THE UNIVERSITY OF THE SOUTH PACIFIC
LIBRARY

Author Statement of Accessibility

Name of Candidate : Sandeep Singh

Degree : MSc in Mathematics

Department/School : FSTE, School of Computing, Information and Mathematical Sciences

Institution/University : USP

Title of Thesis : Developing a Spatially Adjusted CPUE for Albacore fishery in Fiji

Date of award : ____________
(month, year)

This thesis may be:

1. consulted in the Library without the author's permission **YES** ☑ **NO** ☐
2. cited without the author's permission provided it is properly acknowledged **YES** ☑ **NO** ☐
3. photocopied in whole without the author's written permission **YES** ☑ **NO** ☐
   If you answered '**No**', select the percentage that may be photocopied

Under 20% ☑ **20-40%** ☐ **40-60%** ☐ **60-80%** ☐ **Over 80%** ☐

I authorize The University of the South Pacific to:

4. produce a microfilm or microfiche copy for preservation purposes **YES** ☑ **NO** ☐
5. retain a copy in electronic format for archival and preservation purposes **YES** ☑ **NO** ☐
6. make this thesis available on the Internet
   YES ☑
   NO ☐ **Go to Question 7**
7. make this thesis available on the USP Intranet
   YES ☑
   NO ☐ **Give reasons** ____________

Signed : S Singh Date : 05/04/18

Contact Information

Email address : san_jaswani@yahoo.com

Phone
Mobile : 9333588
Home : 3383538
Work : 3220428

Permanent Residential Address

9 Bureta St,
Tamavua,
Suva

DEVELOPING A SPATIALLY ADJUSTED CATCH PER UNIT EFFORT (CPUE) INDEX FOR THE ALBACORE LONGLINE FISHERY IN FIJI

by

Sandeep Singh

A thesis submitted in partial fulfilment of the
requirements for the degree of
Master of Science in Mathematics

School of Computing, Information and Mathematical Sciences
Faculty of Science, Technology and Environment
The University of the South Pacific

July, 2017

Declaration of Originality

Statement by Author:

I hereby declare that the work presented in this thesis, which I now submit for assessment of the award of Master of Science is entirely my own work and has not been taken from the work of others; except in citations and references which have been acknowledged and results in this thesis have not been previously submitted for a University degree either in whole or in part elsewhere.

……………………
Sandeep Singh
S99007579
July 2017

Statement by Supervisors:

The research work in this thesis was accomplished under my supervision and to the best of my knowledge is the sole work of Mr Sandeep Singh.

……………………..
Dr. M.G.M. Khan (Principal Supervisor)
Associate Professor of Statistics and Acting Head of School
School of Computing, Information and Mathematical Sciences
Faculty of Science, Technology and Environment
The University of the South Pacific
Suva, Fiji

………………………
Dr. Cara Miller (Co- Supervisor)
Former Lecturer
School of Biological & Chemical Sciences
Faculty of Science, Technology and Environment
The University of the South Pacific
Suva, Fiji

Acknowledgements

I would like to thank the following people for their invaluable assistance in compiling this thesis:

- Dr M.G.M. Khan – My Principal Supervisor, for his guidance and support rendered throughout the research write-up as well as for his valuable suggestions, motivation and assistance in every phase of the research
- Dr. Cara Miller – My Co-Supervisor, for her guidance and effort put in getting the fisheries data as well as for her valuable assistance towards the clarification of fisheries concepts
- Dr. Simon Nicol – Former Fisheries Scientist, SPC Noumea for his guidance and initiation of this study
- The Fisheries Department of Fiji – for providing the secondary data that was used in this research
- Dr Graham Pilling – Fisheries Scientist, SPC Noumea, for his time, assistance and guidance towards fisheries technical concepts
- Dr Laura Tremblay-Boyer - Fisheries Scientist, SPC Noumea, for her assistance and guidance towards fisheries technical concepts
- Dr Sam McKechnie - Fisheries Scientist, SPC Noumea, for his assistance and guidance towards fisheries technical concepts
- Mrs Sampreet Kaur Singh and Mannat Kaur – my wife and daughter, for their motivation and being a pillar of strength for the research.
- My sincere gratitude goes to all those who have assisted in this research in one way or the other.

Table of Contents

List of Tables

List of Figures

Abbreviations

AIC - Akaike Information Criteria

ANOVA – Analysis of Variance

CLARA – Clustering Large Applications

CPUE – Catch Per Unit Effort

EEZ – Economic Exclusive Zone

FSI – Fishery Survey of India

GAM – Generalized Additive Model

GLM – Generalized Linear Model

GLMm – Generalized Linear Mixed Model

HBF – Hooks between Floats

MLR – Multiple Linear Regression

NARA – National Aquatic Resources and Research and development Agency

PAM – Partitioning Around Medoids

Q-Q Plot – Quantile-Quantile Plot

RTM – Regression Tree Model

SEAPODYM – Spatial Ecosystem and Population Dynamics Model

SOI – Southern Oscillation Index

SPC – Secretariat of Pacific Community

SST – Sea Surface Temperature

TDR – Time Depth Recorders

WCFPC – Western and Central Pacific Fisheries Commission

Abstract

The primary purpose of catch per unit effort (CPUE) standardization is to remove the variation in data not attributed to changes in fish stock abundance. There has been CPUE standardization of albacore tuna done in many parts of the world including the South Pacific by many authors. However, there has been no CPUE standardization done for albacore tuna specifically for Fiji. In this thesis, a primary task is to develop a spatially adjusted CPUE index for albacore longline fishery in Fiji. The data used in the analysis was provided by the Fiji Department of Fisheries. Clustering techniques were used to identify the local fishing vessels that targeted albacore to ensure that only the effort that targeted albacore were considered for the CPUE standardization. The most common method to standardize CPUE is the use of Generalized Linear Model (GLM). These models allow estimating the catch rate (response variable) with a linear combination of a set of explanatory variables (see Bernasconi et al., 2015). The GLM was used to standardize the albacore tuna CPUE for longline fishery in Fiji's EEZ. AIC, pseudo R^2 and ANOVA were used to select the factors for CPUE standardization. The selected model included year-quarter, month and vessel id as the main factors and also the interaction among latitude and longitude aggregated at 1 degree resolution. Diagnostic plots were done to examine the selected model. Box plot was done to identify the outliers in the fitted values. Residuals of the fitted model were normally distributed. The analysis found that all the explanatory variables in the selected model were highly significant. Trends in nominal CPUE and standardized CPUE showed fluctuations and in majority of the cases, the standardized CPUE values were lower.

Preface

This thesis entitled “Developing a Spatially Adjusted Catch Per Unit Effort Index for the Albacore Longline Fishery in Fiji” is submitted to The University of the South Pacific, Suva, Fiji to supplicate the Master of Science in Mathematics.

Generalized Linear Models are widely used to standardize the CPUE in fisheries management. Albacore is one of the common tuna species that are fished by longline vessels in Fiji. Standardized CPUE of albacore tuna will assist in better management of the stock and setting allowable harvest limits to ensure there is sustainable albacore stock. The time series data used was from 2003 to 2014.

In this thesis, we developed a spatially adjusted CPUE index for albacore based on the fisheries data provided by the Fisheries Department of Fiji for the domestic longline fleets. There are two kinds of problems considered which are mainly required in CPUE standardization. The first is whether the domestic longline fishing vessels could be identified that target albacore in Fiji’s EEZ defined by the coordinates 12^o to 20^o south, 177^o east and 177^o west. This is discussed in Chapter 2. The second problem is using GLM to standardize CPUE which is discussed in Chapter 3.

This thesis consists of four chapters. **Chapter 1** provides an introduction of tuna fisheries in the Pacific region as well as the problem of data management in Pacific tuna fisheries, analysis of CPUE, aims and objectives of the research, description of fisheries data, methodology, hierarchical and partitioning clustering. GLMs and a brief review of literature and studies on the problem are also presented in this chapter.

In **Chapter 2**, the problem of identifying the vessels that target albacore tuna was solved using hierarchical and partitioning clustering techniques. The catch and effort data was analysed as well as tuna catch composition from 2003 to 2014. Clustering was done at an interval of 3 years from 2003 to 2014 as done by Molina et al. (2011) for domestic vessels. R software was used to produce the dendogram and the clara function was used to identify the vessels targeting albacore tuna (see Molina et al. 2011). Maps were plotted to show the spatial distribution of fishing sets that targeted albacore tuna.

Chapter 3 begins with the discussion on nominal CPUE and shows how the problem of standardizing CPUE is solved using GLM. GLMs have a common algorithm for the estimation of parameters by maximum likelihood (see Nelder and McCullagh, 1989). Standardization of CPUE using GLMs is a common procedure that removes confounding effects of variables other than abundance (see Bentley et al., 2012). There were different GLMs tested and the model was selected based on the low AIC value (see Crawley, 2015) and R^2 value (see Hoyle et al., 2014). There were appropriate fit and diagnostic residual plots evaluated to check the underlying assumptions on the error distribution and selecting the final GLM (see Bernasconi et al., 2015). The selected model was investigated using Q-Q plot and residual plots.

Finally, a conclusion to this research work is provided in **Chapter 4**. A comprehensive list of references is presented in **Bibliography** at the end of the thesis. The **Appendix** is also presented at the end of the thesis and it includes the various tables and R codes.

Chapter 1

Introduction

1.1 Overview of Tuna fisheries in the Pacific Region

Tuna is a significant natural resource in the Pacific Islands. It not only provides an important source of food for Pacific Island people, but it is part of their cultural heritage. Today, the tuna industry is also an important source of income and employment for Pacific Island Countries, including Fiji (see SPC, 2010). The industry has contributed significantly to Fiji's economic growth through a number of avenues including exports, license fees and employment opportunities.

There are four types of tuna species in the Pacific, namely albacore tuna (*Thunnus alalunga*), skipjack tuna (*Katsuwonus pelamis*), bigeye tuna (*Thunnus obesus*), and yellowfin (*Thunnus albacares*) that are most commonly targeted in the fishing industry. Fishing across the vast Pacific Ocean is done by both domestic and distant water fleets mainly from Japan, Korea, Taiwan and the United States of America (USA). The tuna fishing methods include longline, purse seine, pole and line and surface trolling. Longlines are mostly used to catch bigeye, albacore and yellowfin tuna in tropical waters (see Joseph, 2003). The key tuna fishery in Fijian waters is albacore tuna which is preferentially caught from longline vessels. The longline fishing method involves deploying a main line from a reel with baited hooks on branch lines and floats at regular intervals to ensure that the bait is set at pre-defined depths (see Moffitt, 2008).

1.2 The Problem of Data Management in Pacific Tuna Fisheries

The management of regional tuna fisheries is undertaken by the Western and Central Pacific Fisheries Commission (WCPFC) across the Western and Central areas of the Pacific Ocean. They operate fisheries ranging from small-scale artisanal operations in the coastal waters of Pacific states, to a large-scale, namely industrial purse-seine, pole-and-line and longline operations in both the exclusive economic zones of Pacific states and on the high seas. These fisheries mainly target skipjack tuna, yellowfin tuna, bigeye tuna and albacore tuna (see SPC, 2010). Thus, a huge amount of data is gathered from these fisheries.

Statistical analysis of these data for the South Pacific region including stock assessments are primarily implemented by the Secretariat of the Pacific Community (SPC). The main focus and application of analysis carried out by SPC are to provide region-wide management advice. A key metric that is used within these regional analyses is related to the efficiency and number of fish that are caught in a given area, that is, CPUE. The calculated CPUE for the region or sub-region is then applied on a national basis to come up with an appropriate fishing quota for the coming years. However, there has been no country specific research done on spatially adjusted CPUE index for the albacore longline fishery in Fiji.

1.3 Analysis of Catch Per Unit Effort (CPUE)

The sustainable management of albacore tuna in Fiji requires a detailed understanding of both fishing efficiency and stock status for this species. A key piece of information required for the estimation of these details is the calculation of a measure of how much fishing effort is required to catch a given volume of fish. To date, Fiji fisheries managers have largely relied on regional information to then make informed decisions about the albacore fishery in Fiji. This approach has provided a useful management framework. However, it would seem of added benefit if Fiji-specific information might be available and statistically validated to then consider within an assessment of the status of Fiji albacore.

