Growth	1136	1543	967	1763	1503	3340
Growth	1135	1543	967	1800	1503	4263
Growth	1135	1543	967	1900	1503	6348
Growth	1135	1543	967	2000	1503	9703
Growth	1135	1543	967	2100	1503	14823
Growth	1134	1543	967	2200	1503	25849
Decline	1134	1479	927	1722	823	53845
Decline	1135	1479	927	1800	823	51714
Decline	1135	1479	927	1900	823	51207
Decline	1135	1479	927	2000	823	50697
Decline	1135	1479	927	2100	823	50188
Decline	1136	1479	927	2200	823	48164
Collapse	1135	1481	772	1810	1096	50488
Collapse	1135	1481	772	1900	1096	49985
Collapse	1135	1481	772	2000	1096	51279
Collapse	1135	1481	772	2100	1096	49490
Collapse	1136	1481	772	2200	1096	47674
Pristine	1136	1525	1,213	2226	1384	1196
Pristine	1135	1525	1,213	2000	1384	1287
Pristine	1135	1525	1,213	2200	1384	2583
Dominance	1135	1500	982	1952	1037	42471
Dominance	1135	1500	982	2100	1037	43116
Dominance	1135	1500	982	2200	1037	38790

Table 17: Network query output table for *R. argentea* based on new simulated data and on fisheries data for the various classifications (Source: Author)

Class	*Bagrus*	*Clarias*	*Haplochromis*	*Lates*	*Mormyrus*	*Protopterus*	*Oreochromis*	*Rastrineobola*
Pristine	1147	1756	3743	1	53	2808	20	805
Pristine	1147	1756	3743	10	53	2808	20	806
Pristine	1147	1756	4000	25	53	2808	20	809
Growth	856	2729	4700	40	80	1900	160	3739
Growth	856	2729	3000	60	80	1900	160	3492
Growth	856	2729	2500	200	80	1900	250	3486
Growth	856	2729	2500	800	80	1900	250	3519
Growth	856	2729	2500	1,000	80	1900	250	3530
Dominance	62	1500	3	1,200	140	600	1000	3778
Dominance	62	1500	5000	10,000	140	600	1000	9021
Decline	62	1500	4100	50,000	140	600	1000	36770
Decline	62	1500	4100	70,000	140	600	1000	45151
Decline	62	1500	4100	85,000	140	600	1000	50672
Collapse	88	4000	5000	103,000	160	202	18000	65783
Collapse	88	4000	5100	100,000	160	202	18000	65719
Collapse	88	4000	5100	88,000	160	202	18000	65249
Collapse	88	4000	5100	95,000	160	202	18000	65536

4.3.6 Comparison of Environmental and Fisheries Networks

The two sets of networks: based on environmental variables and on fisheries data were compared first by plotting both the outputs on the same x-axis real time scale (Fig. 19). Both the ouputs could be described by polynomial equations of the third order:

i) Environmental variables: Output = $-2.986x^3 + 147.9x^2 - 4.571x$ *Equation 18*

ii) Fisheries data: Output = $-2.465x^3 + 130.6x^2 + 269.4x$ *Equation 19*

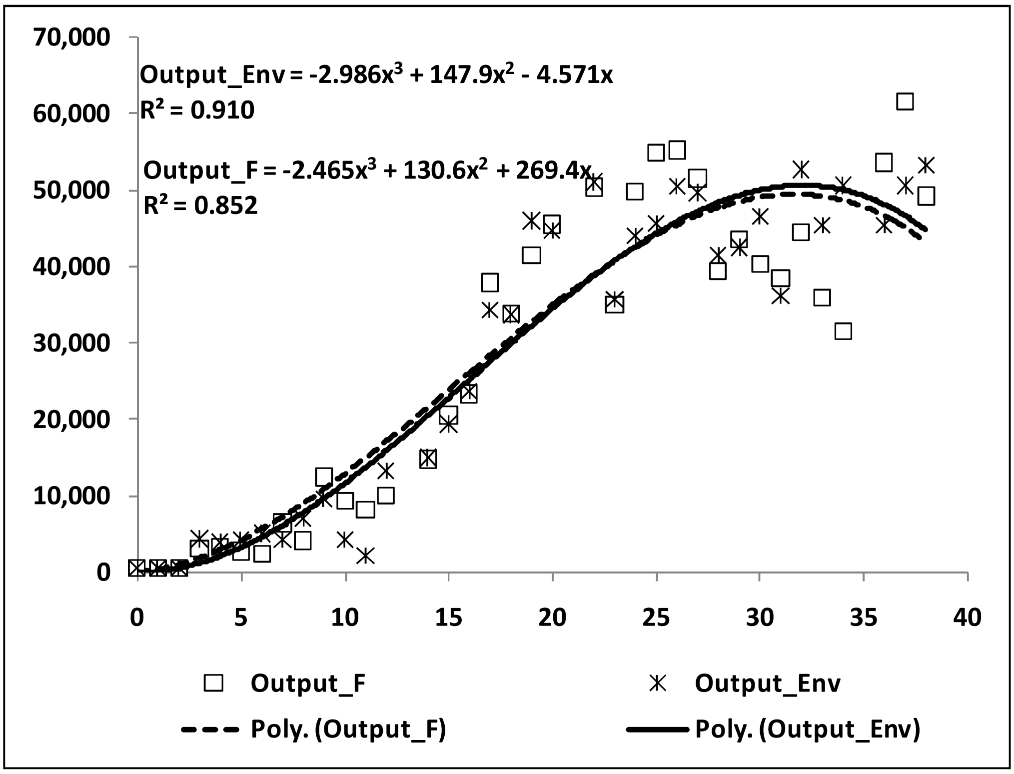

Figure 19: Real time output of catches of *R. argentea* in Lake Victoria (Kenya) based on environmental variables network and fisheries data network (Source: Author)

Since there was no objective way of comparing such polynomial equations, the absolute values of each pair of outputs calculated from environmental variable and fisheries data were analyzed using a linear approach on non real time data. Individual regression of network output on target gave the following linear relationship for the environmental variables: Output = 1057 + 0.952•Target (Fig. 20). The results of the regression coefficient (t=34.71; $p<0.005$, R^2 = 97.3%) and the regression line ($F_{(0.5,\ 1,34)}$=1204; $p<0.0005$) were both significant.

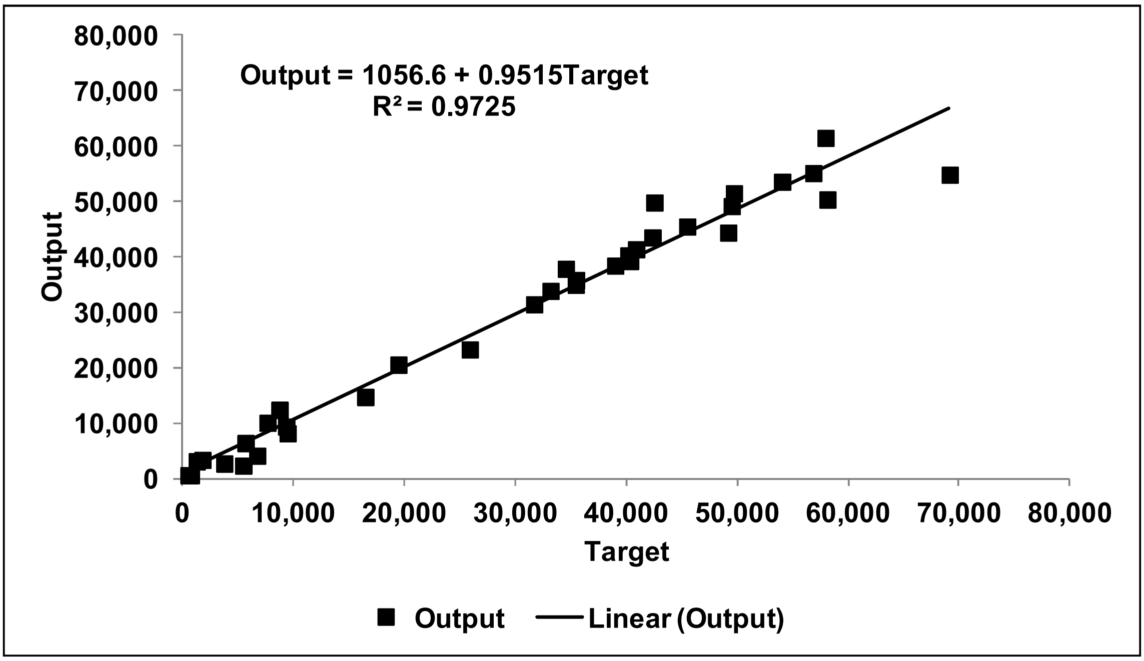

Figure 20: Comparison of network output and target of *R. argentea* catches from Lake Victoria based on environmental variables (Source: Author)

Individual regression of network output on target gave the following linear relationship for the fisheries data: Output = 1,056.6 + 0.9515•Target *Equation 20*

Regression of network output from fisheries data on network output from environmental variables gave the following linear relationship:

$$Output_{\text{Fisheries data}} = 884 + 0.968 \bullet Output_{\text{Environmental variables}}$$ *Equation 21*

The results show that the regression coefficient are statistically significant (t=20.68; $p<0.005$, R^2 = 92.6%). The regression line was also significant ($F_{(0.5,\ 1,34)}$=427.46; $p<0.0005$) (Fig. 21)

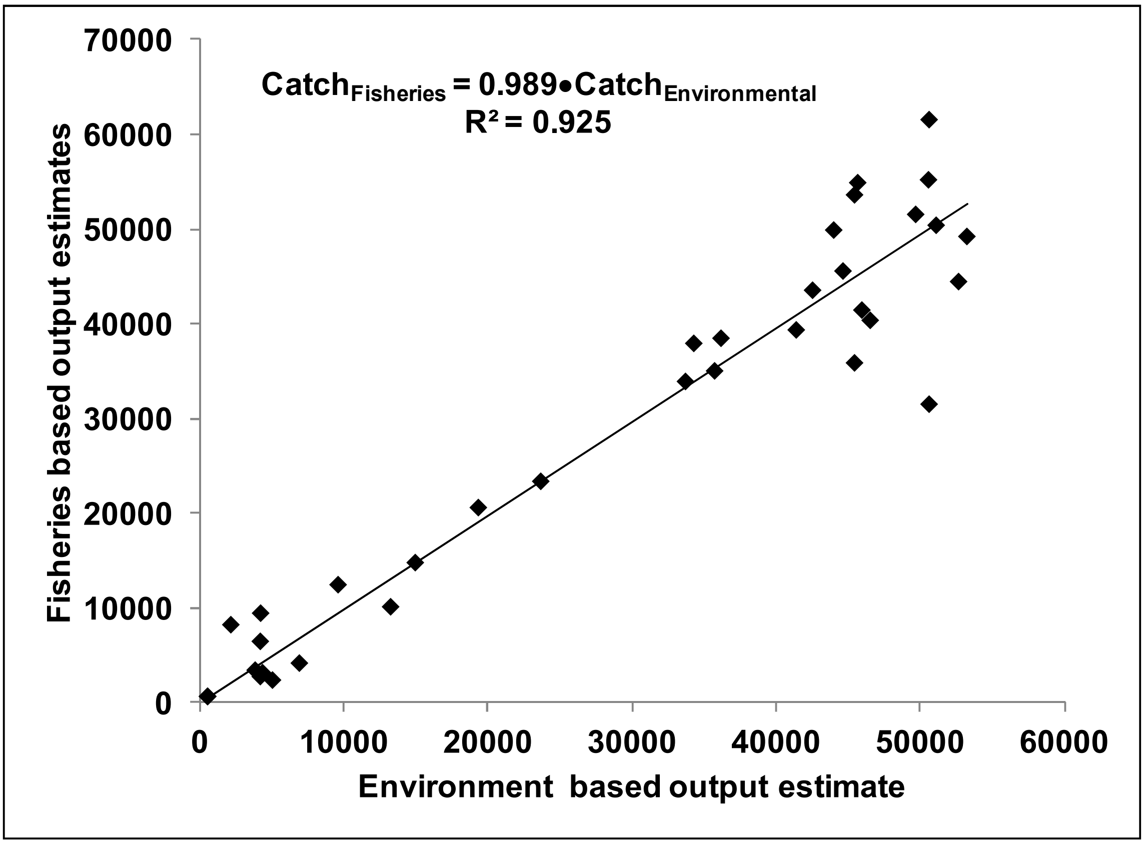

Figure 21: Comparison of network outputs of *R. argentea* catches from Lake Victoria determined from fisheries data and environmental data (Source: Author)