1.4 Aims and Objectives

The aim of the study carried out in this thesis is to develop a spatially adjusted CPUE for the albacore fishery in Fiji. The standardized CPUE for Fiji will provide the index to manage albacore stock effectively and sustainably.

The objectives of the study are:

(i) To identify domestic fleets that target albacore tuna in Fiji by using cluster analysis.

(ii) To develop a CPUE index for longline albacore fishery in Fiji using GLM.

1.5 Fisheries Data

Standardized fisheries data for the Pacific Islands region is collected by observers within the regional observer program. Onboard observers cover a given number of trips in the various fisheries across the region and then provide this information back to SPC and the relevant national governments.

Extensive data is collected on all aspects of fisheries through the observer program. The broad categories in which data is collected include fishing sets, gear, catch, effort, trips, set haul logs and vessel. In-depth information is captured including the date and time of the set and catch, hook between float, hooks in basket, hooks set, float length, vessel speed, fishing gear material, bait information, hook number, species code, condition, length, weight estimate, latitude, longitude, economic exclusive zone code, wind direction, wind knots, vessel id, depart date and time, staff code, return date and time and country code. The fisheries data was available for each set.

1.6 Methodology

The data used in the study was provided by the Fisheries Department of Fiji for the domestic longline fleets from 2000 to 2015. There was information on 3706 sets corresponding to 378 fishing trips and 81 vessels. The data was available for individual fishing set and there were some missing values that were taken into consideration in data preparation.

The study mainly utilizes two statistical techniques; namely cluster analysis and GLMs. The catch and effort data was analysed using cluster analysis on species composition for domestic fleets targeting albacore. After identifying the domestic fleets targeting albacore, GLM was used to standardise CPUE. The study focused from 2003 to 2014 because of completeness of data. The information was reduced to 3655 sets corresponding to 372 fishing trips and 79 vessels.

Nominal CPUE was expressed as albacore catch divided by hook set (effort). Also nominal CPUE per 1000 hooks were used in the comparison with standardized CPUE. The variables included in the explanatory analysis using GLM were year, catch quarter, albacore catch in numbers, hook-set, year-quarter, vessel id, latitude, longitude, catch-month with the two interactions of latitude and longitude aggregated by 1-degree resolution, month and cubic value of latitude.

1.7 Catch Per Unit Effort Standardization: Review of Literature and Studies

1.7.1 Clustering Methods

Cluster is a set of objects that are similar to each other. According to Romesburg (1984), the purpose of cluster analysis is to make classifications based on the similarity among the sets into homogeneous groups. Cohen (1988) also stated that cluster analyses can differentiate between what is common among the variables in relation to the group.

Many clustering methods such as Hierarchical clustering, Partitioning clustering and Ward's method exist in literature (see Cohen, 1988 & Everitt, 1974). The clustering procedure to use depends on the number of cases and types of variables to be used for forming clusters.

Clustering methods is mainly divided into 2 basic types: hierarchical and partitional clustering.

Hierarchical clustering determines the desired clusters by organizing the objects in a dendogram (see Langfelder et al., 2008). Partitioning clustering methods are based on iterative relocation of data points between clusters (see Ayramo and Karkkainen, 2006).

1.7.1.1 Hierarchical Clustering

One of the most common techniques in clustering is the hierarchical cluster analysis. Hierarchical clustering techniques could be subdivided into agglomerative and divisive methods. The agglomerative methods work by a series of successful fusions of N entities into group and the divisive methods partition the set of N entities successfully into finer partitions. The results of both the methods can be presented by a dendogram. A dendogram is a two - dimensional diagram showing the fusions or partitions that are made at each successive level (see Everitt, 1974). The y-axis in the dendogram shows the distance at which the clusters merge and the objects are placed in the x-axis so that the clusters do not mix.

Both the agglomerative and divisive methods finally split the entire set of data into N groups with each containing a single entity. The user decides on the correct number of clusters and at which stage to stop in the analysis.

Agglomerative hierarchical techniques fuse individuals or groups of individuals which are closest or most similar at a particular stage. This technique includes different methods with the basic procedure being similar. In the nearest neighbour or single link method, the groups with the smallest distance are being fused together with each fusion decreasing the number of groups by one. The furthest neighbour or complete linkage method is the opposite of single link method where the maximum distance between groups is used to fuse them together. In the centroid cluster analysis, groups are fused according to the distance between their centroids and the groups with the smallest distances are fused first. In the median cluster analysis, groups are fused according to the distance between their medians and the groups with the smallest distances are fused first. In the group average method, groups are fused together based on the average distance between all pairs of individuals in the two groups. In Ward's method, the union of every pair of clusters at each stage is considered and the two clusters whose fusion results in the minimum increase in the error sum of squares are combined.

In the divisive method, the first task is to split the initial set of individuals into 2. A set of *n* individuals can be divided into two sub-sets in $2^{n-1} - 1$ ways which may be difficult to compute with even small data sets. There is a need to impose a restriction on the number of ways considered for moderately large data sets. Divisive techniques can have two common families; monothetic which are based on the possession or otherwise of a single specified attribute and polythetic methods are based on the values taken by all the attributes. Monothetic techniques are usually used in cases where the data is binary (see Everitt, 1974).

1.7.1.2 Partitioning Clustering

Partitioning clustering works by forming clusters from the centre of the clusters. The centre of a cluster is referred to as centroid. According to Ayramo and Karkkainen (2006) and Rousseeuw et al. (1992), in the partitioning methods, the algorithm divides the dataset into *k* clusters, where the integer *k* needs to be specified by the user. The clustering criterion determines the quality of solutions. Upon changing the clustering criterion, there is a possibility of constructing a robust clustering method that minimizes error and efficiently deals with missing data values (see Ayramo and Karkkainen, 2006).

Most of these techniques use three distinct procedures which include a method of initiating clusters, a method for allocating entities to initiated clusters and a method of relocating some or all of the entities to other clusters once the initial classificatory process has been completed (see Everitt, 1974).

The clustering criterion is derived from the following fundamental matrix equation given below

$$T = W + B \tag{1.1}$$

where T refers to the total scatter or dispersion matrix, W is the matrix of within – groups dispersion, that is, W= $\sum_{i=1}^{p} W_i$, W_i is the dispersion matrix for group i, and B is the between – groups dispersion matrix. T is fixed for any given data set so functions of B and W are sought as clustering criteria. When p =1, the matrix equation (1.1) reduces to an equation involving scalars. Thus, a good grouping index will minimize W or equivalently maximize B. In case where the number of measurements on each entity is greater than 1, group criteria can be found by using various techniques including minimizing the trace of the pooled – within groups matrix of sums of squares cross products, minimization of the determinant of the within – cluster matrix of sums of squares, cross products and average entity stability (see Everitt, 1974).

Partitioning technique mainly includes the K-means and K-medoids method. In K-means, the cluster centers are computed and this may not be the original data point. In K-medoids, each cluster centroid is represented by a point in the cluster.

Three of the algorithms considered of the partitioning type are partitioning around medoids (PAM), clustering large applications (CLARA), and fuzzy (see Rousseeuw et al., 1992). Each object of the data is assigned to exactly one cluster in PAM and CLARA whereas the fuzzy method spreads each object over various clusters (see Rousseeuw et al., 1992).

1.7.2 The Generalized Linear Models (GLMs)

There have been numerous studies done on standardizing CPUE in fishery based on spatial trends using GLMs and other mathematical/statistical modelling techniques. GLMs are one of the most common techniques used to standardize CPUE.

GLMs were introduced by Nelder and Wedderburn in 1972 for the analysis of non-Gaussian data. They include special cases, linear regression and analysis of variance models, logit and probit models for quantal responses, log linear models and multinomial response models for counts and some commonly used models for survival data. These models share a number of properties such as linearity and there is a common method for computing parameters (see Nelder and McCullagh, 1989).

Ordinary linear models take the form:

$$u = \sum_1^p x_j \beta_j,$$

where βs represent the parameters whose values are usually unknown and have to be estimated from the data. Let i index the observations then the systematic part of the model may be written as

$$E(Y_i) = u_i = \sum_1^p x_{ij\beta_j\,;i=1,\ldots,n},$$

where x_{ij} is the value of the jth covariate for the observation i .

Generalized Linear Models have three components:

1. The random component: the components of Y have independent Normal Distributions with E(Y) = μ and constant variance σ^2;

2. The systematic component: covariates $x_1, x_2, \ldots x_p$ produce a linear predictor φ given by $\varphi = \sum_1^p x_j \beta_j$;

3. The link function between the random and systematic components; $\mu = \varphi$.

The generalization introduces a new symbol φ for the linear predictor and according to the third component, μ and φ are in fact identical. If we write $\varphi_i = g(u_i)$, then g(.) will be called the link function.

From the above formulation, classic linear models have a normal or Gaussian distribution in Component 1 and the identity function for the link in Component 3.

An important characteristic of GLMs is they assume independent observations or are at least uncorrelated (see Nelder and McCullagh, 1989). The distribution of the response variables (normal, binomial, Poisson and gamma distribution) can be written in the form of a univariate exponential family. The density of a univariate exponential family for the response variable *y* is defined by

$$f(y\,\theta) = \exp\left(\frac{y\theta - b(\theta)}{\emptyset}\, w + \, c(y, \emptyset, w)\right).$$

The log density is given by

$$\log\, f\,(y\,\theta) = \frac{y\theta - b(\theta)}{\emptyset}\, w + c(y, \emptyset, w).$$

The parameter θ is called the natural or canonical parameter. For the function $b(\theta)$, it is required that $f(y\,\theta)$ can be normalized and the first as well as the second derivative $b'(\theta)$ and $b''(\theta)$ exist. The parameter $\emptyset$ is a dispersion parameter, while *w* is a known value, usually a weight. In the case of the normal distribution, $\emptyset = \sigma^2$ (see Fahrmeir et al., 2013).

1.7.3 Catch Per Unit Effort Standardization

The broad research questions dealt with identifying fishing fleets that target albacore tuna in Fiji and develop a spatially adjusted CPUE for albacore fishery in Fiji using statistical analyses. The fisheries data used in the literature review ranges from 1952 to 2013. The geographical area in the literature review is mostly for the South Pacific and the Western and Central Pacific Ocean (WCPO).

Many authors such as Molina et al. (2011), Bigelow and Hoyle (2012), Mourato et al. (2011), Lee et al. (2012), Hazin et al. (2007) used cluster analysis to identify the vessels targeting a particular tuna species. Molina et al. (2011), Bigelow and Hoyle (2008, 2012), Hoyle et al. (2013), Hoyle (2009), Bigelow (2006), Bromhead et al. (2009), Coelho et al. (2014), Travassos et al. (2009), Lee et al. (2014), Lee et al. (2012), Yeh and Chang (2013), Hazin et al. (2007), Pons and Domingo (2014), Okamoto et al. (2001, 2011), Bentley et al. (2012), Meneses et al. (2004), Kell et al. (2011), Glazer and Butterworth (2016), Goni and Arrizabalaga (2005) and Gulati and Premchand (2015) have used GLMs to standardize CPUE.

Most of the studies in the literature review dealt with longline fishery. Majority of the studies dealt with standardization of tuna species mainly albacore, bigeye and yellowfin. The data has mainly been collected by the fisheries observer program and secondary data were used from the Fisheries Agencies and Councils. Domestic and distant water fleets were used in the literature review from mainly Japan, Korea, Taiwan, Brazil, Chinese Taipei, Fiji, American Samoa, Vanuatu, Cook Islands, Tonga and Samoa. Almost all studies in the literature review used non-experimental research design where secondary data was analysed. As stated by Bigelow and Hoyle (2008), South Pacific albacore are mainly landed at canneries in Pago Pago, American Samoa and Levuka, Fiji. A wide range of fishing data has been collected ranging from catch, length, weight, fish species, vessel information, environmental factors and spatial information.

Most studies used the traditional method known as GLMs to standardize CPUE. The GLM has been used to assess the tuna stock status which has been presented in many international expert forums in recent years (see Gulati and Premchand, 2015). There were also cases where cluster analysis has been carried out to identify vessels targeting certain tuna species such as albacore. In most of the studies, analysis was done at 5 degrees spatial resolution due to the large geographical area covered. In some studies, the whole area was divided into different regions and the analysis was done in the different regions. The overall analysis considering the whole area was also done. In some studies, the whole area was analysed to develop spatially adjusted CPUE. Almost all studies used linear relationships in analysing fisheries data. Bigelow and Hoyle (2008) selected their model based on Bayesian Information Criteria. While comparing the models with Akaike Information Criteria (AIC) using log (effort) rather than effort, a much better fit to the data was seen (see Hoyle, 2009).

Molina et al. (2011) recommended that the inclusion of spatially adjusted CPUE could improve the Spatial Ecosystem and Population Dynamics Model (SEAPODYM) performance and therefore its use in fisheries management. Through further statistical analyses, it can explore the in-depth relationship between catch rates and environment variables. Hoyle et al. (2013) mentioned that given the lack of data on swordfish, there were highly variable and imprecise indices developed for most regions and as such, the authors recommended using Japanese index for the western central region and Taiwanese index for eastern central region as indices of abundance. Chang et al. (2011) stated that there is a need to carry out research on the spatial distribution, biology and exploitation of albacore in the eastern Mediterranean Sea through a regional research program in order to have sufficient data required to design a sustainable fisheries management plan. Coelho et al. (2014) recommended collecting and compiling additional data, especially operational details which could be used to identify the fishery targeted. Teo and Block (2010) recommended reducing variability of the data by modelling the data on a trip by trip basis.

Bigelow and Hoyle (2008) suggested that in the cases of distant–water fleets, there are long term trends in unstandardized CPUE which are not consistent with all fisheries that have constant catchability. The previous standardization methods indices for specific fleets were produced for Japan, Korea and Taiwan and no standardized indices were done for domestic fleets (see Bigelow and Hoyle, 2012). Lack of the required information may potentially affect CPUE standardization as mentioned by Hoyle et al. (2013) that the lack of vessel information could potentially bias CPUE as it is not possible to account for possible increases in efficiency over time such as the phasing out of old vessels and introduction of new ones. It is difficult to work with aggregate data to use clustering in identify targeting Hoyle (2009). Also aggregated data provide limited information on the factors that affect CPUE. Chang et al. (2011) mentioned that targeting of albacore has declined, with a reduction in the number of vessels and shifting of target species to bigeye and yellowfin tunas which currently have higher commercial value. Coelho et al. (2014) pointed out that one possible shortcoming of their work was that the targeting effects were not included directly in the models and there is a possibility that the relative importance of the fishery may have changed over time.

An adjusted CPUE index with spatial structure has been developed for albacore tuna in the South Pacific using the operational log sheet data from distant waters and domestic fleets (See Molina et al., 2011). In this thesis, an attempt is made to develop a spatially adjusted CPUE index for the albacore longline fishery in Fiji as very little study has been carried out for Fiji. For this study, we extend the techniques used in Molina et al. (2011) as the research questions are similar. The use of GLM by Bigelow and Hoyle (2008) in CPUE standardization of distant water fleets targeting South Pacific albacore is a similar concept that was used in this study. The only difference is that they used GLMs in four regions by sub-dividing the study area. In order to address the issue of targeting, cluster analysis has been done to disaggregate albacore and bigeye tuna targeting operations based on species composition (see Bigelow and Hoyle, 2012). A Ward Hierarchical clustering (hclust) was applied to the different regions to produce a dendogram to illustrate the number of clusters in the data (see Bigelow and Hoyle, 2012). Appendix A shows the summary of key findings from literature review regarding the relevant CPUE standardizations.

Chapter 2

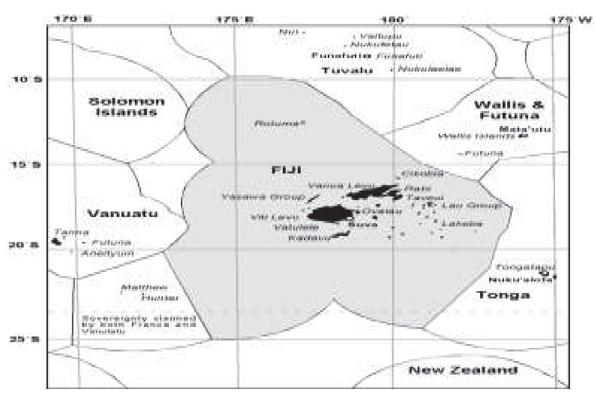
Identifying Domestic Fleets that Target Albacore Tuna in Fiji using Cluster Analysis

2.1 Introduction

Many authors have used the cluster analysis techniques to identify the target species based on the catch composition of tuna (see Molina et al., 2011 & Bigelow and Hoyle, 2009). In this Chapter, cluster analysis is carried out to avoid bias in CPUE standardization that can occur due to targeting change. The original catch and effort data from the Fiji Fisheries Department was investigated and summarized. The clustering techniques used were both hierarchical and partitioning (see Molina et al., 2011). As a result of cluster analysis, the catch and effort retained was for those trips targeting albacore tuna.

2.2 Fiji's Economic Exclusive Zone (EEZ)

This research focused on local longline vessels targeting albacore tuna in Fiji's Economic Exclusive Zone (EEZ). EEZ is the sea territory within 200km from the shore that belongs to a country. The country owns the resources found in its EEZ. Fiji's EEZ extends from 12° and 20° S latitude and 177° E to 177° W longitude with an area of approximately 1,281,122 km^2.

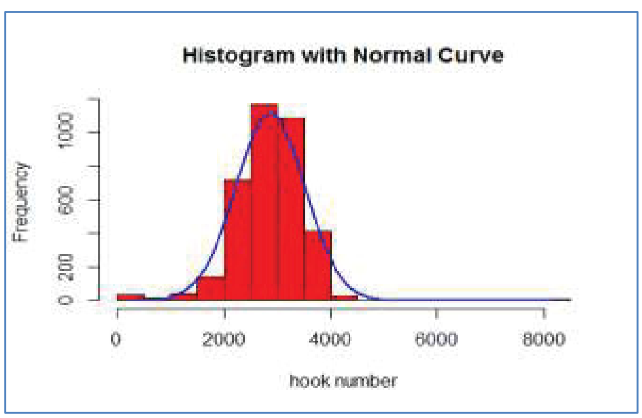


Source: http://fisherymanagement.wikia.com/wiki/Fiji

Figure 1: Map showing Fiji's EEZ

2.3 Data Preparation

Fisheries data was represented by a two dimensional array in which rows are indexed by sets. The columns of the data matrix are the variates such as albacore number, hook set, latitude, longitude and so on. Variates are regarded as responses or dependent variates whose values are believed to be affected by the explanatory variables or covariates or independent variates. In our study, the covariates were all quantitative, taking numerical values. The Fiji Fisheries Department provided operational level data from 2000 to 2015. The data from 2000, 2002 and 2015 were removed as they had missing trips for more than 1 quarter which could potentially bias the CPUE standardization. The data retained was from 2003 to 2014. The full data set was split into a period of three years and the cluster analysis was done separately on each time period. Those sets assigned outside Fiji's EEZ were discarded. The additional data that was not required for the CPUE standardization, blank columns and columns with very few data as well as erroneous data were removed. The information regarding catch quarter, month and year was extracted from the variable *catch_dtime*. Geographical position of individual sets was available. The function *floor.precise* was used in MS Excel 2010 to aggregate the latitude and longitude to 1 degree resolution. The year-quarter explanatory variable was constructed by concatenating year and quarter. There were 11 sets with no information on the number of hook sets and 5 sets with inconsistent number of hook sets and as such these were removed. The descriptions of the variables used in the analysis are listed in the Appendix B. Figure 2 below shows the distribution of hooks for the fishing sets.



Source: The Fisheries Department of Fiji

Figure 2: Histogram showing the distribution of hooks

Majority of the hooks used in the fishing sets were around 2000 to 4000. There were more sets with the number of hooks lower than 2000 than hooks greater than 4000.

2.4 Catch Analysis

Table 1 below shows the number of fishing trips, sets and vessels that operated in Fiji from 2003 to 2014 for domestic fleets.

Table 1: Fishing Trips, Sets and Vessels

Year	Trips	Sets	No. of Trips made by Vessels
2003	13	138	13
2004	10	113	9
2005	34	327	30
2006	29	336	24
2007	20	227	17
2008	27	296	17
2009	26	207	10
2010	22	161	7
2011	10	64	8
2012	6	96	5
2013	73	698	22
2014	102	992	27
Total	372	3655	189

Source: The Fisheries Department of Fiji

The above table shows that there were a total of 372 fishing trips and 3655 sets from 2003 to 2014 carried out by the domestic fleets from the data provided by the Fiji Fisheries Department. There were 79 domestic vessels that participated in tuna fishery from 2003 to 2014. There were also some vessels that made more than 1 trip within the year. The information for quarter 4 for 2003 and 2010 as well as quarter 3 for 2011 was not available. The maximum number of trips and sets were in 2014. On the other hand, the minimum number of trips and sets were in 2012 and 2011, respectively. Table 2 below shows the total tuna catch number, hooks (effort), tuna weight estimate and average tuna length from 2003 to 2014.

Table 2: Total tuna catch and effort

Year	Total Tuna Catch Number	Total Hooks (Effort)	Total Tuna Weight Estimate (Kg)	Average Tuna Length (cm)
2003	4936	319976	2197.71	91.61
2004	4410	292625	1844.78	93.23
2005	16839	905365	5744.81	95.70
2006	18726	919132	5816.9	95.06
2007	11176	642585	3976.64	96.01
2008	10946	893307	5852.33	100.10
2009	6863	547843	4202.89	99.80
2010	3796	404821	2948.92	94.78
2011	1309	168929	1442.79	104.10
2012	2374	295738	1882.21	96.18
2013	14047	2087051	12321.75	91.16
2014	33108	2997403	18820.21	97.96

Source: The Fisheries Department of Fiji

The total values are calculated from the total fishing sets in the single year from 2003 to 2014. Each time series had different number of sets. Total hooks represent the effort used in tuna catch. The maximum number of tuna catch and hooks used were in 2014 while the minimum were in 2011.Total hooks refer to the sum of hooks used in the sets. Total maximum weight estimate for tuna was in 2014 and the minimum in 2011. The minimum values in 2011 are attributed to no fishing trips by the domestic fleets in quarter 3 according to the data received from the Fisheries Department of Fiji. The average tuna length was maximum in 2011 and minimum in 2013.

Table 3 shows the total catch and effort by the vessels targeting albacore during the study period.

Table 3: Total albacore catch and effort

Year	Total Albacore Caught	Total hook_set (Effort)
2003	1847	168833
2004	1772	197135
2005	11454	833393
2006	14661	890012
2007	6688	576356
2008	7103	772519
2009	4832	515804
2010	1358	280584
2011	543	116997
2012	1724	287454
2013	7998	1417103
2014	17804	2401642
Total	**77784**	**8457832**

Source: The Fisheries Department of Fiji

Based on the data provided by the Fiji Department of Fisheries, the albacore catch in numbers in 2003 was 1847 and in 2004 was 1772. There was a big increase in the albacore catch in 2005 and 2006 with 11,454 and 14,661, respectively. In 2007, albacore catch decreased to 6688, while in 2008 the albacore catch was 7103. From 2009 to 2012, the albacore catch declined from around 4832 to 1724. There was an increase seen in albacore catch in 2013 and 2014 with 7998 and 17804 respectively. The effort (number of hooks) increased from 2003 to 2005, fluctuated from 2006 to 2008, decreased from 2009 to 2011 and increased in the last period from 2012 to 2014. The effort varied between 116,997 and 2,401,642 hooks from 2003 to 2014.

Figure 3 shows the trend of catch and effort made during the study period.