CHAPTER FIVE

5.0 DISCUSSION

5.1 Food Types and Food Selection

Results from this study show that the number of food items in the stomach can be analyzed in two ways. The first method is independent of the food items in the environment to produce results on ontogenic changes in feeding habits based on probability approach while the second method uses the count of the food items in the environment per unit volume to determine food selection, based on prior probabilities. This study has shown that the changes in both food type and quantity consumed can be predicted using nominal logistic regression, which also provides the relative change in stomach content against fish size. Many reports on ontogenic changes in food and feeding habits are descriptive but this study provides an objective and quantitative way of explaining ontogenic changes in food in *R. argentea* from Lake Victoria as compared to food selection indices developed by Ivlev (1961) and Strauss (1979).

The study shows that more of the larger sized zooplankton species among the copepods (*Thermocyclops oblingatus*), the rotifers (*Brachionus falcutus*) and the cladocera (*Moina macrurus*) are selected. This finding is in line with the feeding strategies that reduce the energy cost of prey capture and maximize the returns according to Ainsworth *et al*. (2010). The size dependent food selection strategy in this study is reflected by probability, providing an estimate of uncertainty other than the level of variation as provided by the classical food selection indices (Chesson, 1978; Strauss, 1979).

In view of the fact that *R. argentea* exhibits extended vertical migration for different size cases (Wanink, 1989), the shift in diet with size could also be interpreted as a feeding strategy in response to the vertical migration of zooplankton prey species. *R. argentea* is limited in vertical distribution in the water column and is more sensitive to hypoxia (<1–2 mg O_2 L^{-1}) than Nile perch (Wanink *et al*., 2002). It can then be argued that the oxycline-dwelling *R. argentea* are not seeking a predation refugium but that they are limited by low oxygen levels in reaching their traditional feeding areas near the bottom and hence have adopted a feeding strategy that is both predator and prey size dependent. Victor and Brown (1990) reported similar changes in diet in relation to size in *Brycinus nurse* and *B. longipinnis* in a perturbed river in Benin. Similarly, *Oreochromis niloticus* and fry of *Sarathoredon melanotheron, Heterotis niloticus* and *Brycinus nurse* in Asa reservoir in Nigeria all exhibit similar feeding behaviours (Akintunde, 1986; Ugwumba and Adebisi, 1992; Sailu, 2002).

Whereas this study has made no assumptions about a feeding model, observed size-related difference can be attributed to two possibilities: i) the encounter rate of *R. argentea* with prey ii) capture efficiency, i.e. the time and energy spent by *R. argentea* in pursuing a given prey type and successful pursit are the major determinants of prey selection. Drenner *et al*. (1978) found that the probability of escape was highest for *Chaoborus* and calanoid copepods, intermediate for cyclopoid copepods and lowest for cladocera in food selection of *Xenomelaniris Venezuelae* in Lake Valencia. The use of probability approach (BBN) in analysing the food selection of *R. argentea* in Lake

Victoria therefore offers a better option in dealing with ontogenti changes in food and feeding habits.

The ability to use BBN in the analysis of stomach content has demonstrated that there is very little change in the uncertainty (0.051% to 0.014%), also known as entropy, associated with selection of all the food items. In information theory, entropy is a measure of the uncertainty associated with a random variable (Shannon, 1948) and in this context; the term quantifies the expected change in the value of the information on food selection (Brillouin, 2004). Equivalently, the Shannon entropy is a measure of the average information content one is missing when one does not know the value of the different food items in either the stomach or the environment.

5.2 Fecundity and Recruitment

The basic assertion of recruitment is that if there is no spawning biomass (SSB), there would be no recruitment (Ricker, 1954). In view of the available data for this study, the conversion of length into weight and computation of spawning biomass was based on: i) sex ratio ii) size at maturity iii) the proportion of mature fish (Stage 3-6). The spawning biomass is therefore considered to be mature females of reproductive size that contribute to egg production and hence recruitment. The estimated spawning biomass index of 56,754 is 90% of the estimated total biomass (62,997) of active females in the fishery. With units in grammes and sample data, the results are indices of spawning stock biomass that can be converted to actual biomass fro reliable estimates such as hydroacoustics. This SSB index provides a possible explanation of the resilience of the species in Lake Victoria. A number of models have set the reference point for spawning biomass at 40% of the total biomass (Pitcher *et al.*, 1996). The high estimated spawning biomass of 90% determined in this study compared to the reference spawning biomass of 40% provides a good explanation for the resilience of *R. argentea* in Lake Victoria despite the shifts in species composition (Balirwa *et al.*, 2005; Bayona *et al.*, 2005; Masai *et al.*, 2005; Manyala and Ojuok, 2007), changes in water quality and eutrophication (Hecky, 1984; 1993; Ochumba and Kibaara, 1989; Lung'ahyia *et al.*, 2000; Gikuma-Njuru and Hecky, 2005; Hecky *et al.*, 2010) and water hyacinth invasion of the lake (Wawire and Ochiel, 2005; Njiru *et al.*, 2005).

The analysis presented in this study provides a method of determining SR relationship using egg production as a proxy and taking into consideration the limitation of egg and larval surveys in Lake Victoria. All the three classical approaches to egg production studies; annual egg production (Saville, 1980), daily egg production (Parker, 1985) and daily fecundity reduction (Lo *et al.*, 1992), rely on egg and larval surveys and the estimation of fecundity. Even though the spawning stock biomass (SSB) is commonly used to fit such SR models, the SSB is often based on VPA and does not discriminate between the males and females (Beverton and Holt, 1956). VPA as classical method make the same assumptions of constant recruitment and constant mortality (state of dynamic equilibrium) that are never achieved in reality. The present study has attempted to move away from these classical methods to the alternative non-classical approaches and this means that present study can be applied to the biomass estimate without assuming constant mortality.

This study used the Total Egg Production (TEP) that takes into account both the size-dependent capacity of females to produce eggs and the demographical structure of the spawning stock. The TEP has been used and found to be a more relevant stock reproductive potential index for the European hake *Merluccius merluccius* (Mehault *et al*., 2010).

It has been observed by Petterson (1999) and Frøysa *et al*. (2002) that regardless of any management measures instituted for any fisheries, if there is recruitment failure, then that stock or fishery faces an eminent collapse. However, it has not been possible to develop suitable recruitment models because data in egg and larval surveys in many stocks and many fisheries are scarce (Sparre *et al*., 1989). This study has attempted to bridge the gap using diverse sources of information that relate to fecundity (Okedi, 1971; Wanink, 1989; Manyala *et al*., 1992), sex ratio (Okedi, 1973; Wandera, 1992; Manyala *et al*., 1992) and length-weight relationship (Manyala *et al*., 1995a) on length-structured size frequency (LVFO, 2005) to model recruitment. This study has therefore shown that recruitment can be modelled using analytical methods and non-classical stock-recruitment models. This method can be applied on a geo-spatial scale if the biomass distribution of *R. argentea* is known since the conversion of size to weight (Manyala, 2005a) has been established and the relationship between egg production and size has also been established in this study.

Obviously, there can be no recruits if no fish are left to mature, spawn and produce eggs which hatch and grow to become recruits. According to a review by Manyala and Ojuok (2007), the females of 27 fish species of Lake Victoria are extremely fecund, producing thousands, even millions of eggs during their adult life. This enormous fecundity has generally given a false impression to fisheries biologists that even a very small parental stock should be able to rebuild the stock after each spawning season. In his review on fisheries management, Caddy (1999) observes that until S-R models developed by Ricker (1954), it was assumed that features of the abiotic environment are the major factors determining how many of the spawned eggs would survive to become recruits. It was believed that the spawning stock biomass (biomass of mature fish) was virtually an irrelevant factor for the determination or recruit numbers, except in cases of stock sizes close to zero. This lack of a definite S-R relationship in previous fisheries management models and measures was discussed in Beverton and Holt (1957) and Berverton (1963). Later works (Parrish, 1973 and Saville, 1980) suggested that many fish stocks do display S-R relationship and that recruitment overfishing (Murphy, 1977) was responsible for the depletion and subsequent collapse of many fisheries

However, S-R relationships generally cannot be established directly by plotting the number of recruits (or some index of recruitment) on spawning stock biomass. Rather, it is necessary to simultaneously account for S-R relationship and the biotic and/or abiotic factors which may affect it (Csirke, 1980). The present study determined the threshold in recruitment of *R. argentea* as: Number of females times the average egg production to be an index between 7,057 and 7,139. This threshold in recruitment is adequately described in both the Ricker (1954) and Beverton and Holt (1957) S-R models. Both the Ricker (1954) and Beverton and Holt (1957) S-R models states that recruitment decreases from a maximum level towards zero as the production of eggs increases. Even though the two models seem to revolve around the same principles, the present study follows more closely the Ricker (1954) model where there is always a relationship between spawning stock biomass as opposed to the Beverton and Holt (1957) model where no relationship axists above some spawning stock biomass. The present study showed clearly that recruitment in *R. argentea* does not follow a normal distribution as indicated by Skewness-Kurtosis tests. The three parameter Gamma or Weibull distribution best describes the S-R relationship of *R. argentea* in Lake Victoria. Unfortunately, classical S-R models also do not follow a normal distribution and the difficulty in using them arise from the lack of suitable techniques of estimating the parameters of a non-normal distribution when the probability density functions (pdf) are not proberly known. The pdf of both Gamma and Weibull are well known for given moment generating functions.

From these results we can address the following issues:

i) Can the survival and resilience of *R. argentea* in Lake Victoria be explained through its reproductive potential and ecological strategies?
ii) How does the relative reproductive potential of *R. argentea* (based on weight) compare with that of Nile perch and other commercially important species?
iii) Does the S-R relationship of *R. argentea* explain the resilience of this species in Lake Victoria?