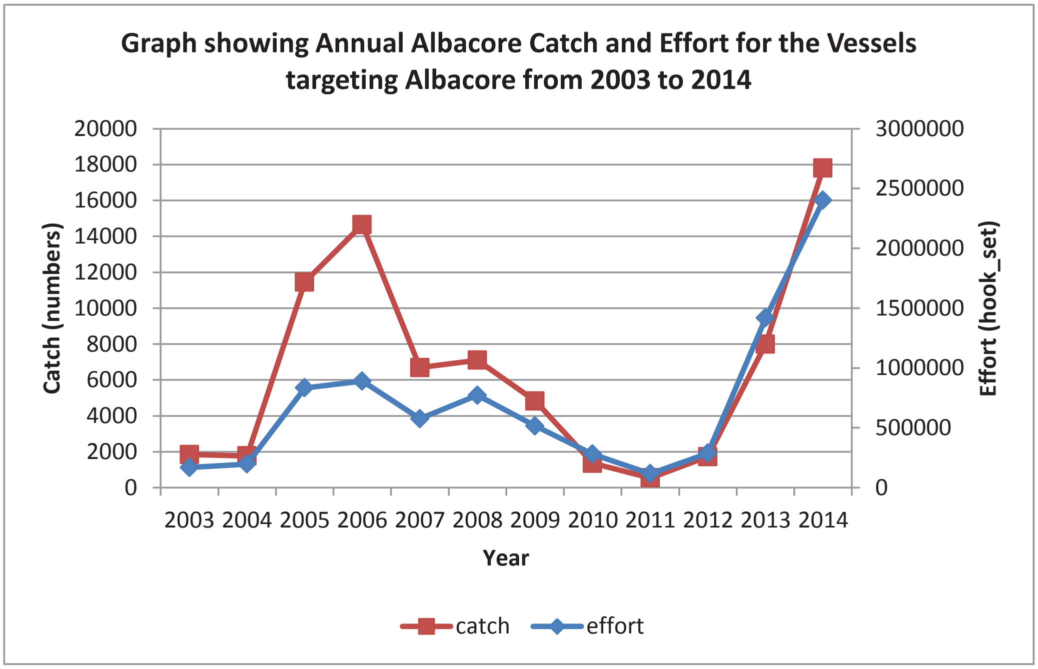


Figure 3: Albacore annual catch and effort in Fiji's EEZ for the vessels targeting albacore

From Table 3 and Figure 3, it can be seen that generally the total albacore annual catch from domestic fleets targeting albacore increased from 2003 to 2006. From 2006 to 2008, there was generally a decreasing trend. From 2009 to 2011, there was a declining trend. From 2012 to 2014, there was a significant increase in the total annual albacore catch. The annual albacore catch in the entire time series to some extent is attributed to the number of fishing trips made by the vessels within the year.

2.5 Cluster Analysis

The clustering analyses are used to partition the fishing effort into discrete groups that represent different modes of operation with respect to species targeting (see McKechnie et al., 2013, 2014). Molina et al. (2011) carried out cluster analysis to identify fleets targeting albacore to avoid bias due to targeting change (see Bigelow and Hoyle, 2009) also stated that cluster analysis is used to remove bias through separating the catch and effort data according to target species. The factors that can cause bias include the fleets switching target, spatial changes in fishery, use of deeper longline gear and higher catch rates of bigeye tuna. In cases where the information on targeting strategy is not available, cluster analysis is ideal to identify the target species and therefore assists to generate the standardised CPUE series (Hazin et al., 2007; Travassos et al., 2009; Mourato et al. 2011).

In this study, Figure 4 shows the tuna catch composition from 2003 to 2014. It reveals that albacore tuna had the highest catch proportion throughout the time series used in the study. Yellowfin also had a significant percentage of catch proportion compared to bigeye tuna.

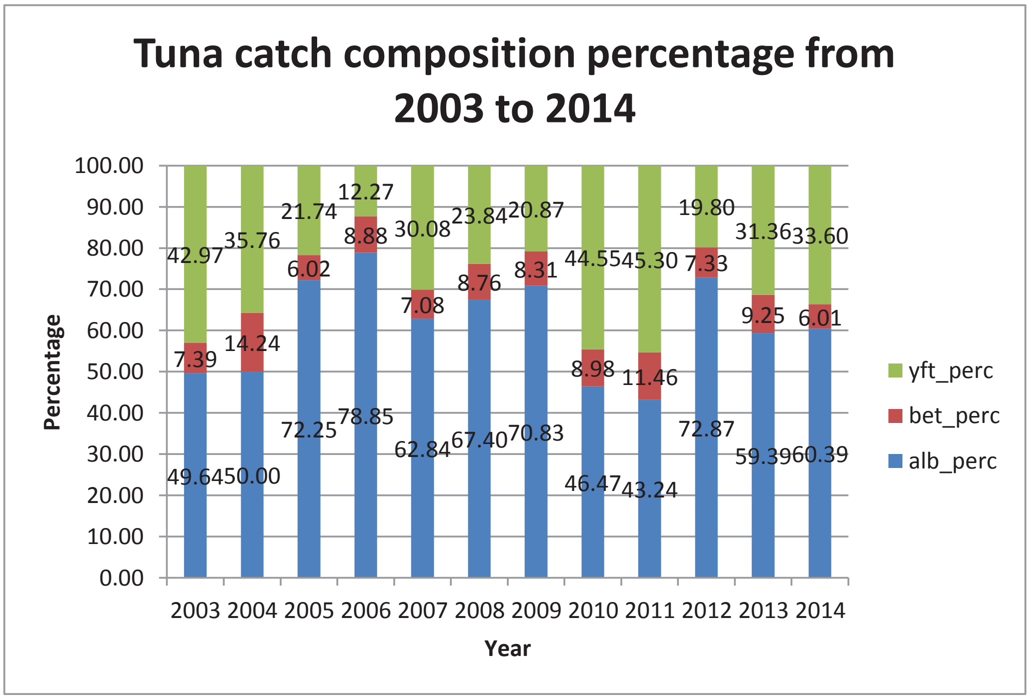


Figure 4: Tuna catch composition

Then, the first step for building the new spatially adjusted catch per unit effort (CPUE) was the identification of fleets targeting albacore with cluster analysis on species composition of albacore, bigeye and yellowfin tuna. Statistical package R (Version 3.2.2) was used to do the cluster analysis. According to Molina et al. (2011), the underlying assumption is that within a trip targeting did not change thus the cluster analysis focused on trip rather than set. Also, Bigelow and Hoyle (2012) discussed that clustering was conducted by trip as clustering on each set may confuse the chance of failure of a set to capture albacore with active targeting of other species whereas this is unlikely at the trip level.

Bigelow and Hoyle (2009) had removed longline sets with zero tuna catch as they are uninformative in the cluster analysis. Bigelow and Hoyle (2012) removed trips that caught zero tuna from the cluster analysis as zero proportions were uninformative in cluster analysis. However, we included all the 44 sets with zero tuna catch as they were part of the trips that had tuna catch. All the 372 trips had tuna catch from 2003 to 2014. There were 255 sets out of 3655 that had zero albacore catch.

Two clustering routines *hcluster* and *clara* from the package cluster were used in the analysis. *hcluster* routine is a mixture of routines dist and clust from the package cluster. This technique works by defining the Euclidean or geometric distance between points as a distance function. The 372 fishing trips can be considered as points $x_1,\ x_2,\ x_3, \ldots\ x_n$. According to Cohen (1988), Euclidean or geometric distances between the points, $x_1,\ x_2,\ x_3, \ldots\ x_n$ are mathematically given by:

$$(Distance)^2 = \frac{(difference\ in\ x_1)^2}{variance\ of\ x_1} + \frac{(difference\ in\ x_2)^2}{variance\ of\ x_2} + \frac{(difference\ in\ x_3)^2}{variance\ of\ x_3} + \cdots + \frac{(difference\ in\ x_n)^2}{variance\ of\ x_n}$$

Taking the square root, we get Euclidean distances as the root sum-of-squares of differences. The *hcluster* routine produced dendograms for determining the appropriate number of clusters based on species targeted from the data (see Molina et al., 2011).

The cluster analysis has been done in periods of three years since there was no suspected change in targeting.

The second clustering routine, *clara* was used to partition the data sets into appropriate number of clusters as determined by dendograms. *Clara* partitions the data set with respect to medoid points and efficiently handles large data sets and only accepts input of an n × p data matrix (Rousseeuw et al., 1992; Ayramo and Karkkainen, 2006). Upon investigation, *clara* was used to represent the clustering of the tuna catch composition by fishing trips into 3 clusters. Traditional partitioning clustering methods such as K-means and hierarchical methods produce separated clusters, which imply that each data point is assigned to only one cluster (see Ayramo and Karkkainen, 2006).

The longline gear setup for targeting albacore also catches yellowfin tuna because these two species are found at the same depth and are attracted by the same bait (see

Molina et al., 2011). As a result of that, sometimes vessels that target albacore catch mostly yellowfin in some seasons. Clusters were validated where the data was divided into different time series and each clustered independently (see Everitt, 1974).

2.6 Results and Discussion

The clustering was done at an interval of three years from 2003 to 2014. A dendogram showing the clusters was produced for every period. This study only considers Fiji's EEZ so this provided us with the luxury to examine clusters in detail. Thus, there were 3 clusters used for each time interval to identify fishing trips targeting albacore tuna.

The criterion used was that if the sum of the proportion of the yellowfin and albacore tuna species was less than 80%, then we defined the fishery as not targeting albacore (See Molina et al., 2011).

According to Bigelow and Hoyle (2012), cluster results pertaining to South Pacific albacore and South Pacific albacore and yellowfin were retained for analysis and bigeye and yellowfin tuna were discarded. However, in our case all the clusters either had the highest proportion of albacore or yellowfin. Thus we only considered those clusters where the albacore catch composition was the highest as a criterion for targeting albacore.

The clustering was done by the trips that had tuna catch. All the 372 trips had tuna catch. The first period was from 2003 to 2005 inclusive. There were 57 trips with tuna catch in this time series.

The hierarchical cluster was used to form the tree diagram called the dendogram for the period 2003 to 2005 as shown below.

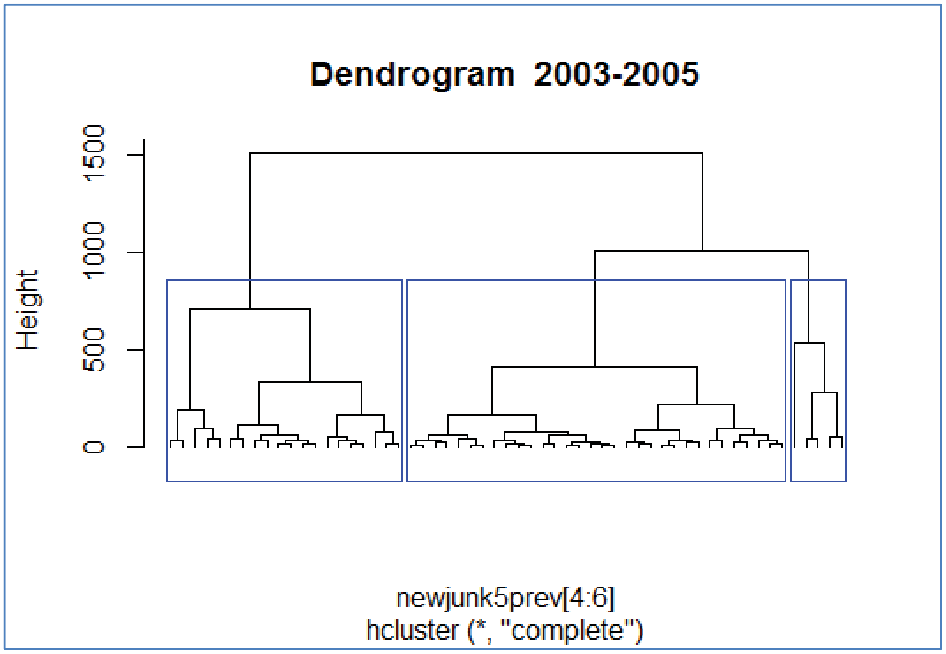


Figure 5: Dendogram from 2003 to 2005

The y-axis shows the distance between the clusters and the x-axis shows the trips and clusters. The dendogram lists all the trips and shows the level of similarity between the clusters. The rectangles denote the three clusters shown by the dendogram as the three branches occur at the same vertical distance.

The *clara* function in R software was used to partition the data sets into three clusters as determined by the dendogram. The 3 clusters for the period 2003 to 2005 are given by the Table 4 .

Table 4: Result of cluster analysis from 2003 to 2005

Cluster	Bigeye (%)	Yellowfin (%)	Albacore (%)	Total (%)
1	11.15	62.69	26.16	100
2	9.27	27.73	63.00	100
3	4.70	8.87	86.43	100

In the first period, all the clusters had the proportion of yellowfin and albacore tuna species greater than 80%. The first cluster was targeting yellowfin tuna as it had the highest percentage catch. The second and the third clusters were targeting albacore as they had the highest catch composition. The catch composition was dominated by albacore in clusters 2 and 3. Cluster 1 did not target albacore because the percentage of yellowfin tuna is more than albacore. There were 14 trips in Cluster 1, 27 in Cluster 2, and 16 in Cluster 3. As a result of the cluster analysis, 43 out of 57 trips were considered to be targeting albacore.

Spatial data analysis is involved when data are spatially located and explicit consideration is given to the possible importance of their spatial arrangement in the analysis or in the interpretation of the results (see Bailey and Gatrell, 1995).

Figure 6 shows the map of spatial distribution of clusters targeting albacore from 2003 to 2005.

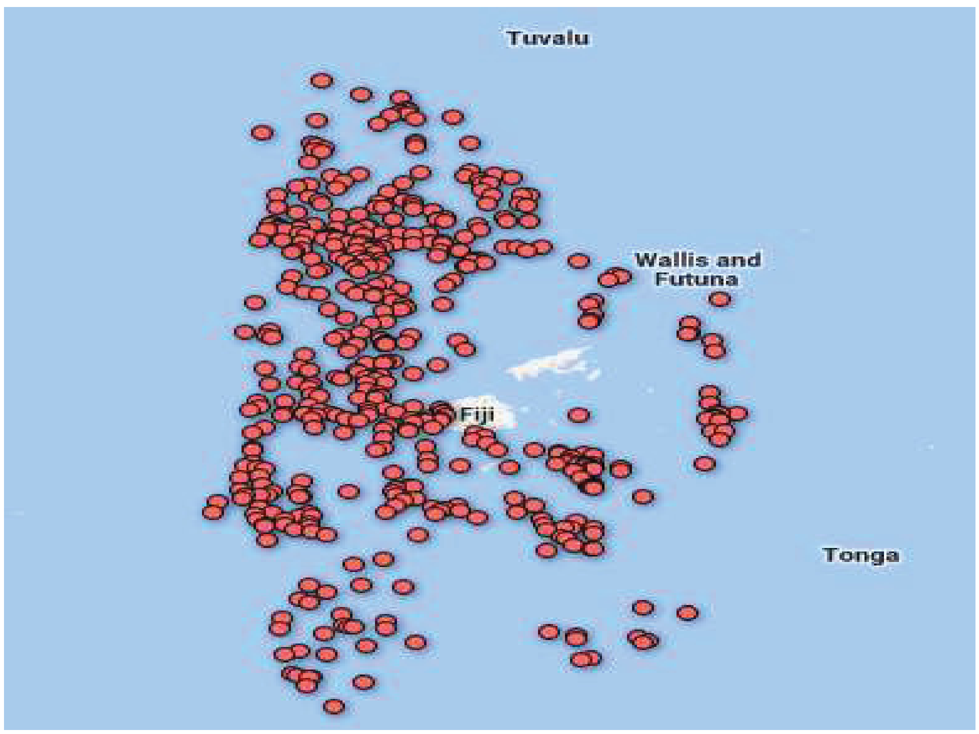


Figure 6: Spatial distribution of fishing sets targeting albacore from 2003 to 2005

The map shows that the majority of the albacore longline fishing activity concentrated in the western side of Fiji's EEZ with latitudinal effect being seen.

The second period considered for clustering was from 2006 to 2008 inclusive. There were 76 trips with tuna catch.

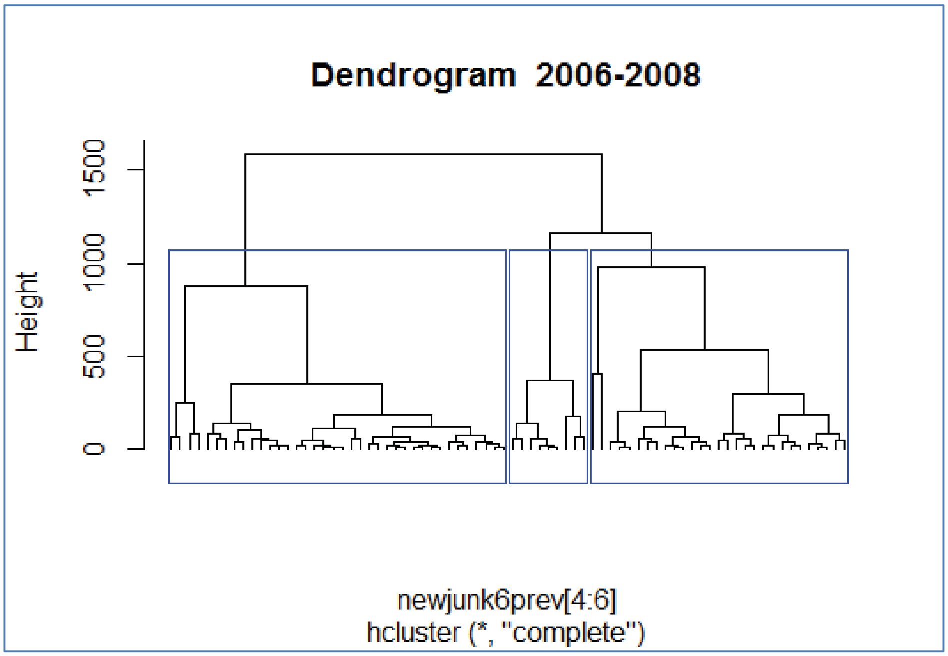


Figure 7: Dendogram from 2006 to 2008

The rectangles denote the three clusters formed by the dendogram. The results of using the *clara* function in R software for the three clusters for the period 2006 to 2008 is given in the Table 5.

Table 5: Result of cluster analysis from 2006 to 2008

Cluster	Bigeye (%)	Yellowfin (%)	Albacore (%)	Total (%)
1	11.90	31.65	56.45	100
2	7.00	11.00	82.00	100
3	15.01	62.72	22.27	100

In the second period, all the clusters had the proportion of yellowfin and albacore tuna species greater than 80%. The first and the second clusters were targeting albacore as it had the highest catch composition. The third cluster was targeting yellowfin tuna as it had the highest percentage catch. Cluster 3 did not target albacore because the percentage of yellowfin tuna is more than albacore. There were 24 trips in Cluster 1, 43 in Cluster 2, and 9 in Cluster 3. As a result of the cluster analysis, 67 out of 76 trips were considered to be targeting albacore.

Figure 8 shows the map of spatial distribution of clusters targeting albacore from 2006 to 2008.

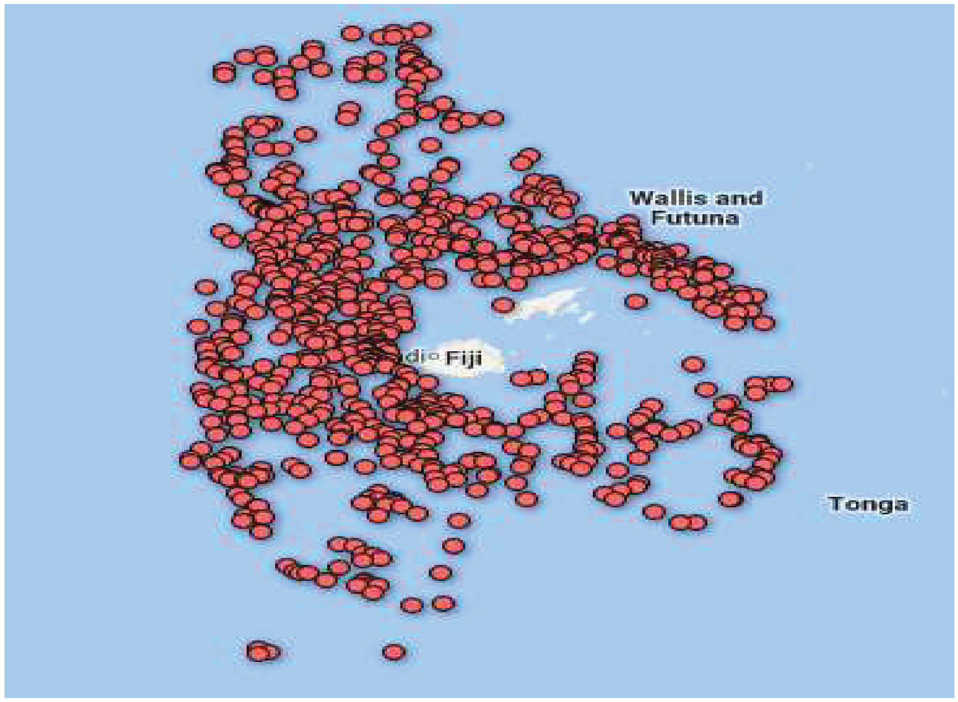


Figure 8: Spatial distribution of fishing sets targeting albacore from 2006 to 2008

Almost similar pattern in albacore fishing activity was seen as in 2003 to 2005. Majority of the albacore fishing activity concentrated in the Western EEZ of Fiji.

The third period considered for clustering was from 2009 to 2011 inclusive. There were 58 trips with tuna catch.

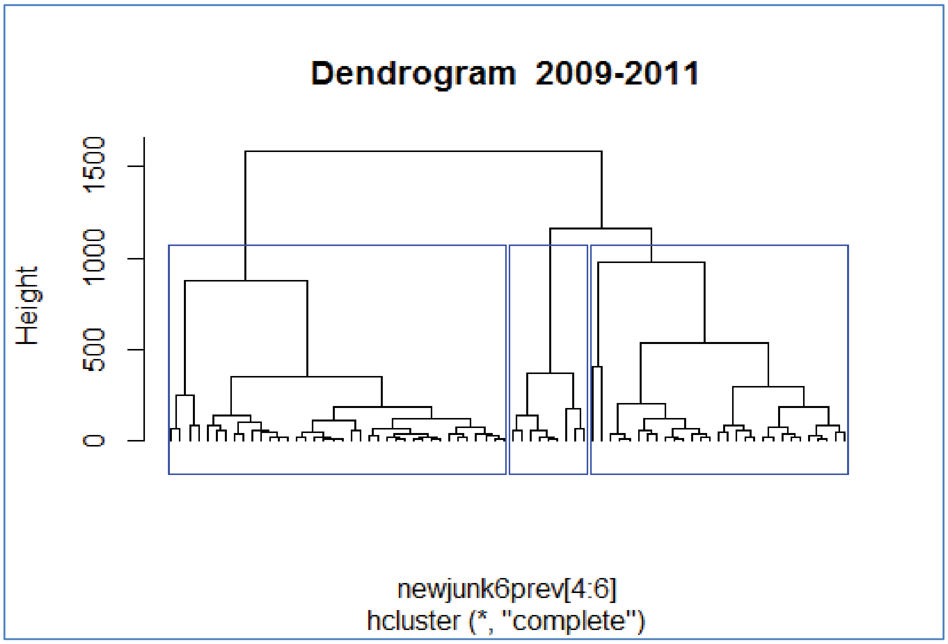


Figure 9: Dendogram from 2009 to 2011

The rectangles denote the three clusters formed by the dendogram.