Wanink (1989) report that the absolute fecundity of *R. argentea* has halved since Okedi's (1971) study and ascribes this to the dwarfing of this species in Lake Victoria. For this reason, the fecundity-length relationship of Manyala *et al*. (1992) was used in this study to account for the apparent reduction in the number of eggs per female. These results showed that at a sex ratio of 1:1, and 22.5 million eggs from only 56 kg sample of mature fish in this study, egg production can be estimated at about 22.5 trillion eggs per spawning season. At about 400,000 eggs kg^{-1}, we expect an annual production of 2.25×10^{13} eggs from *R. argentea* at the current catch of about 50,000 tonnes in the Kenya part of Lake Victoria. However, at a production to biomass (P/B) ratio of 3 to 4, the actual egg production would only be ¼ to ⅓ of this estimate. This estimate could be used, together with mortality/survival rates, to estimate total annual recruitment of *R. argentea* in Lake Victoria.

Due to predation and fishing, the increased mortality of *R. argentea* is probably responsible for reduction in its size at maturity in Lake Victoria according to Manyala and Ojuok (2007). The population growth characteristics of *R. argentea* in Lake Victoria show characteristic typical of r–selection with a high Von Bertalanffy growth coefficient (K), short lifespan (2–3 years), a high natural mortality rate independent of population size (rarely reaching the maximum carrying capacity), high rates of egg production at low trophic levels and high P/B ratio with small body size. The finding in this study of egg production distribution Kurtosis of up to 5.6 therefore statistically confirms that the species is highly fecund. Based on the results of this study, it is expected that translation of S-R from length to weight is adequate for direct estimation of S-R using SSB.

5.3 Catch and Production Analysis

Since the network acts as a dynamic database, it can be queried with new datasets either in the range or outside the range of the original data but when out of range datasets is used for querying the network, larger errors are sometimes produced (Alyuda Research, 2005).

Assuming that environmental variables are limiting in the production of *R. argentea* in Lake Victoria, then the network would predict production when the fishing regime remains relatively constant over a long period of time. In such a case, productivity will be limited only by the physical environment. On the other hand, if the physical environment does not change over several years, then the production will be influenced by the fishing regime. Changes in catches reflected in the fishing regime are likely to be a consequence of absolute changes in gear types and gear-vessel combinations, spatial changes in the distribution of effort, changes in absolute effort or even changes in fishing power.

Changes in fishing regime is also often related to socio-economic and technological factors. In a single-stock, single-fleet perspective, classical population dynamics models (Schaefer, 1934; Ricker, 1949; Beverton and Holt, 1956) provide appropriate answers. But when it comes to multispecies, multi-fleet fisheries, fleets depend on several fish stocks according to Daurès *et al*. (2009) and the different coexisting fish stocks are exploited by many fleets of different design, efficiency and fishing power. It has been noted that fleet behaviour can change considerably changes in response to various factors, including technological progress, management regulations, social needs and resource availability (Baelde 2001; Christensen and Raakjær, 2006), thereby affecting the overall yield or catch.

For *R.argentea* in the Winam Gulf of Lake Victoria, there is a good advantage to fisheries management by specifying management objectives for each or any of the species used in the ANN model and possible alternative objective for *R. argentea* estimated by manipulating the fishing regime. Even though the use of ANN is rare in fish stock assessment, this study has shown that the method can be applied to Lake Victoria for *R. argentea*. It is also possible to use this approach for total catch on condition that there are no data gaps according to computational requirements. One method of handling missing data is to use the residuals with the mean value to estimate the missing observation.

CHAPTER SIX

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

6.1.1 Food Types and Food Selection

Based on the results of the logistic regression and Bayesian Belief Network, the following conclusions are made:

i) *R. argentea* is likely to significantly select the food items *T. oblingatus* and *M. macrourus* as compared to the food item *Epiphanes*.

ii) The selection of the food item *T. oblingatus* and *M. macrourus* is 3.8 and 3.5 times respectively higher as compared to the food item *Epiphanes* as the fish size change from 10 mm to 40 mm.

iii) The selection of the food item *T. oblingatus* and *M. macrourus* is 10.3 and 11.4 times respectively higher as compared to the food item *Epiphanes* as the fish size change from 10 mm to 50 mm.

iv) The food selection was found to be heterogeneous between *Epiphanes* and *T. oblingatus, B. lonirostris, M. macrourus* and *T. emini.*

v) The generated feeding model was found to be stable, based on minimum percentage entropy reduction, variance reduction and belief variance while the root mean square change was less then 0.0027 for all the prey items..

6.1.2 Sexual Maturity, Fecundity and Recruitment

From this study, it is concluded that:

i) Modelling of maturity, fecundity, sex ratio and length-weight relationship provided a robust way of determining the S-R relationship in the study.

ii) The best S-R relationship was provided by the Gamma and Weibull distributions while the Log normal distribution did not give consistent parameters.

iii) The Gamma distribution produces a higher shape parameter as compared to the Weibull distribution hence higher recruitment over a narrow range of size classes while the Weibull distribution produces a relatively lower recruitment over a larger range of size classes.

6.1.3 Catch and Production

Based on the environmental variables and fisheries data, the study concludes that:

i) Artificial Neural Networks developed from environmental and catch data adequately predicted catches of *R. argentea* in Lake Victoria.

ii) The catches predicted by the ANN compare well with observed catches for both environmental and catch data.

iii) For environmental data, the lake level and evaporation were the most important variables that determined the performance of the ANN.

iv) *Lates niloticus* and *Haplochromis* catches were the most important species that determined the performance of the ANN.

v) For both environmental and catch data, the class benchmark was accounted for 40% and 50% importance to the ANN respectively.

6.2 Recommendations

6.2.1 Food Types and Food Selection

From this study and the results obtained, it is recommended that food and feeding habit studies of *R. argentea* in Lake Victoria should focus on the following:

i) Structure and relationships between food items in the stomach and the environment using probability approaches such as BBN.

ii) Logistic regression should be used as one of the methods of studying food habits and food selection to provide concrete evidence on size dependent intraspecific resource partitioning in *R. argentea* in Lake Victoria.

iii) The study on food habits and prey selection based on comparison of the main species such as *T. oblingatus*, *M. macrourus* and *T. emini* with other prey species as well as among larger taxonomic groups (copepods, cladocera and rotifers).

6.2.2 Fecundity and Recruitment

In view of the approaches and results obtained in this study, it is recommended that:

i) The results obtained in this study be used for planning of egg and larval surveys of *R. argentea* in Lake Victoria for the purposes of determining the Total Egg

Production (TEP) and Female Spawning Stock Biomass (FSSB) for the purpose of monitoring annual recruitment strength.

ii) The management of *R. argentea* should be based on the annual recruitment estimate for setting fishing targets as well as on Biological Reference Points (BRP).

6.2.3 Catch and Production

Based on the dynamic predictive power of Artificial Neural Network, future analysis of production of *R. argentea* in Lake Victoria should:

i) Concentrate on non-classical methods such as ANN using both catch and environmental data for *R. argentea* in Lake Victoria.

ii) In the use of ANN method for analysis of catch and production, emphasis should be placed on the analysis of network errors and the importance of input variables in explaining the quality of the network.

iii) Testing and querrying of the network should be done using the range of datasets in the range used for developing the network as well as datasets outside the range used so as to be able to analyze the sensitivity of different network architectures.

REFERENCES

Abira, M. A., C. Oleko, J. O. Okungu, J. O. Z. Abuodha and R. E. Hecky (2005). Industrial and municipal effluent loadings into the Lake Victoria catchment, Kenya. *Lake Victoria Environmental Management Project, National Water Quality Synthesis Report. KARI-LVEMP*

Acere, T. (1988). Recent trends in the fisheries of Lake Victoria (Uganda and northern part). *FAO Fish. Rep.,* ***388:*** *36-41***.**

Agresti, A. (1984). Analysis of Ordinal Categorical Data, John Wiley & Sons, Inc.

Agresti, A. (1990). *Categorical Data Analysis.* John Wiley and Sons, Inc.

Ahrens, J. H. and U. Dieter (1982). Generating gamma variates by a modified rejection technique. *Comm. ACM*, ***25:*** *47–54*.

Ainsworth, C., I. Kaplan, P. Levin and M. Mangel (2010). A statistical approach for estimating fish diet compositions from multiple data sources: Gulf of California case study. *Ecol. Appl.,* ***20(8):*** *2188–2202.*

Alyuda Research (2005). *Alyuda NeuroIntelligence Ver. 2.2*. Alyuda Research Inc. (2001-2005)

Anderson J. and E. Rosenfeld (1990). NeuroComputing: Foundations for Research. *MIT Press, (Vol I, 1988) and (Vol II, 1990).*

Argeant B. L. S., E. E. Gaiser and J. C. T. Rexler (2012). Indirect and direct controls of macroinvertebrates and small fish by abiotic factors and trophic interactions in the Florida Everglades. *Freshwat. Biol.,* ***56:*** *2334 – 2346.*

Asila, A. A. and J. Ogari (1988). The growth parameters and mortality of Nile perch (*Lates niloticus* L.) estimated from length-frequency in the Nyanza Gulf (Lake Victoria). *FAO Fish. Rep.,* ***389:*** *272 - 287.*

Asila, A. A., S. O. Dache and C. O. Rabuor (1990). Influence of beach and mosquito seine on the fisheries of the Nyanza Gulf. In: *Proceedings of the symposium on socio-economic aspects of Lake Victoria fisheries (Vol. 1 Unedited papers 1-7). FAO/UNDP RAF/87/099-WP/05/89 Eng.* ***p*** *1 – 18*

Baelde, P. 2001. Fishers' description of changes in fishing gear and fishing practices in the Australian South East Trawl Fishery. Mar. Freshw. Res.52(4): 411–417.

Balirwa, J. S. and F. W. B. Bugenyi (1980). Notes on the fisheries of the River Nzoia, Kenya. *Biol. Cons.,* ***18:*** *53 – 58.*

Banneheka, B. M. S. G. and G. Ekanayake (2009). A new point estimator for the median of Gamma distribution. *Viyodaya J. Science,* ***14:*** *95 – 103*.

Baranov, F. I. (1918). On the question of the biological basis of fisheries (Translated from Russian by W.E. Ricker, 1945). *Izvestiya* ***1:*** *81–128.*

Bellows Jr., T. S. (1981). The descriptive properties of some models for density dependence. *J. Anim. Ecol.,* ***50:*** *139 – 156*.

Benda, R. S. (1981). A comparison of bottom trawl catch rates in the Kenyan waters of Lake Victoria. *J. Fish Biol.,* ***18: 13 -24***.

Beverton, R. J. H. and S. J. Holt (1956). A review of methods for estimating mortality rate of exploited populations, with special reference to source of bias in catch sampling. *Rapp. P.-V. Reun. Cons. Int. Explor. Mer.,* ***140(1):*** *67-83*.

Beverton, R. J. H. and S. J. Holt (1957). On the dynamics of exploited fish populations. *Fish. Invest. Min. Agric. Fish. Food G.B. (2 Sea Fish.),* ***19:*** *533 p*.

Bhattacharya, C. G. (1967). A simple method of resolution of a distribution into Gaussian components. *Biometrics,* ***23:*** *115-135*.

Bishop, C. M. (1992). Exact Calculation of the Hessian Matrix for the Multi-layer Perceptron. *Neural Comp, **4:** 494 – 501.*

Branstrator, D. K., J. T. Lehman and L. M. Ndawula (1996). Zooplankton dynamics in Lake Victoria. In: *The Limnology Climatology and Paleoclimatology of the East African Lakes* (Johnson and T.C., Odada O. (Eds)*, Gordon and Breach, Toronto. **p** 337-355.*

Brosse, S., S. Lek and C. R. Townsend (2001). Abundance, diversity, and structure of freshwater invertebrates and fish communities: an artificial neural network approach. *N.Z. J. Mar. Freshwat. Res., **35:** 134-145.*

Brown, C. C. (1982). "On a Goodness of Fit Test for the Logistic Model Based on Score Statistics," *Comm. in Stat.*, ***11:** 1087–1105.*

Burnham, K. P., D. R. Anderson and K. P. Huyvaert (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol Sociobiol., **65:** 23–35.*

Bwathondi, P. O. J. (1988). State of Lake Victoria fisheries, Tanzanian sector. *FAO Fish. Rep. **388:** 29-35.*

Caddy, J. F. (1999). Fisheries management in the twenty-first century: will new paradigms apply. *Rev. in Fish Biol. and Fish., **9:** 1–43.*

Calamari, D., M. O. Akech and P. B. O. Ochumba (1995). Pollution of Winam Gulf, Lake Victoria, Kenya: A case study for preliminary assessment. *Lakes and Res. Res. and Manag., **1:** 89–106.*

Callan, R. (1999). *The Essence of Neural Networks*. Prentice Hall Publishers, *125 pp*.

Chabot, F. and E. J. Maly (1986). Variation in diet of yellow perch (*Perca flavescens*) in a Quebec reservoir. *Hydrobiologia, **137(2):** 117-124.*

Chesson, J. (1978). Measuring preference in selective predation. *Ecology, **59:** 211-215*.

Choi K. P. (1994). On the medians of Gamma distributions and an equation of Ramanujan. *Proc Amer. Math. Soc ., **121 (1):** 245–251*

Christensen, A.-S., and J. Raakjær 2006. Fishermen's tactical and strategic decisions: a case study of Danish demersal fisheries. Fish. Res.81(2–3): 258–267.CIFA (Committee for Inland Fisheries of Africa) (1988). Report of the fourth session of the sub-committee for the development and management of Lake Victoria. Kisumu, Kenya, 6-10 April 1987. *FAO Fish. Rep., **388: 112p***.

Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge University Press, New York.

Clark, J. D., J. E. Dunn and K. G. Smith (1993). A multivariate model of female bear habitat use for a geographic information system. *J. Wildl. Manag., **57:** 519-526*.

Copley, H. (1953). The tilapia fishery of Kavirondo Gulf. *J. E. Afr. Nat. Hist. Soc., **94:** 1-5.*

Csirke, J. and J. F. Caddy. (1983). Production modelling using mortality estimates. *Can. J. Fish. Aquat. Sci., **40:** 43–51.*

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function," *Math. Contr., Signals, Syst., **2:** 303–314*.

Daurès, F., Rochet, M.J., Van Iseghem, S., and Trenkel, V.M. 2009. Fishing fleet typology, economic dependence, and species landing profiles of the French fleets in the Bay of Biscay, 2000–2006. Aquat. Living Resour.22(4): 535–547.

Deriso, R. B. (1980). Harvesting strategies and parameter estimation for an age-structured model. *Can. J. Fish. Aquat. Sci., **37:** 268–282*.

Deriso, R. B., T. J. Quinn and P. R. Neal (1985). Catch-age Analysis with auxiliary information. *Can. J. Fish. Aquat. Sci.,* ***42:*** *815 – 824*.

Dickson, P. and A. M. Ellison (1996). Introduction: ecological applications of Bayesian inference. *Ecol. Appl.,* ***6:*** *1034–1035*.

Dunn, P. O. and C. E. Braun (1986). Summer habitat use by adult female and juvenile sage grouse, *J. Wildl. Manag.,* ***50:*** *228-235*.

Efron, A and C. Gong (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross Validation. *The Am. Stat.,* ***45:*** *211-219.*

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters, 7****:*** *509–520*.

Fish, G. R. (1957). A section movement and its effect on hydrology of Lake Victoria. *Fisheries Publication, London,* ***10: 1 - 68***.

Fitzpatrick J. (1 995): Technology and Fisheries legislation. Paper presented to the International Technical Consultation on the Precautionary Approach to Capture Fisheries. Lysekil, Sweden, 6-1 3 June 1995.

Fournier, D. and C. P. Archibald (1982). A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.,* ***39:*** *1195 – 1207*.

Frøysa, K. G., B. Bogstand and D. W. Skagen (2002). Fleksibest – an age-length structured fish stock assessment model. *Fish. Res.* ***55:*** *87 – 101*.

Fry, F. E. J. (1949). Statistics of a lake trout fishery. *Biometrics,* ***5:*** *27–67*.

Geman, S., E. Bienenstock and R. Doursat (1992). Neural networks and the bias: variance dilemma. *Neural Comp.,* ***4:*** *1–58*.

Getabu, A. (1988). Aspects of the Lake Victoria fisheries, with emphasis on *Oreochromis niloticus* and *Alestes sadleri* from the Nyanza Gulf. *FAO Fish. Rep.,* ***389:*** *416-431*.

Getabu A., R. Tumwebaze and D. N. MacLennan (2003). Spatial distribution and temporal changes in the fish populations of Lake Victoria. *Aquat. Liv. Res.,* ***16:*** *159–165.*

Gikuma-Njuru P., R. E. Hecky (2005). Nutrient concentrations in Nyanza Gulf, Lake Victoria, Kenya: light limits algal demand and abundance. *Hydrobiologia,* ***534:*** *131-140*.

GoK (2008). *Fisheries Statistical Bulletin 2008*. Ministry of Livestock and Fisheries Development Report.

GoK (2009). *Fisheries Statistical Bulletin 2009*. Ministry of Fisheries Development Report.

GoK (2010). *Fisheries Statistical Bulletin 2010*. Ministry of Fisheries Development Report.

Gor, S. A., P. O. Peterlis, J. O. Okungu, J. O. Z. Abuodha and R. E. Hecky (2005). Non-point loading from the Lake Victoria Catchment: Kenya Section. *Lake Victoria Environmental Management Project (LVEMP), National Water Quality Synthesis Report KARI-LVEMP.*

Gordon, H. S. (1954). Economic theory of a common property resource: the fishery. *J. Political Econon.****75:*** *274–286.*

Graham, M., (1929). The Victoria Nyanza and its fisheries. *A report on the fishing surveys of Lake Victoria (1927–1928). Crown Agents Colonies, London*.

Graham, M. (1935). Modern theory of exploiting a fishery and application to North Sea trawling. *J. Cons. Int. Explor. Mer****10*****:** *264–274*

Grossman, G. D., R. E. Ratajczak, M. Crawford and M. C. Freeman (1998). Assemblage organization in stream fishes: effects of environmental variation and interspecific interactions. *Ecol. Monogr.,* ***68:*** *395–420*.

Guégan, J. F., S. Lek and T. Oberdorff (1998). Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature,* ***391:*** *382-384*.

Gulland, J. A. (1983). Fish stock assessment: a manual of basic methods. *FAO/Wiley Series on Food and Agriculture, V.1,* ***223 pp***.

He, F., J. Zhou and H. Zhu (2003). Autologistic regression model for the distribution of vegetation. *J. Agric. Biol. Environ. Stat.,* ***8:*** *205–222*.

Hecky, R. E. (1984). Africa lakes and their trophic efficiencies: a temporal perspective. In: *Trophic interactions within aquatic ecosystems. AAAS Selected Symposium 85 (D. G. Meyers and J. R. Stricker (eds.),* ***pp.*** *405 - 448.*

Hecky R. E. (1993). The Eutrophication of Lake Victoria. *Ver. Intern. Ver. Limn.. Sturtgart.,* ***25:*** *39-48*.

Hecky, R. E., F. W. B. Bugenyi, P. B. O. Ochumba, J. F. Talling, R. Mugiddde, M. Gophen and L. Kaufman (1994). Deoxygenation of the deep water of Lake Victoria, East Africa. *Limn and Ocean.,* ***39: 1476-1480***.

Hilborn, R., and C. J. Walters (1992). *Quantitative fisheries stock assessment: choice, dynamics, and uncertainty*. Chapman and Hall, New York, N.Y. ***570 pp***

Hjort, J. and C. G. J. Petersen (1905). Short review of the results of the International Fisheries Investigations. *Rapp. Cons. Perm. Int. Explor. Mer 3: Appendix G*.

Holgate, P. (1989). "The lognormal characteristic function". *Commun. in Stat. - Theory and Methods.,* ***18 (12):*** *4539–4548.*

Horne, J. S., E. O. Garton and J. L. Rachlow (2008). A synoptic model of animal space use: simultaneous estimation of home range, habitat selection and inter/intra-specific relationships. *Ecol. Model.,* ***214:*** *338-348*.

Hornik, K., M. Stinchcombe and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks,* ***2:*** *359-366*.

Ivlev, V. S. (1961). *Experimental ecology of the feeding of fish.* Yale University Press, New Haven. ***302 pp*** (translated from Russian).

Jacobs, J. (1974). Quantitative measurement of food selection. *Oecologia (Berl.).,* ***14:*** *413 – 417.*

Jacobs, R. (1988). "Increased rates of convergence through learning rate adaptation." *Neural Networks****, 1:*** 295 – 307.

Jain, A. K., R. C. Dube and C. Chen (1987). Bootstrap techniques for error estimation. *IEEE Trans. Patt. Anal. Mach. Intell. PAMI,* ***9: 628–633***.

Johnson, J. B. and K. S. Omland (2004). Model selection in ecology and evolution. *Trends in Ecol. and Evol.,* ***19(2:)*** *101 – 108.*

Johnson, D.S., D. L.Thomas, J. M. Ver Hoef and A. Christ (2008). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics,* ***64:*** *968-976.*

Kaedi, M., and N. Ghasem-Aghaee (2011). Biasing Bayesian Optimization Algorithm using Case Based Reasoning. *Knowl. Based Syst.* ***24:*** *1245–1253.*

Kalff, J. (1983). Phosphorous limitation in some tropical African inland Lakes. *Hydrobiologia,* ***100:*** *101-112.*

Kitaka, G. E. B. (1971). An instance of cyclonic upwelling in southern offshore waters of Lake Victoria. *Afr. J. Trop. Hydrobiol.,* ***1:*** *85 - 92.*

Kite, G. W. (1981). Recent changes in level of Lake Victoria. *Hydrolog. Soc. Bull.,* ***9:*** *233 - 243.*

Kohavi, R. (1995). A study of cross-validation and bootstrap for estimation and model selection. In: *Proceeding of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers. p. 1137–1143*

Kudhongania, A. W. and A. J. Cordone (1974). Bathospatial distribution pattern and biomass estimate of major demersal fishes in Lake Victoria. *J. Trop. Hydrobiol. Fish.,* ***3:*** *15 - 31.*

Kudoja, W. M., C. A. Muhando, J. D. Michael and B. C. M. Sayi (1992). The vertical distribution of *Haplochromis* species in relation to the phytoplankton and zooplankton distribution. *Paper presented at the IDEAL meeting on biodiversity, people and socio-economics of Lake Victoria, August, 1992, Jinja, Uganda.*

Lafebre C. and J. Principe (1992). "Object-oriented artificial neural network implementations." *World Cong. Neural Networks,* ***4:*** *436-439.*

Lafebre, C., J. Principe, J. D. Samson, D. Wooton, G. Geniesse, M. Lucas, C. Funcecourt, J. Gerstenberger, N. Euliano, G. Lynn, M. Allen and D. Morossero (2005). *NeuroSolution*. NeuroDimension Inc. 1994 – 2005. Ganesville, FL.