The results of using the *clara* function in R software for the three clusters for the period 2009 to 2011 is given in Table 6.

Table 6: Result of cluster analysis from 2009 to 2011

Cluster	**Bigeye (%)**	**Yellowfin (%)**	**Albacore (%)**	**Total (%)**
1	5.33	74.91	19.76	100
2	11.24	43.64	45.12	100
3	10.63	15.52	73.85	100

In the third period, all the clusters had the proportion of yellowfin and albacore tuna species greater than 80%. The second and the third clusters were targeting albacore as it had the highest catch composition. The first cluster was targeting yellowfin tuna as it had the highest percentage catch. Cluster 1 did not target albacore because the percentage of yellowfin tuna is more than albacore. There were 11 trips in Cluster 1, 19 in Cluster 2, and 28 in Cluster 3. As a result of the cluster analysis, 47 out of 58 trips were considered to be targeting albacore. Figure 10 shows the map of spatial distribution of clusters targeting albacore from 2009 to 2011.

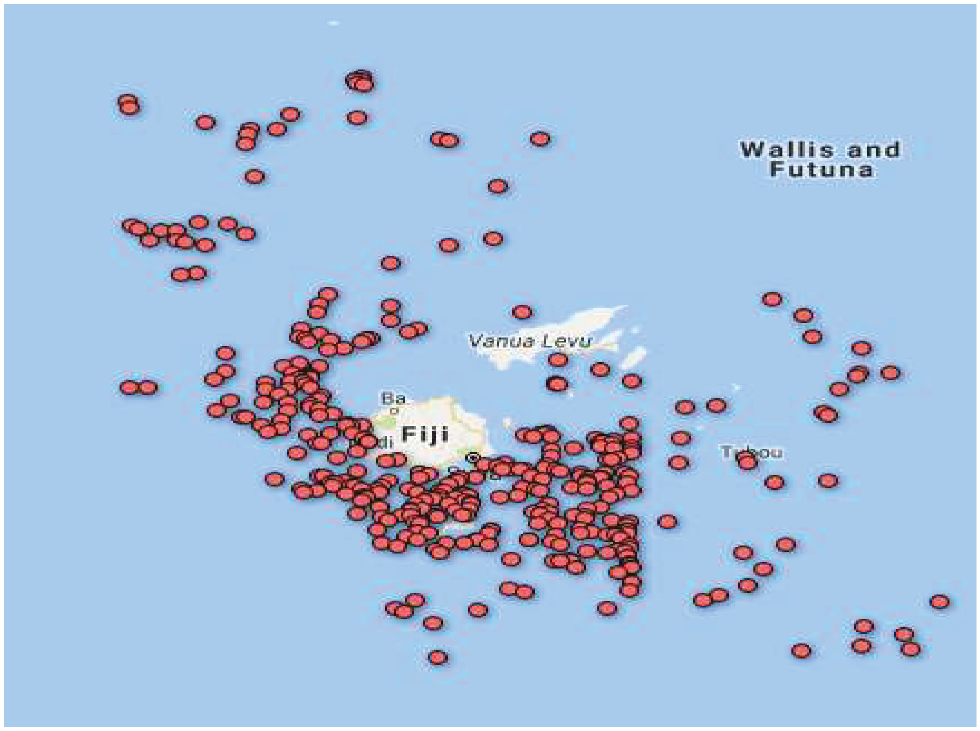


Figure 10: Spatial distribution of fishing sets targeting albacore from 2009 to 2011

The map shows significant change in albacore fishing when compared to the previous period. The albacore fishing activity mainly concentrated in the Southern EEZ of Fiji during this period.

The fourth period considered for clustering was from 2012 to 2014 inclusive. There were 181 trips with tuna catch.

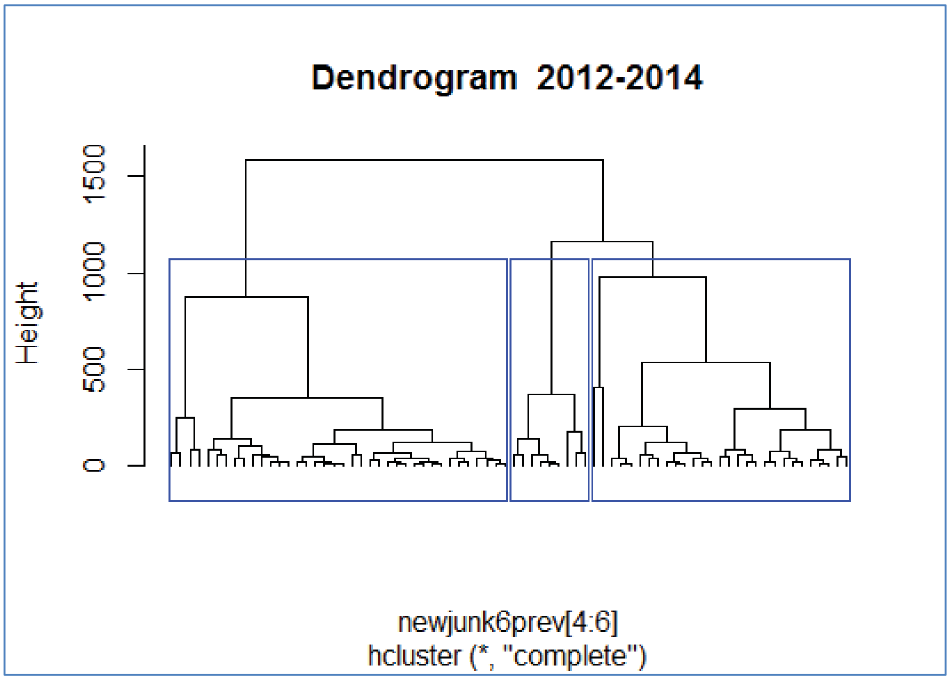


Figure 11: Dendogram from 2012 to 2014

The rectangles denote the three clusters formed by the dendogram.

The results of using the *clara* function in R software for the three clusters for the period 2012 to 2014 is given in the Table 7 below.

Table 7: Result of cluster analysis from 2012 to 2014

Cluster	**Bigeye (%)**	**Yellowfin (%)**	**Albacore (%)**	**Total (%)**
1	4.39	11.62	83.99	100
2	11.62	30.13	58.25	100
3	7.23	70.08	22.69	100

In the fourth period, all the clusters had the proportion of yellowfin and albacore tuna species greater than 80%. The first and the second clusters were targeting albacore as it had the highest catch composition. The third cluster was targeting yellowfin tuna as it had the highest percentage catch. Cluster 3 did not target albacore because the percentage of yellowfin tuna is more than albacore. There were 41 trips in Cluster 1, 86 in Cluster 2, and 54 in Cluster 3. As a result of the cluster analysis, 127 out of 181 trips were considered to be targeting albacore. Figure 12 shows the map of spatial distribution of clusters targeting albacore from 2012 to 2014, which reveals that that the albacore fishing sets was throughout Fiji's EEZ.

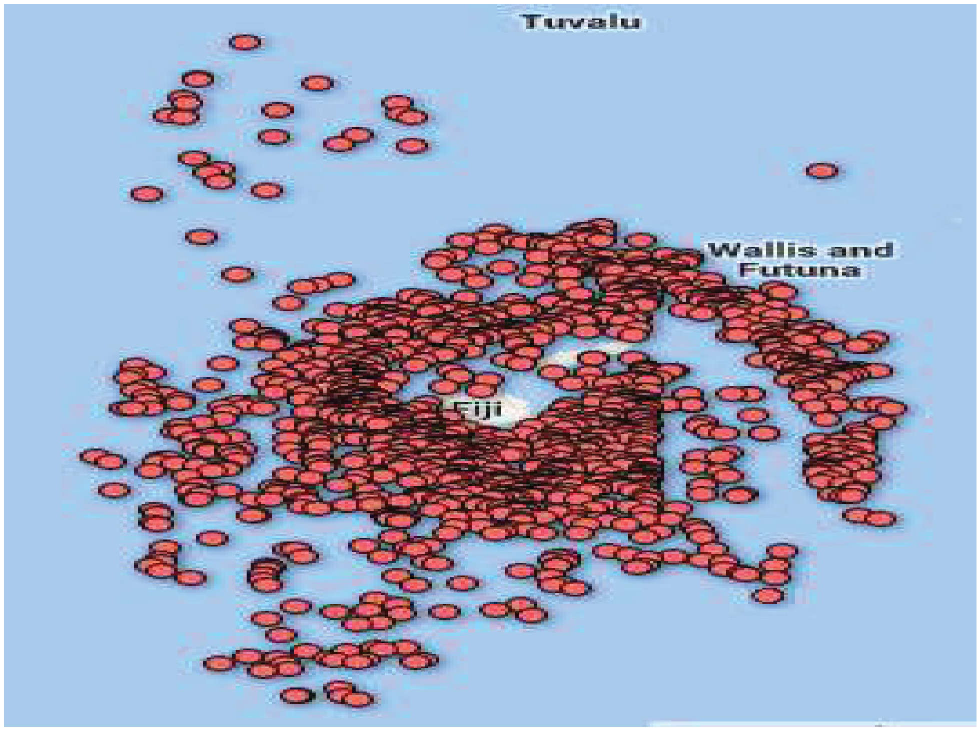


Figure 12: Spatial distribution of fishing sets targeting albacore from 2012 to 2014

The spatial distribution of fishing sets targeting albacore had a different fishing activity distribution compared to the other three periods. Figure 12 shows fishing sets targeting albacore were scattered throughout Fiji's EEZ in the last period in the study.

Figure 13 shows the map of spatial distribution of clusters targeting albacore from 2003 to 2014 in Fiji's EEZ.

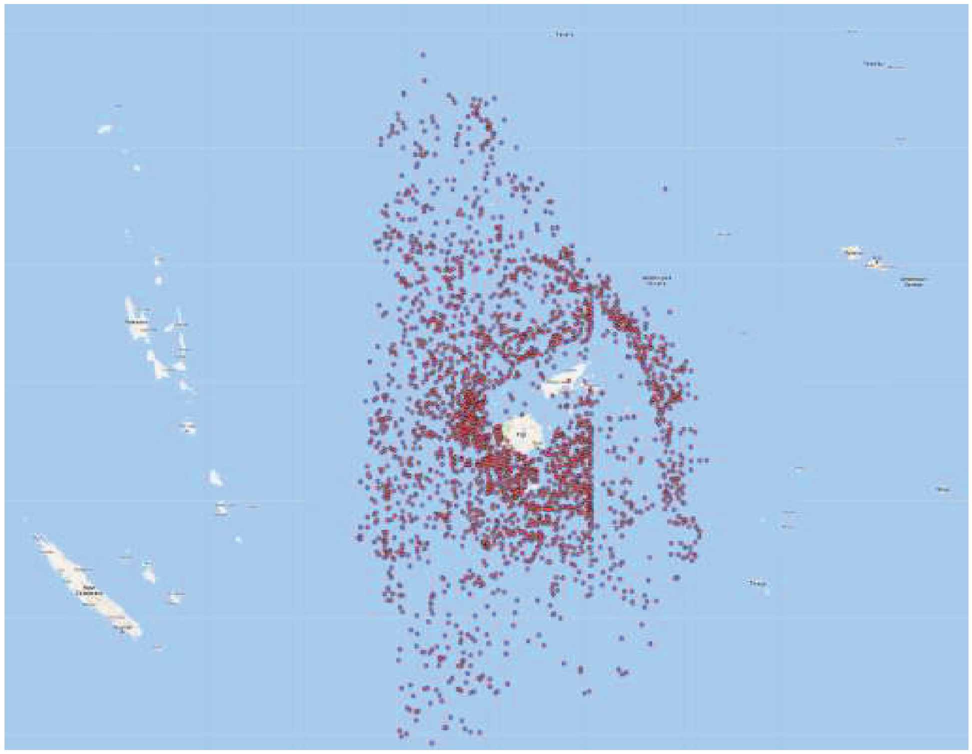


Figure 13: Spatial distribution of fishing sets targeting albacore for the entire time series

Figure 13 reveals that the fishing sets targeting albacore are scattered around Fiji's EEZ. The map reveals that majority of the fishing sets targeting albacore are more towards the south of Fiji.