Larkin, P.A. and Ricker, W. (1964). Further information on sustained yields from fluctuating environments. *J. Fish. Res. Bd Can.,* ***21:*** *1–7.*

Lecowicz, M. J. (1982). The sampling characteristics of electivity indices. *Oecologia (Berl.).*, **52:** 22 – 30.

Lek, S., J. L. Giraudel and J. F. Guégan (2000). Neuronal networks: algorithms and architectures for ecologists and evolutionary ecologists. In: *Artificial Neuronal*

Networks: Application to Ecology and Evolution (Lek, S., Gue ´gan, J. F. (Eds.), *Springer-Verlag, Berlin.* ***p*** *3-27.*

Lele, S. R. (2007). A new method of estimation of resource selection probability function, *J. Wildl. Manag.,* ***73:*** *122-127.*

Levin, P.S., Fogarty, M.J., Murawski, S.A., and Fluharty, D. 2009. Integrated ecosystem assessments: developing the scientific basis for ecosystem-based management of the ocean. PLoS Biol. 7(1): e1000014. doi: 10.1371/journal.pbio.1000014. PMID:19166267

Li, J., E. K. Burke and R. Qu (2011). Integrating neural networks and logistic regression to underpin hyper-heuristic search. *Knowledge-Based Systems,* ***24:*** *322–330*

Limpert, E., W. Stahel and M. Abbt (2001). Log-normal Distributions across the Sciences: Keys and Clues. *BioScience,* ***51(5):*** *341–352.*

Luger, G. F. (2005). *Artificial Intelligence*. Addison Wesley.

Lung'yia, H. B. O., A. M'ttarzi, M. Tackx, J. Gichuki and J. J. Symoens (2000). Phytoplankton Community Structure and Environment in the Kenyan Waters of Lake Victoria. *Freshwat. Biol.,* ***43:*** *529 – 543*.

LVFO (2005). Report of the Regional Working Group on *Rastrineobola argentea* in Lake Victoria. Jinja, **p. 37**

Madan, D. B. and E. Seneta (1990). The variance gamma (V.G.) model for share market returns, *Journal of Business*, ***63:*** *511 - 524*.

Manel, S., J. M. Dias and S. J. Ormerod (1999). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecol. Model.,* ***120:*** *337–347*.

Manly, B. F. J., P. Miller and L. M. Cook (1972). Analysis of a selective predation experiment, *Am. Nat.,* ***106:*** *719-36,*

Manyala, J. O. (1993). Growth, mortality and mesh size selection of dagaa *Rastrineobola argentea* (Pellegrin 1904) in the Winam Gulf of Lake Victoria (Kenya). In: *Papers presented at the symposium on biology, stock assessment and exploitation of small pelagic fish species in the African Great Lakes region* (B. E. Marshall, R. Mubamba (Eds.))*, 1992 November 25–28, Bujumbura, Burundi. CIFA* p ***246–256****.Occasional Paper No. 19. GCP/RAF/271/FIN-TD/06 (En.). CIFA Occ. Pap.* ***19.***

Manyala, J. O., 1994. Knowledge gaps in the biology and ecology of *Rastrineobola argentea* (Pellegrin) in Lake Victoria*. FIRI/IDRCWorkshop on Environment, Fisheries and Socioeconomic changes in Lake Victoria Basin. 1993 November 15–20. Jinja, Uganda.*

Manyala, J. O. (2005). Fisheries Synthesis Report for Kenya. *Lake Victoria Environmental Management Project (LVEMP) Report. KARI-LVEMP Technical Report, 2005, Kisumu, Kenya.*

Manyala, J. O., C. O. Nyawade and C. O. Rabuor (1992). The Dagaa (*Rastrineobola argentea* Pellegrin) fishery in the Kenyan waters of Lake Victoria: A national review and proposal for future research. In: *The Lake Victoria Dagaa (Rastrineobola argentea). Report of the First Meeting of the Working Group on Lake Victoria* Rastrineobola argentea (Mannini, P. (ed))*, 9-11 December 1992, Kisumu, Kenya. UNDP/FAO Regional Project for Inland Fisheries Planning. RAF/89/099-TD/38/92 (En). .* ***p*** *18 – 35.*

Manyala, J. O., E. Vanden Berghe and S.. Dadzie. (1995a). Growth, mortality, exploitation rate and relative yield of *Rastrineobola argentea* (Pellegrin 1904) in the Winam Gulf of Lake Victoria (Kenya). *Scientia Marina, **59(3-3):** 555-563*.

Manyala, J. O., E. Vanden Berghe and S.. Dadzie. (1995b). Morphometrics of *R. argentea* in Lake Victoria. *East Afr. J. Trop. Hydrobiol. Fish **6:** 35–42*.

Manyala, J. O. (2006). Linking poverty to biodiversity assesment of existing data within Lake Victoria basin. *World Wildlife Fund (WWF), Netherlands Environmental Assessment Agency and Kenya Wildlife Service (KWS). Nairobi, October 2006. **p.** 64*

Manyala, J. O. and J. E. Ojuok (2007). Survival of the Lake Victoria *Rastrineobola argentea* in a rapidly changing environment: Biotic and abiotic interactions. *J. Aqu. Ecosy. Health and Manag., **10(4): 407-415***.

Marsaglia, G and W. Tsang (2000). A simple method for generating gamma variables. ACM Transactions on Mathematical Software, 26(3):363-372, 2000

Marten, G. G., B. Wanjala and L. T. Galuka (1976). Exploratory trawling of the Lake Victoria fishery in Kenya during 1975. *EAFFRO manuscript, **19 pp***.

Marzluff, J.M, Millspaugh, J.J., Hurvitz, P. and Handcock, M.S. (2004). Relating resources to a probabilistic measure of space use: forest fragments and Stellar's jays. *Ecology **85:** 1411 – 1427*.

Maynard-Reid, P., U. Chajewska (2001). Aggregating learned probabilistic beliefs. In: *Proceedings of the 17^th^ Conference in Uncertainty in Artificial Intelligence (UAI'01), Seattle, WA. **pp.** 354–361.*

McDonald, T. L., B. F. J. Manly, R. M. Nielson and L. Diller (2006). Discrete choice modeling in wildlife studies exemplified by Northern spotted owl nighttime habitat selection, *J. Wildl. Manag., **70:** 375-383*.

Melack, J. M. (1976). Photosynthesis rates in four tropical African freshwaters. *Freshwat. Biol., **9:** 555 - 571.*

Melack, J. M. (1979). Primary productivity and fish yields in the tropical Lakes. *Trans. Am. Fish. Soc., **105:** 575 - 380*.

Minitab (1997). *Minitab User's Guide 2: Data Analysis and Quality Tools*. Minitab Inc., State College, Pennsylvania.

Mohammadzaheri, M., L. Chen, and S. Grainger (2012). Acritical Review of the Most Popular Types of Neuro Control. *Asian Journal of Control, **14(1):** 1-11.*

Muhoozi, I. L. (2002). Exploitation and management of the artisanal fisheries in the Ugandan waters of Lake Victoria. *PhD Theses, University of Hull. **pp**. 260*

Muller, R. G. and R. S. Benda (1981). A comparison of bottom trawl stock densities in the inner Kavirondo Gulf of Lake Victoria *Journal Fish Biology, **19: 339 - 401***.

Mwirigi, P. M., P. Gikuma-Njuru, J. O. Okungu, J. O. Z. Abuodha and R. E. Hecky (2005). Lake Victoria monitoring of the pelagic, littoral, river mouths and near shore urban environments, Kenya. *Lake Victoria Environmental Management Project (LVEMP), National Water Quality Synthesis Report. LVEMP-KARI.*

Myers, R. A. and P. Pepin. (1990). The robustness of lognormal-based estimators of abundance. *Biometrics **46:** 1185 – 1192*.

Myers, R. A., J. Bridson and N. J. Barrowman (1995). Summary of Worldwide Spawner and Recruitment Data. *Can. Tech. Rep. Fish. and Aqua. Sci., **2024:** iv + 327 p.*

Newell, B. S. (1960). The hydrology of Lake Victoria. *Hydrobiologia, **15:** 363-383.*

North, M. P. and Reynolds, J. H. (1996). Microhabitat analysis using radiotelemetry locations and polytomous logistic regression, *J. Wildl. Manag.,* ***60:*** *639-653*.

Ochumba, P. B. O. (1984). The status of Lake Victoria environment in Kenya, Lake Victoria. In: *Proceedings of the Shiga conference 1984 on conservation and management of world lake environments. Shiga Prefectural Government, Otsu, Japan.* ***p.*** *86-89.*

Ochumba, P. B. O. and D. I. Kibaara (1989). Observations on blue-green algal blooms in the open waters of Lake Victoria, Kenya. *Afr. J. Ecol.,* ***27:*** *23-34.*

Ogari, J. (1985). Distribution, food and feeding habits of *Lates niloticus* (L.) in Nyanza Gulf Lake Victoria (Kenya). *FAO Fish. Rep.,* ***335:*** *68 - 80.*

Ogari, J. and S. Dadzie (1988). The food of Nile perch, *Lates niloticus* (L.) after the disappearance of the haplochromine cichlids in the Nyanza Gulf of Lake Victoria. *J. Fish Biol.,* ***32:*** *571-577.*

Okach, J. I. O. and S. Dadzie (1988). The food, feeding habits and distribution of a siluroid catfish, *Bagrus docmac* (Forskal) in Kenyan waters of Lake Victoria. *J. Fish Biol.,* ***32:*** *21-26.*

Okedi, J. 1971. The food and feeding habits of the small mormyrid fishes of LakeVictoria, East Africa. *Afr. J. Trop. Hydrobiol. Fish.* ***1(1):*** *1–12*.

Okedi, J. (1973). Preliminary observations on *Engraulicypris argenteus* from Lake Victoria. *EAFFRO annual report 1973*.

Olsson, O. and N. M. A. Holmgren (1999). Gaining ecological information about Bayesian foragers through their behaviour. I. Models with predictions. *Oikos,* ***87:*** *251–263*.

Paloheimo, J. E. (1979). Indices of food type preference by a predator, *J. Fish. Res. Bd Can., **36:** 470-473*

Park, S. Y., and A. K. Bera (2009). "Maximum entropy autoregressive conditional heteroskedasticity model". *J. Econometrics, **29:** 219–230*.

Paulik, G. J. (1973). Studies of the possible form of the stock–recruitment curve. *Rapp. P.-v. Réun. Cons. Perm. Int. Explor. Mer **164:** 302–315.*

Pella, J. J. and P. K. Tomlinson (1969). A generalized stock produc-tion model. *Bull. Inter-Am. Trop. Tuna Comm., **13:** 419–496.*

Peterman, R.M. 2004. Possible solutions to some challenges facing fisheries scientists. ICES J. Mar. Sci. 61(8): 1331–1343.

Phillips, S. J., R. P. Anderson and R. E. Schapire (2006). Maximum entroy modeling ofspecies geographic distributions. *Ecol. Model., **190:** 231-259*.

Pinkas, L., M. S. Olipham and I. L. K. Iverson (1971). Food habits of albacore, bluefin tuna and bonito in Californian waters. *California Fish Game **152:** 1-105.*

Pope, J. G. (1972). An investigation of the accuracy of virtual population analysis using cohort analysis. *Bull. ICNAF, **9:** 65–74*.

Pope, J. G. and B. J. Knight (1982). Simple models of predation in multi-age multispecies fisheries for considering the estimation of fishing mortality and its effects. In: *Multi-species Approaches to Fisheries Management Advic* (Mercer, M.C. ed.). *Can. Spec. Publ. Fish. Aquat. Sci., **59:** 64–69*.

Prado, J., R. J. Beare, J. Siwo Mbuga and L. E. Oluka, 1991. A catalogue of fishing methods and gear used in Lake Victoria. UNDP/FAO Regional Project for Inland Fisheries Development (IFIP), FAO RAF/87/099-TD/19/91 (En). Rome, Food and Agricultural Organisation.

Rabuor, C. O. (1988). First Report of the Fish Stock Assessment on Artisanal Fishery of Winam Gulf of Lake Victoria (Kenya). *K.M.F.R.I Tech. Rep. 20 February 1988.*

Raymond, L., S. Lek and J. Moreau (1999). Predicting fish yield of African lakes using neural networks. *Ecological Modelling, **120:** 325–335*.

Reuter, U. and B. Möller. (2010). Artificial Neural Networks for Forecasting of Fuzzy Time Series. *Comp. Aid. Civ. and Infra. Eng., **25:** 363–374*

Ricker, W. E. (1949). Mortality rates in some little-exploited populations of freshwater fishes. *Trans. Am. Fish. Soc., 77**:** 114–128*.

Rijnsdorp, A.D., Poos, J.J., Quirijns, F.J., HilleRisLambers, R., De Wilde, J.W., and Den Heijer, W.M. 2008. The arms race between fishers. J. Sea Res. 60(1-2): 126–138

Ruck, D. W., S. K. Rogers and M. Kabrisky (1990). Feature Selection Using a Multilayer Perceptron. *J. Neur. Net. Comp., **2(2):** 40-48.*

Salamó, M., and M. López-Sánchez (2011). Adaptive case-based reasoning using retention and forgetting strategies. *Knowl. Based Syst. **24(2):** 230–247*.

Schaefer, M. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bull.I-ATTC/Bol. CIAT, **1(2):** 27-56*

Schnute, J., (1985). A general theory for analysis of catch and effort data. *Can. J. Fish. Aquat. Sci., **42:** 414-429.*

Schnute, J. (1987). A general fishery model for a size-structured fish population. *Can. J. Fish. Aquat. Sci., **44(5):** 924-940.*

Silsbe , G. M., R. E. Hecky, S. J. Guildford , R. Mugidde (2006). Variability of chlorophyll a and photosynthetic parameters in a nutrient-saturated tropical great lake. *Limnol. Oceanogr.* , ***51(5)****: 2052-2063*.

Sims, S. E. (1985). Selected computer programmes in FORTRAN for fish stock assessment. *FAO Fish. Tech. Pap.,* ***259:*** *183 pp*.

Sokal, R. R. and F. J. Rohlf (1995). *Biometry, the principles of statistics in biological research*. W. H. Freeman Company, New York.

Sparre, P. (1987). Computer programmes for fish stock assessment: length based fish stock assessment for Apple II computers. *FAO Fish. Tech. Pap.* ***101, suppl. 2***.

Sparre, P., Ursin, E. and Venema, S. C. (1989). Introduction to tropical fish stock assessment, part 1 – manual. *FAO Fish. Tech. Pap.* ***No. 306/1****: 337 pp*

Stager, C. J., R. E. Hecky, D. Grzesik , B. F. Cumming , H. Kling (2009). Diatom evidence for the timing and causes of eutrophication in Lake Victoria, East Africa. *Hydrobiologia,* ***636(1):*** *pp. 463-478*,

Strauss, R. E. (1979). Reliability estimates for Ivlev's electivity index, the forage ratio, and a proposed linear index of food selectivity. *Tran. Am. Fish. Soc.,* ***108:*** *344–352*.

Sumaila, U.R., Teh, L., Watson, R., Tyedmers, P., and Pauly, D. 2008. Fuel price increase, subsidies, overcapacity, and resource sustainability. ICES J. Mar. Sci. 65(6): 832–840

Swamee, P. K. (2002). Near Lognormal Distribution, *J. Hydrol. Eng. 7****(6):*** *441–444*

Talling, J. F. (1957). Some observations on the stratification of Lake Victoria. *Limnol. and Oceanog., 2: 213-221.*

Talling, J. F. (1966). The annual cycle of stratification and phytoplankton growth in Lake Victoria (East Africa). *Inernationale revue der Gesamten HydroBiologie Und Hydrography* ***51:*** *545 – 621.*

Talling, J. F. (1969). The incidence of vertical mixing and some biological and chemical consequences in tropical African lakes. *Verh Internat. Verein Limnol.,* ***17:*** *998-1012*.

Tamatamah, R. A., R. E. Hecky and H. C. Duthie (2004). The Atmospheric Deposition of Phosphorus in Lake Victoria.

Taylor, W. A. (2007). *Guide to Normality Testing and Transformations*.Taylor Enterprises, Inc., Libertyville, IL.

Thiam, S. E. (1986). Some improvements and corrections to Sim's version of ELEFAN I. *ICLARM Fishbyte,* ***4(3):*** *6-10*.

Thompson, W. F. and F. H. Bell (1934) Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per unit of gear. *Rep. Int. Fish. (Pacific Halibut) Comm.,* ***6:*** *108 pp*

Tumwebaze R. (2003). Hydro acoustic abundance assessment and population characteristics of *R. argentea* in Lake Victoria. Ph. D Thesis, University of Hull.

Tumwebaze, R., I. Cowx, S. Ridgway, A. Getabu, and D. N. MacLennan (2007). 'Spatial and temporal changes in the distribution of *Rastrineobola argentea* in Lake Victoria'. *Aqu. Ecos. Health and Manag.,* ***10(4):*** *398 – 406.*

vanden Bossche, J.-P., G. M. Bernacsek (1990). "Lake Victoria". *Source book for inland fishery resources of Africa. Volume 1*. CIFA Technical Paper No. 18/1. Food and Agricultural Organization. pp. 83–87.Vanderploeg, H. A. and D.

Scavia (1979). Two electivity indices for feeding with spatial reference to zooplankton grazing. *J. Fish. Res. Bd Can.,* ***36:*** *362-365.*

van Gills, J. A., I. W. Schenk, O. Bos and T. Piersma (2003). Incompletely informed shorebirds that face a digestive constraint maximize net energy gain when exploiting patches. *Am. Nat.,* ***161:*** *777–793*.

van Someren, V. D. (1959) A study of a small basket-trap fishery in Kenya. *E. Afr. Agr. For. J.,* ***24:*** *257 - 267.*

von Bertalanffy, L. 1938. A quantitative study of organic growth. *Hum. Biol.,* ***10:*** *181–123.*

Wade, P. R. (2001). Bayesian methods in conservation biology. *Conserv. Biol.,* ***14:*** *1308–1316*.

Wandera, S. B. (1990). The population exploitation of small pelagic fishes of the great lakes of Africa with reference to the mukene (*Rastrineobola argentea*) fishery of the northern waters of Lake Victoria. In**:** *Fisheries of the African Great Lakes. Research papers presented at the International Symposium on Resource Use and Conservation of the African Great Lakes. Bujumbura, 29 November - 2nd December 1989. International Agricultural Centre, Wagenigen, The Netherlands, Fisheries and Aquaculture Unit. Occasional Paper,* ***3.*** **p** 67-74

Wandera, S. B. (1992). A study of *Rastrineobola* in the Uganda lakes. **p 36-50**. In: Mannini, P. (ed). The Lake Victoria Dagaa (*Rastrineobola argentea*). Report of the First Meeting of the Working Group on Lake Victoria *Rastrineobola argentea*, 9-11 December 1992, Kisumu, Kenya. *UNDP/FAO Regional Project for Inland Fisheries Planning (IFIP) RAF/87/099-TD/38/92 (En).*

Wandera, S. B. (1999). The reproductive biology of *Rastrineobola argentea* (Pellegrin) in the Northern Waters of Lake Victoria. In: *Report of the FIDA WOG Workshop, I. G. Cowx and D. Tweddle (Eds.). Kisumu 16 – 20 August 1999. FIDAWOG (1999).*

Wanink, J. H. (1989). The ecology of dagaa, *Rastrineobola argentea* (Pellegrin) *1904*. *In:* Report of the Haplochromis Ecology Survey Team (HEST) operating in Lake Victoria. HEST/TAFIRI/FAO/DANIDA workshop on the fish stock in Lake Victoria. January/February 1989, Mwanza, Tanzania.

Wanink, J. H. (1995). Prospects for the fishery on the small pelagic Rastrineobola argentea in Lake Victoria. *Symposium on Lake Tanganyika Research, Kuopio, Finland.*

Wanink, J. H., E. F.B. Katunzi, P. Kees, C. Goudswaard, F. Witte and W. L. T. van Densen (2002). The shift to smaller zooplankton in Lake Victoria cannot be attributed to the 'sardine' *Rastrineobola argentea* (Cyprinidae). *Aquat. Living Resour.* ***15:*** *37–43*

Weibull, W. (1951) "A statistical distribution function of wide applicability" *J. Appl. Mech.-Trans. ASME* ***18(3):*** *293-297*.

Werbos, P. (1990). "Backpropagation through time: what it does and how to do it." *Proc. IEEE* ***78****(10)*: *121-125.*

Whitehead, P. J. P. (1958). Indigenous river fishing methods in Kenya. *E. Afr. Agr. For. J.,* ***24:*** *111.*

Whitehead, P. J. P. (1959). The river fisheries of Kenya. 1. Nyanza Province. *E. Afr. Agr. For. J.,* ***24:*** *274 - 278.*

Williams, R. and D. Zipser (1989). "A learning algorithm for continually running fully recurrent neural networks." *Neur. Comput., **1:** 270-280, 1989*.

Worthington, E. B. (1930). Observations on the temperature, hydrogen-ion concentration, and other physical conditions of the Victoria and Albert Nyanzas. *Inter Rev Ges Hydrobiol, **24:** 328–357*.