2.7 Conclusion

In this chapter, there is discussion of data preparation and inspection. Both hierarchical and partitioning clustering techniques were used to identify domestic vessels that targeted albacore tuna in Fiji. Majority of the clusters were mostly found to be targeting albacore with some targeting yellowfin. This implies that albacore is a common tuna specie targeted in Fiji. After the cluster analysis, the information reduced to 2954 sets corresponding to 284 trips and 69 vessels. The results from the cluster analysis were plotted on maps to show the spatial distribution of fishing sets targeting albacore in Fiji's EEZ.

Chapter 3

Generalized Linear Model

3.1 Introduction

In this chapter, we developed a Generalized Linear Model (GLM) to standardize CPUE for Fiji Fisheries. The model was selected based on low AIC value and lower adjusted R^2 value.

3.2 Nominal Catch Per Unit Effort

Mathematically, catch per unit effort is defined by

$$CPUE_i = \sum_{i=1}^{n} \frac{catch_i}{effort_i},$$

where n is the number of years, *catch* is the number of the species caught, and *effort* is the number of hooks used in a trip. There are studies that have also used nominal CPUE as catch divided by 100 hooks and catch divided by 1000 hooks as given as

$$Nominal\ CPUE = \frac{total\ albacore\ catch}{1000}$$

or

$$Nominal\ CPUE = \frac{total\ albacore\ catch}{100}.$$

As used by Molina et al. (2011), in this study the temporal trend of the albacore nominal CPUE was estimated by dividing the total albacore catch by the total number of hooks aggregated by year using the information from the domestic fleets, that is,

$$Nominal\ CPUE = \frac{total\ albacore\ catch}{total\ number\ of\ hooks\ (effort)}.$$

Table 8: Nominal CPUE for albacore

Year	Albacore Catch	Hooks (Effort)	Nominal Albacore CPUE (Albacore Catch/ effort)	Nominal Albacore (per 1000 Hooks)
2003	1847	168833	0.01	1.85
2004	1772	197135	0.01	1.77
2005	11454	833393	0.01	11.45
2006	14661	890012	0.02	14.66
2007	6688	576356	0.01	6.69
2008	7103	772519	0.01	7.10
2009	4832	515804	0.01	4.83
2010	1358	280584	0.01	1.36
2011	543	116997	0.01	0.54
2012	1724	287454	0.01	1.72
2013	7998	1417103	0.01	8.00
2014	17804	2401642	0.01	17.80

The nominal CPUE for both years 2003 and 2004 were 2 albacore/1000 hooks. The nominal CPUE increased in 2005 and 2006 with 12 and 15 albacore/1000 hooks, respectively. For 2007 and 2008, the nominal CPUE was 2 albacore/1000 hooks. From 2009 to 2012, there was a decrease in nominal CPUE from 5 to 2 albacore/1000 hooks. In 2013 and 2014, the nominal CPUE increased to 8 and 18 albacore/1000 hooks.

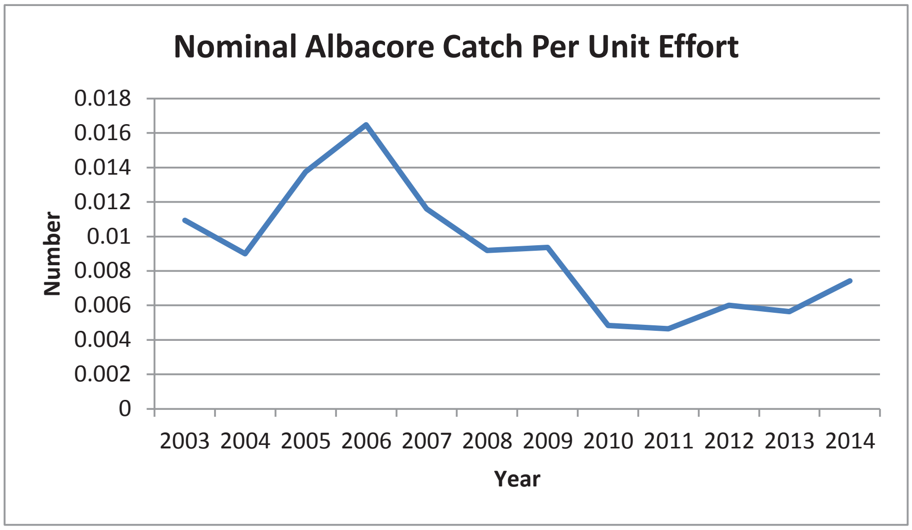


Figure 14: Nominal albacore CPUE during 2003 to 2014

Table 8 and Figure 14 show the nominal albacore CPUE for the entire study period. It reveals that the nominal albacore CPUE decreased from 2003 to 2004. There was an increase till 2006 with a decline thereafter till 2011.Then, there is an increasing trend again from 2012. The maximum nominal CPUE was in 2006 with the minimum in 2011.

3.3 Standardization Models

A key assumption in CPUE standardization is that catch rate is proportional to fish density. The relationship is $CPUE = \frac{C}{E} = qD$, where C refers to catch, E is the effort, D is the density, and q is the catchability coefficient that is related to the efficiency of fishing. The population size of fish in the area of study is a function of average density D and the size of area A, with the equation $N = DA$. The changes in CPUE through time are due to either changes in the stock density or to changes in the catchability coefficient. There is a need to standardize because q and D are not constant. Both catchability and density vary due to many factors such as vessels, skippers, changes in fishing gear, fish detection technology, spatial distribution, season and overall changes in population size. Standardization accounts for changes in q and for spatial and seasonal changes in D, so the remaining changes in CPUE can be related to longer term changes in stock density (see Hoyle et al., 2014).

3.4 GLM Applications in Standardizing Catch Per Unit Effort

Time series data has been standardized using Generalized Linear modelling techniques (see Nelder and McCullagh, 1989 and Hoyle et al., 2013). Bigelow and

Hoyle (2012) considered four predictors namely, year-quarter, vessels and two interactions, month and latitude, latitude and longitude. GLMs are considered when the response variable satisfies the following conditions: count data expressed as proportions, for example logistic regressions count data that are not proportions, for example, log linear models of count binary response variables, for example dead or alive data on time-to-death where the variance increases faster than linearly with the mean, for example time data with gamma errors (see Crawley, 2005).

The most appropriate link function is the one which produces the minimum residual deviance (see Crawley, 2005). GLMs allow the specification of variety of different errors; Poisson errors, useful with count data; binomial errors, useful with data on proportions; gamma errors, useful with data showing a constant coefficient of variation; exponential errors, useful with data on time-to-death series. GLMs fit the data using categorical or continuous explanatory variables by specifying one of a family of error structures (for example, Poisson for count data or binomial for proportion data) and a particular link function (see Crawley, 2005).

The relationship between the function of the expected value of the response variable and the explanatory variables can be written in the linear form as $g(u_i) = X_i^T \beta + \varepsilon_i$, where g is the link function, $u_i = E(Y_i)$, Y_i is the i^{th} response variable, X_i is the vector of size p that specifies the explanatory variables for the i^{th} value of Y, β is the vector of size p parameters and ε_i is the i^{th} residual $(Y_i - u_i)$, which is assumed to be normally distributed with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$ (see Maunder and Punt, 2004).

Previous work (Bigelow and Hoyle, 2008, 2009) tested several generalized linear models containing different types of predictor variables. Molina et al. (2011) selected the following model in standardizing CPUE for albacore tuna in the South Pacific:

$$\ln(C/E) = year\_quarter + month * lat^3 + lat5 * long5 + vessel\_id + \varepsilon,$$

where the response variable is the natural logarithm of CPUE, C is the catch and E is the effort.

The predictor variables were all considered as factors and lat represented the local latitude. Lat5 and long5 represented the latitude and longitude aggregated at 5 degrees resolution.

3.5 Results and Discussion

The data provided by the Fiji Fisheries Department was used to standardize CPUE for albacore tuna in Fiji. Only the vessels targeting albacore after the cluster analysis were used in the generalized linear model to standardize the CPUE. According to Bernasconi et al. (2015), GLMs allow to estimate CPUE with a linear combination of a set of explanatory variables. There were 2954 sets with 15 variables used in the GLM analysis. The main effects considered in the GLM analysis are year, quarter, month, albacore catch, hook set, latitude, longitude, weight, year-quarter, longitude and latitude aggregated by 1 degree resolution and vessel id. In the model, the interactions for the main effects considered were month with latitude cubed and latitude aggregated at 1 degree resolution with longitude aggregated at 1 degree resolution. The CPUE standardization for albacore did not allocate the data into different regions in Fiji's EEZ but used the whole region as one as done by Molina et al. (2011) in the case for CPUE standardization for albacore tuna in the South Pacific.

3.5.1 Removing outliers using boxplot

Thorough checks need to be done on the fit of a model to the data by using residuals and other statistics derived from the fit to look for outlying observations (see Nelder and McCullagh, 1989). A Box-and-Whisker plot is useful in visualizing the uncertainty of a variable (see de Vries and Meys, 2015).

The boxplot shown in Figure 15 presents the logarithmic CPUE.

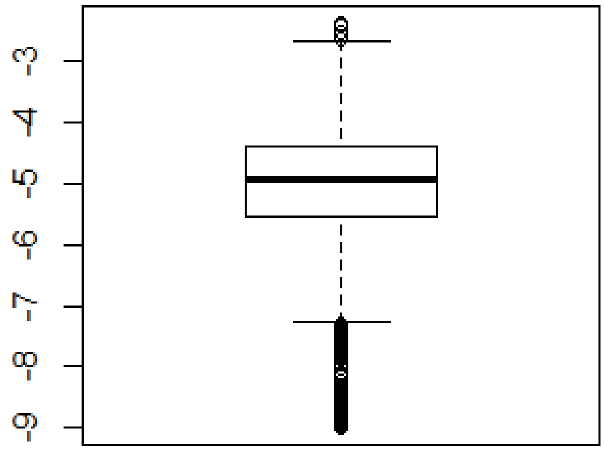


Figure 15: Boxplot to remove outliers

The bold horizontal bar in the middle of the box shows the median value of CPUE. The top of the box shows the 75th percentile and the bottom of the box shows the 25th percentile. interquartile range which is between about -4.5 and -5.5. The whiskers show the maximum and minimum values of CPUE. The summary of boxplot is presented in Table 9.

Table 9: The summary of boxplot

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-9.000	-5.545	-4.917	-5.044	-4.385	-2.367

Outliers are unusual points at extremes of the *x* range in a region where the model does not apply or there is a possibility that the values may be wrong, the result of misrecording or errors. From the boxplot the outliers identified were the values more than -2.645 and less than -7.285. There were more small outliers than the large outliers in the dataset. Removing the outliers, the data was reduced to 2861 sets, 284 trips and 69 vessels for developing the GLM model. There were 16 sets with zero albacore catch with tuna catch. Also there were four sets with no albacore and tuna catch.

3.5.2 Model Selection

Several GLMs containing different types of predictor variables were investigated to develop the models for the Fiji Fisheries. The GLMs that were considered in the study as well as their AIC values together with pseudo R^2 values are presented in the Table 10.