Yang J., H. Xu, L. Pan, P. Jia, F. Long, M. Ji, (2011). Task scheduling using Bayesian optimization algorithm for heterogeneous computing environments. *Appl, Soft Comput. **11(4):** 3297–3310.*

Zar, J. H. (1984). *Biostatistical analysis*. Prentice-Hall International, Inc., Englewood Cliffs, New Jersey. **pp** 718.

APPENDICES

Appendix I: Processing of data for modeling of stock-recruitment relationship (Source: Author)

SL mm	TL mm	Weight	Frequency	Males	Females	Mature	%age	Total Biomass	Mature Biomass	Fecundity	Egg production
20	23.9	0.12275	80	40	40	-	-	5	-	-	-
21	25.1	0.14320	60	30	30	-	-	4	-	-	-
22	26.2	0.16595	30	15	15	-	-	2	-	-	-
23	27.3	0.19113	30	15	15	-	-	3	-	-	-
24	28.4	0.21890	40	20	20	-	-	4	-	-	-
25	29.5	0.24940	580	290	290	-	-	72	-	-	-
26	30.6	0.28278	200	100	100	-	-	28	-	-	-
27	31.7	0.31920	270	135	135	-	-	43	-	-	-
28	32.8	0.35882	700	350	350	-	-	126	-	-	-
29	33.9	0.40178	570	285	285	-	-	115	-	-	-
30	35.0	0.44825	1600	800	800	-	-	359	-	-	-
31	36.2	0.49839	270	135	135	-	-	67	-	-	-
32	37.3	0.55236	900	450	450	-	-	249	-	-	-
33	38.4	0.61034	340	170	170	22.67	0.1333	104	14	276.65	6270.75
34	39.5	0.67248	480	240	240	66.67	0.2778	161	45	300.93	20062.13
35	40.6	0.73897	2160	1080	1080	475.20	0.4400	798	351	326.58	155191.64
36	41.7	0.80996	750	375	375	225.00	0.6000	304	182	353.64	79568.21
37	42.8	0.88563	1100	550	550	330.00	0.6000	487	292	382.13	126103.76
38	43.9	0.96617	2680	1340	1340	786.52	0.5870	1295	760	412.11	324130.80
39	45.0	1.05174	3100	1550	1550	1048.53	0.6765	1630	1103	443.59	465122.18
40	46.1	1.14252	13240	6620	6620	5443.11	0.8222	7563	6219	476.63	2594369.21
41	47.3	1.23870	5650	2825	2825	2340.71	0.8286	3499	2899	511.26	1196711.44
42	48.4	1.34046	8950	4475	4475	4147.56	0.9268	5999	5560	547.51	2270821.68
43	49.5	1.44798	7850	3925	3925	3700.71	0.9429	5683	5359	585.42	2166456.05
44	50.6	1.56145	5540	2770	2770	2620.27	0.9459	4325	4091	625.02	1637718.72

Appendix I (Contd.): Processing of data for modeling of stock-recruitment relationship (Source: Author)

SL mm	TL mm	Weight	Frequency	Males	Females	Mature	%age	Total Biomass	Mature Biomass	Fecundity	Egg production
45	51.7	1.68106	9820	4910	4910	4795.81	0.9767	8254	8062	666.35	3195709.45
46	52.8	1.80699	3560	1780	1780	1780.00	1.0000	3216	3216	709.46	1262832.54
47	53.9	1.93944	2800	1400	1400	1400.00	1.0000	2715	2715	754.36	1056107.81
48	55.0	2.07860	2600	1300	1300	1300.00	1.0000	2702	2702	801.11	1041441.24
49	56.1	2.22466	1810	905	905	905.00	1.0000	2013	2013	849.73	769005.85
50	57.2	2.37782	4420	2210	2210	2210.00	1.0000	5255	5255	900.26	1989582.07
51	58.4	2.53828	830	415	415	415.00	1.0000	1053	1053	952.74	395388.77
52	59.5	2.70623	920	460	460	460.00	1.0000	1245	1245	1007.21	463315.67
53	60.6	2.88188	640	320	320	320.00	1.0000	922	922	1063.69	340381.16
54	61.7	3.06543	370	185	185	185.00	1.0000	567	567	1122.23	207612.40
55	62.8	3.25707	730	365	365	365.00	1.0000	1189	1189	1182.86	431743.17
56	63.9	3.45703	140	70	70	70.00	1.0000	242	242	1245.61	87192.93
57	65.0	3.66549	30	15	15	15.00	1.0000	55	55	1310.53	19657.96
58	66.1	3.88268	50	25	25	25.00	1.0000	97	97	1377.65	34441.15
59	67.2	4.10880	20	10	10	10.00	1.0000	41	41	1446.99	14469.94
60	68.3	4.34405	40	20	20	20.00	1.0000	87	87	1518.61	30372.24
61	69.5	4.58866	30	15	15	15.00	1.0000	69	69	1592.53	23888.02
62	70.6	4.84283	20	10	10	10.00	1.0000	48	48	1668.80	16687.97
63	71.7	5.10678	20	10	10	10.00	1.0000	51	51	1747.44	17474.35
64	72.8	5.38073	10	5	5	5.00	1.0000	27	27	1828.48	9142.42
65	73.9	5.66489	20	10	10	10.00	1.0000	57	57	1911.98	19119.80
66	75.0	5.95948	10	5	5	5.00	1.0000	30	30	1997.96	9989.79
67	76.1	6.26472	10	5	5	5.00	1.0000	31	31	2086.45	10432.27
68	77.2	6.58084	10	5	5	5.00	1.0000	33	33	2177.50	10887.51
69	78.3	6.90805	10	5	5	5.00	1.0000	35	35	2271.14	11355.69
70	79.4	7.24659	10	5	5	5.00	1.0000	36	36	2367.40	11836.99

Appendix II: **Raw data used in the ANN modeling. Class is based on Manyala (2006), fisheries data (GoK, 2008) and environmental data (Mwirigi *et al*., 2005)**

Year	Class	*Rastrineobola* (mt)	*Bagrus* (mt)	*Clarias* (mt)	*Haplochromis* (mt)	*Lates* (mt)	*Mormyrus* (mt)	*Protopterus* (mt)	*Oreochromis* (mt)	Discharge (cusecs)	Rainfall (mm)	Evaporation (mm)	Outflow (cusecs)	Level (m a.s.l)
1968	Pristine	731	1147	1756	3743	1	53	2808	2141	1213.43	2226.06	1525.83	1383.80	1135.66
1969	Pristine	520	966	1354	6427	17	73	1626	3951	744.41	1732.64	1523.63	1462.70	1135.94
1970	Pristine	524	1091	1592	5357	28	82	1629	3686	837.07	1682.28	1519.04	1403.13	1135.83
1972	Growth	1255	856	2729	4644	38	78	1915	1480	657.68	1719.92	1527.47	1183.45	1135.17
1973	Growth	1768	1563	2885	5451	246	182	1841	792	650.53	1679.03	1561.87	1224.71	1135.18
1974	Growth	3742	1103	2913	6013	136	89	2750	468	678.53	1807.13	1626.94	1117.22	1135.14
1975	Growth	5448	1389	2989	4620	51	58	1935	230	672.00	1736.21	1521.15	1110.31	1135.02
1976	Growth	5652	1025	2686	6368	97	89	941	470	507.89	1578.19	1567.86	1105.67	1134.88
1977	Growth	6704	1141	1755	5378	203	102	773	507	972.62	1813.89	1487.77	1082.77	1134.86
1978	Growth	8710	183	2047	6621	1066	132	653	2521	1160.32	2073.59	1497.01	1297.34	1135.31
1979	Growth	9321	1769	3205	6599	4286	359	472	1056	966.66	1763.24	1542.83	1502.75	1135.61
1980	Growth	9443	642	1223	3636	4310	333	370	1274	586.44	1343.19	1556.98	1303.81	1135.14
1981	Dominance	7635	430	1328	916	22836	209	323	1997	820.20	1479.64	1489.38	1079.22	1134.82
1982	Dominance	10419	2532	2062	2546	33134	2678	239	2980	929.20	1708.66	1477.66	1068.94	1134.70
1983	Dominance	16444	1243	1336	612	55572	218	374	2904	768.72	1477.55	1481.64	1115.33	1134.79
1984	Dominance	19437	88	877	41	44698	89	95	6235	520.16	1531.91	1614.39	1072.44	1134.50
1985	Dominance	25866	61	590	6	53011	49	179	7615	892.74	1600.35	1543.62	934.16	1134.49
1986	Dominance	34518	62	1697	3	58806	51	216	7853	654.67	1674.03	1550.59	888.84	1134.37
1987	Dominance	33145	40	345	183	68545	12	58	9027	704.52	1720.84	1524.55	961.19	1134.53
1988	Dominance	40861	75	300	1338	61210	300	25	16347	981.96	1952.18	1437.08	1037.78	1134.74

Appendix II (Contd.): Raw data used in the ANN modeling. Class is based on Manyala (2006), fisheries data (GoK, 2008) and environmental data (Mwirigi *et al*., 2005)

Year	Class	*Rastrineobola* (mt)	*Bagrus* (mt)	*Clarias* (mt)	*Haplochromis* (mt)	*Lates* (mt)	*Mormyrus* (mt)	*Protopterus* (mt)	*Oreochromis* (mt)	Discharge (cusecs)	Rainfall (mm)	Evaporation (mm)	Outflow (cusecs)	Level (m a.s.l)
1989	Dominance	45464		403	4759	56810	403	24	13101	797.94	2100.62	1510.00	1172.36	1135.01
1990	Dominance	46738	134	507	1	71514	578	84	38305	1138.68	1675.68	1572.36	1085.70	1135.34
1991	Decline	58098	174	2115	3615	51262	444	123	27475	783.40	1779.84	1514.87	1178.85	1135.23
1992	Decline	35414	78	589	3018	77599	175	1544	16769	675.76	1340.18	1564.89	1048.42	1134.83
1993	Decline	42505	34	264	3506	100037	102	146	12670	635.36	1443.84	1576.78	1009.27	1134.55
1994	Decline	69134	2	263	4196	103995	150	202	11821	926.55	1721.54	1478.99	822.58	1134.31
1995	Decline	56827	127	234	4822	102546	141	408	12363	880.05	1819.18	1455.52	924.27	1134.56
1996	Decline	49670	157	405	3914	97145	113	119	10903	784.19	1745.49	1463.87	975.93	1134.76
1997	Decline	40318	206	2049	2454	73549	53	1704	13953	1409.88	1943.97	1504.91	925.02	1134.63
1998	Decline	42336	324	2586	2577	77967	57	1895	14652	1543.41	1795.63	1550.53	1224.66	1135.58
1999	Decline	40168	57	1200	528	115324	4	776	23701	682.15	1753.21	1547.13	1158.38	1135.38
2000	Decline	38968	60	1070	527	109221	4	733	23226	552.86	1320.17	1581.8	1118.56	1134.86
2001	Collapse	49165	88	2063	1198	78939	21	1854	7292	771.60	1809.65	1481.13	1096.34	1134.70
2002	Collapse	35455	57	1874	1029	59007	2	1178	16251	850.10	1813.21	1557.31	1178.93	1134.71
2003	Collapse	31659	63	1545	1020	55175	14	867	15982	750.26	1775.91	1524.91	1203.17	1134.63
2004	Collapse	34679	88	1710	1066	61440	14	854	18121	548.13	1713.58	1560.18	1329.33	1134.38
2005	Collapse	54019	69	1353	4832	52368	9	777	22231	850.10	1813.21	1557.31	1178.93	1134.71
2006	Collapse	57929	88	4387	5198	55706	23	2914	19038	750.26	1775.91	1524.91	1203.17	1134.63
2007	Collapse	49472	150	2092	5690	47067	29	3146	13090	548.13	1713.58	1560.18	1329.33	1134.38

Appendix III: Pre-processed data – catch (Source: Author)

Pristine	Growth	Dominance	Decline	Collapse	*Bagrus*	*Clarias*	*Haplochromis*	*Lates*	*Mormyrus*	*Protopterus*	*Oreochromis*	*Rastrineobola*
1	-1	-1	-1	-1	0.28874	-0.26704	0.13025	-1.00000	-0.76923	0.78347	-1.00000	-0.99385
1	-1	-1	-1	-1	0.08176	-0.46063	0.94137	-0.99972	-0.67873	0.02627	-0.99782	-1.00000
1	-1	-1	-1	-1	0.22470	-0.34602	0.61801	-0.99953	-0.63801	0.02819	-0.99417	-0.99988
-1	1	-1	-1	-1	-0.04403	0.20154	0.40254	-0.99936	-0.65611	0.21140	-0.99053	-0.97858
-1	1	-1	-1	-1	0.76444	0.27667	0.64642	-0.99575	-0.18552	0.16400	-0.96591	-0.96362
-1	1	-1	-1	-1	0.23842	0.29015	0.81626	-0.99766	-0.60634	0.74632	-0.97152	-0.90608
-1	1	-1	-1	-1	0.56547	0.32675	0.39529	-0.99913	-0.74661	0.22422	-0.98674	-0.85636
-1	1	-1	-1	-1	0.14923	0.18083	0.92354	-0.99834	-0.60634	-0.41256	-0.97079	-0.85041
-1	1	-1	-1	-1	0.28188	-0.26752	0.62436	-0.99650	-0.54751	-0.52018	-0.96758	-0.81975
-1	1	-1	-1	-1	-0.81361	-0.12690	1.00000	-0.98153	-0.41177	-0.59705	-0.93065	-0.76127
-1	1	-1	-1	-1	1.00000	0.43077	0.99335	-0.92569	0.61539	-0.71300	-0.93138	-0.74346
-1	1	-1	-1	-1	-0.28874	-0.52372	0.09792	-0.92527	0.49774	-0.77835	-0.91521	-0.73991
-1	-1	1	-1	-1	-0.53116	-0.47315	-0.72409	-0.60398	-0.06335	-0.80846	-0.86611	-0.79261
-1	-1	1	-1	-1	0.39851	-0.46930	-0.81596	-0.03626	-0.02262	-0.77579	-0.79778	-0.53584
-1	-1	1	-1	-1	-0.92224	-0.69034	-0.98852	-0.22484	-0.60634	-0.95452	-0.55447	-0.44860
-1	-1	1	-1	-1	-0.95312	-0.82856	-0.99909	-0.08067	-0.78733	-0.90071	-0.44979	-0.26120
-1	-1	1	-1	-1	-0.95197	-0.29545	-1.00000	0.01983	-0.77828	-0.87700	-0.42939	-0.00901
-1	-1	1	-1	-1	-0.97713	-0.94655	-0.94560	0.18873	-0.95475	-0.97822	-0.34409	-0.04903
-1	-1	1	-1	-1	-0.93711	-0.96822	-0.59656	0.06152	0.34842	-0.99936	0.18390	0.17588
-1	-1	1	-1	-1	-0.90852	-0.91861	0.43729	-0.01479	0.81448	-1.00000	-0.04710	0.31005
-1	-1	-1	1	-1	-0.82390	-0.09415	0.09157	-0.11100	1.00000	-0.93658	1.00000	0.67832
-1	-1	-1	1	-1	-0.93368	-0.82904	-0.08885	0.34575	-0.21720	-0.02627	0.22011	0.01711

Appendix III (Contd.): Pre-processed data – catch (Source: Author)

Pristine	Growth	Dominance	Decline	Collapse	*Bagrus*	*Clarias*	*Haplochromis*	*Lates*	*Mormyrus*	*Protopterus*	*Oreochromis*	*Rastrineobola*
-1	-1	-1	1	-1	-0.98399	-0.98555	0.05863	0.73488	-0.54751	-0.92185	-0.07849	0.22380
-1	-1	-1	1	-1	-1.00000	-0.98603	0.26715	0.80353	-0.33032	-0.88597	-0.14034	1.00000
-1	-1	-1	1	-1	-0.87764	-1.00000	0.45633	0.77840	-0.37104	-0.75400	-0.10086	0.64127
-1	-1	-1	1	-1	-0.84334	-0.91765	0.18193	0.68473	-0.49774	-0.93914	-0.20721	0.43265
-1	-1	-1	1	-1	-0.78731	-0.12593	-0.25929	0.27551	-0.76923	0.07623	0.01497	0.16006
-1	-1	-1	1	-1	-0.65237	0.13268	-0.22212	0.35213	-0.75113	0.19859	0.06589	0.21888
-1	-1	-1	1	-1	-0.95769	-0.53479	-0.84134	1.00000	-0.99095	-0.51826	0.72508	0.15568
-1	-1	-1	1	-1	-0.95426	-0.59740	-0.84164	0.89416	-0.99095	-0.54580	0.69048	0.12070
-1	-1	-1	-1	1	-0.92224	-0.11919	-0.63886	0.36899	-0.91403	0.17233	-0.47026	0.41793
-1	-1	-1	-1	1	-0.95769	-0.21021	-0.68994	0.02332	-1.00000	-0.26073	0.18237	0.01831
-1	-1	-1	-1	1	-0.95083	-0.36865	-0.69266	-0.04314	-0.94570	-0.45996	0.16278	-0.09234
-1	-1	-1	-1	1	-0.94397	-0.46111	0.45935	-0.09182	-0.96833	-0.51762	0.61799	0.55942
-1	-1	-1	-1	1	-0.92224	1.00000	0.56996	-0.03393	-0.90498	0.85138	0.38539	0.67339
-1	-1	-1	-1	1	-0.85134	-0.10523	0.71865	-0.18375	-0.87783	1.00000	-0.04790	0.42688

Appendix IV: Pre-processed data – environmental

Pristine	Growth	Dominance	Decline	Collapse	Discharge (cusecs)	Rainfall (mm)	Evaporation (mm)	Outflow (cusecs)	Level (m a.s.l)	*Rastrineobola* (mt)
1	-1	-1	-1	-1	0.36268	1.00000	-0.06510	0.65024	0.65644	-0.99385
1	-1	-1	-1	-1	-0.54319	-0.08936	-0.08828	0.88224	1.00000	-1.00000
1	-1	-1	-1	-1	-0.36422	-0.20054	-0.13663	0.70707	0.86503	-0.99988
-1	1	-1	-1	-1	-0.71070	-0.11744	-0.04783	0.06112	0.05522	-0.97858
-1	1	-1	-1	-1	-0.72451	-0.20772	0.31455	0.18244	0.06749	-0.96362
-1	1	-1	-1	-1	-0.67043	0.07510	1.00000	-0.13363	0.01841	-0.90608
-1	1	-1	-1	-1	-0.68304	-0.08148	-0.11440	-0.15395	-0.12883	-0.85636
-1	1	-1	-1	-1	-1.00000	-0.43035	0.37765	-0.16759	-0.30061	-0.85041
-1	1	-1	-1	-1	-0.10242	0.09002	-0.46603	-0.23493	-0.32515	-0.81975
-1	1	-1	-1	-1	0.26010	0.66338	-0.36869	0.39600	0.22699	-0.76127
-1	1	-1	-1	-1	-0.11393	-0.02180	0.11398	1.00000	0.59509	-0.74346
-1	1	-1	-1	-1	-0.84829	-0.94918	0.26304	0.41503	0.01841	-0.73991
-1	-1	1	-1	-1	-0.39681	-0.64793	-0.44907	-0.24537	-0.37423	-0.79261
-1	-1	1	-1	-1	-0.49623	-0.65254	-0.53060	-0.13919	-0.41104	-0.53584
-1	-1	1	-1	-1	-0.97630	-0.53253	0.86780	-0.26530	-0.76687	-0.44860
-1	-1	1	-1	-1	-0.25670	-0.38143	0.12230	-0.67191	-0.77914	-0.26120

Appendix IV: Pre-processed data – environmental

Pristine	Growth	Dominance	Decline	Collapse	Discharge (cusecs)	Rainfall (mm)	Evaporation (mm)	Outflow (cusecs)	Level (m a.s.l)	*Rastrineobola* (mt)
-1	-1	1	-1	-1	-0.71651	-0.21876	0.19572	-0.80517	-0.92638	-0.00901
-1	-1	1	-1	-1	-0.62023	-0.11541	-0.07858	-0.59243	-0.73006	-0.04903
-1	-1	1	-1	-1	-0.08438	0.39534	-1.00000	-0.36722	-0.47239	0.17588
-1	-1	1	-1	-1	-0.43980	0.72306	-0.23186	0.02851	-0.14110	0.31005
-1	-1	-1	1	-1	-0.46788	0.01485	-0.18055	0.04759	0.12883	0.67832
-1	-1	-1	1	-1	-0.67578	-0.95582	0.34636	-0.33593	-0.36196	0.01711
-1	-1	-1	1	-1	-0.75381	-0.72697	0.47161	-0.45105	-0.70552	0.22380
-1	-1	-1	1	-1	-0.19140	-0.11387	-0.55852	-1.00000	-1.00000	1.00000
-1	-1	-1	1	-1	-0.28121	0.10170	-0.80575	-0.70099	-0.69325	0.64127
-1	-1	-1	1	-1	-0.46636	-0.06099	-0.71779	-0.54908	-0.44785	0.43265
-1	-1	-1	1	-1	0.74210	0.37721	-0.28547	-0.69878	-0.60736	0.16006
-1	-1	-1	1	-1	1.00000	0.04971	0.19509	0.18229	0.55828	0.21888
-1	-1	-1	1	-1	-0.66344	-0.04395	0.15928	-0.01260	0.31288	0.15568
-1	-1	-1	1	-1	-0.91315	-1.00000	0.52449	-0.12969	-0.32515	0.12070
-1	-1	-1	-1	1	-0.49067	0.08066	-0.53597	-0.19503	-0.52147	0.41793
-1	-1	-1	-1	1	-0.33906	0.08852	0.26651	0.04783	-0.50920	0.01831
-1	-1	-1	-1	1	-0.53189	0.00617	-0.07479	0.11910	-0.60736	-0.09234
-1	-1	-1	-1	1	-0.33906	0.08852	0.26651	0.04783	-0.50920	0.55942
-1	-1	-1	-1	1	-0.53189	0.00617	-0.07479	0.11910	-0.60736	0.67339
-1	-1	-1	-1	1	-0.92228	-0.13144	0.29675	0.49007	-0.91411	0.42688