Table 10: AIC and pseudo R squared values of the GLM models tested

GLM Model	$\boldsymbol{R^2}$	**AIC value**
$\ln(C/E) = year + \varepsilon$	0.18	6648
$\ln(C/E) = catch\_month + \varepsilon$	0.58	7040
$\ln(C/E) = year\_quarter + \varepsilon$	0.25	6442
$\ln(C/E) = vessel\_id + \varepsilon$	0.27	6440
$\ln(C/E) = month{*}lat^3 + \varepsilon$	0.25	6442
$\ln(C/E) = lat1{*}long1 + \varepsilon$	0.31	6363
$\ln(C/E) = year\_quarter + month + \varepsilon$	0.28	6337
$\ln(C/E) = year\_quarter + month + vessel\_id + \varepsilon$	0.44	5771
$\ln(C/E) = year\_quarter + month + lat1{*}long1 + \varepsilon$	0.45	5838
$\ln(C/E) = year\_quarter + month + lat1{*}long1 + vessel\_id + \varepsilon$	0.52	5604
$\ln(C/E) = year\_quarter + month{*}lat^3 + \varepsilon$	0.38	5991
$\ln(C/E) = year\_quarter + month{*}lat^3 + lat1{*}long1 + \varepsilon$	0.48	5766
$\ln(C/E) = year\_quarter + month{*}lat^3 + vessel\_id + \varepsilon$	0.48	5630
$\ln(C/E) = year\_quarter + month{*}lat^3 + lat1{*}long1 + vessel\_id + \varepsilon$	0.54	5543
$\ln(C/E) = year\_quarter + lat1{*}long1 + vessel\_id + \varepsilon$	0.50	5702
$\ln(C/E) = year\_quarter + month + vessel\_id + \varepsilon$	0.43	5771

Akaike Information Criteria (AIC) provide a measure of the fit of the model. The general assumption when comparing two models is that the lower the AIC value, the better is the fit of the model. AIC use a penalty of two per parameter and for a given model, it is calculated as $\text{AIC} = deviance + 2p$ where p refer to the parameters used in the model. In case where the deviance goes down by less than two, the inclusion of extra the parameter is not justified (see Crawley, 2015).

In model selection, the AIC value is useful but sensitive and limiting the number of variables requires another criterion. There is a recommendation to test only variables for which there is good *a priori* reason to expect a relationship based on expert understanding of the data and fishery. In cases where there is a concern with oversensitivity, it is better to use the AIC and apply an R^2 criterion to reduce the sensitivity (see Hoyle et al., 2014). Pseudo R^2 or adjusted R^2 is mainly used to measure the goodness of fit for GLMs. In using R^2 as improvement from null to the fitted model, a smaller ratio indicates a better improvement for pseudo R^2. (see Mbachu et al., 2012). The smaller ratio implies a higher Pseudo R^2 when subtracted by 1. Pseudo R^2 was calculated using the formula:

$$Pseudo\ R^2 = 1 - \frac{Residual\ deviance}{Null\ deviance}.$$

The exclusion of outliers produced an improvement in the goodness of fit. The two models with the minimum AIC from Table 10 were as follows:

$$\ln(C/E) = year\_quarter + month*lat^3 + lat1*long1 + vessel\_id + \varepsilon. \quad (3.1)$$

and

$$\ln(C/E) = year\_quarter + month + lat1*long1 + vessel\_id + \varepsilon. \quad (3.2)$$

The dependent variable in the GLM was the natural logarithm of albacore CPUE with a small constant 0.5 added to the catch to allow for zero values, *C* is the catch and *E* is the effort. According to Maunder and Starr (2003), natural logarithmic CPUE ensure that CPUE is generally log normally distributed. The predictor variables were all considered as factors and *lat* represented the local latitude, *lat1* and *long1* represented the latitude and longitude aggregated by 1 degree resolution. Each longline set was weighted by $\frac{1}{\sqrt{number\ of\ longline\ sets\ per\ trip}}$ because individual longline sets within a trip are highly correlated (Molina et al., 2011; Bigelow and Hoyle, 2012) and the weightings given to each set prevented the changes in the spatial distribution of effort from biasing CPUE indices (see McKechnie et al., 2014).

The model in (3.1) was similar to the model selected by Molina et al. (2011) to develop a spatially adjusted CPUE for the albacore fishery in the South Pacific. The only difference was that in this study, latitude and longitude were aggregated by 1 degree resolution as only Fiji's EEZ was considered. The R^2 for this model is 0.54. The coefficients produced for the interaction of month with latitude cubed has very high values compared to other predictor variables. The use of this coefficient would give big values when predicting standardized CPUE. This would have significantly increased or decreased the standardized value of the CPUE.

The model in (3.2) is similar to model (3.1) except that it does not take into account the effects of cubic value of latitude. The R^2 for this model is 0.52. The coefficients produced by this model are more reliable. Thus, the model (3.2) is the best model that can be selected to develop a spatially adjusted CPUE for the albacore fishery in Fiji. Although the model (3.2) has a slightly lower pseudo R squared value and a slightly higher AIC value than (3.1), the coefficients produced by the model (3.2) were more reliable. Since in this study only looks at Fiji's EEZ there was lesser impact of the latitudinal effect compared to the study for the South Pacific by Molina et al. (2011).

The lognormal components and the full GLM is given by

$$\log y_i \sim \text{Normal}(\log u_i, \delta^2),$$

where $\log u_i = \beta_0 + \beta_{year-quarter(i)} + \beta_{(month)(i)} + \beta_{(lat1*long1)(i)} + \beta_{vessel\_id(i)}$,

where $\log y_i$ is the log CPUE and $\log u_i$ is the expected CPUE which was the albacore catch in numbers divided by the hook sets.

3.5.3 Diagnostic Tests for Selected Model

The selected model has to be evaluated to determine how well the model fits the fisheries data that was available and whether the data meets the assumptions of the model. There are various diagnostic plots to do that and in this study we have chosen the common diagnostic plots used with fisheries data. These include Q-Q plot and th e different residual plots. For normal models, the dependent variate can be expressed in the form $y = u + (y - u)$, that is

data = fitted value + residual.

Residuals can be used to explore the adequacy of fit of a model, in respect of choice of variance function, link function and terms in the linear predictor. Q-Q plot is a common diagnostic plot as shown in Figure 16.

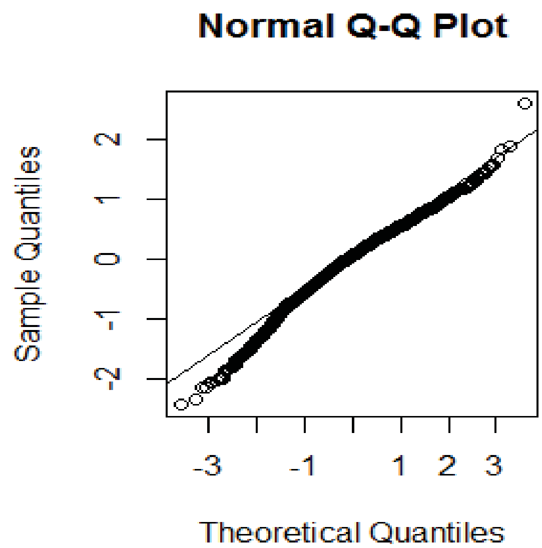


Figure 16: Q-Q Plot of the residuals of GLM model

A Q-Q plot is a plot for the test of normality. It plots the ranked samples from the distribution against a similar number of ranked quartiles taken from a normal distribution (see Crawley, 2015). Figure 16 shows the Q-Q plot of the residuals of GLM model selected. It reveals that the assumption of normality looks sufficient from the plot. There is some deviation at the lower tail.

Residual Plot

One of the common diagnostic tools in CPUE standardization is the residual plot shown in Figure 17.

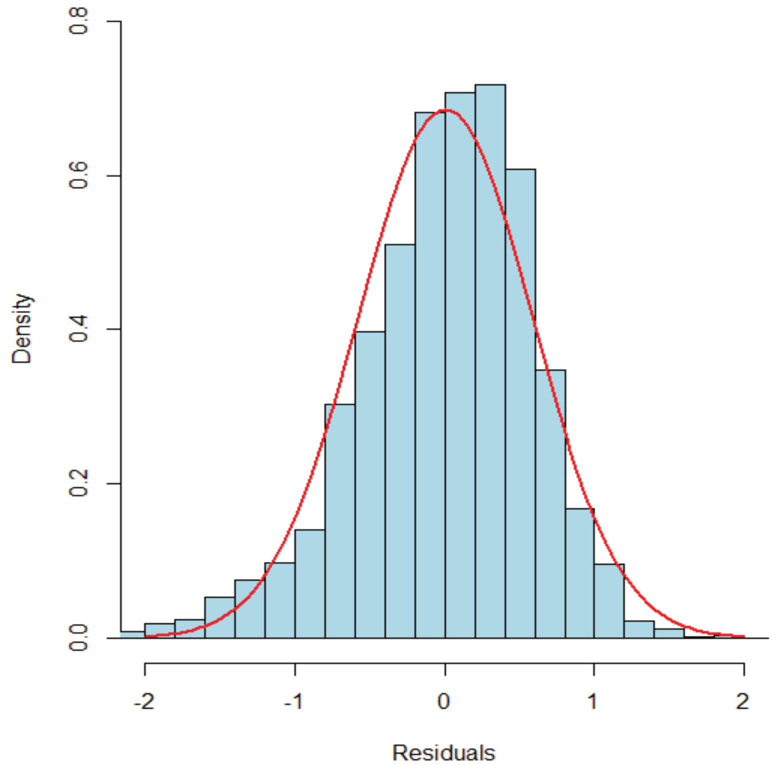


Figure 17: Histogram Showing Residual Plot with Normal Curve

Furthermore, a histogram of the residuals is presented in Figure 17 to see the distribution of the residuals. The histogram shows that the overall pattern of residuals conforms to the bell-shaped assumption of normally distributed data.

Residuals plotted against the fitted values are a common diagnostic plot used with Fisheries data.

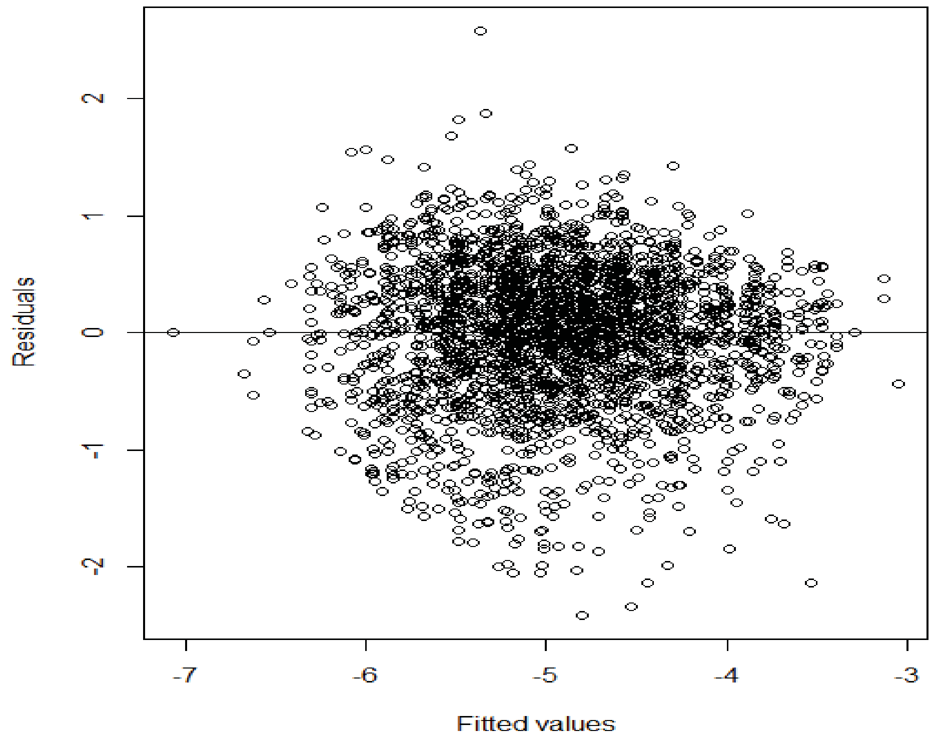


Figure 18: Residual Plot of the fitted values

The residual plot in Figure 18 shows the residuals on the y- axis against the fitted values on the x- axis. The positive values for the residuals indicate the prediction was too low, negative values mean the prediction is too high, and zero means the prediction was exactly correct. The distribution generally shows normal distribution and looks appropriate for the constancy of variance. The horizontal line at zero helps to determine if the points are symmetrical. If the fit is good, the pattern expected is null indicating no relationship between residuals and fitted values (see Nelder and McCullagh, 1989).

Absolute standardized residuals plotted against the fitted values are also a common diagnostic plot used with Fisheries data.

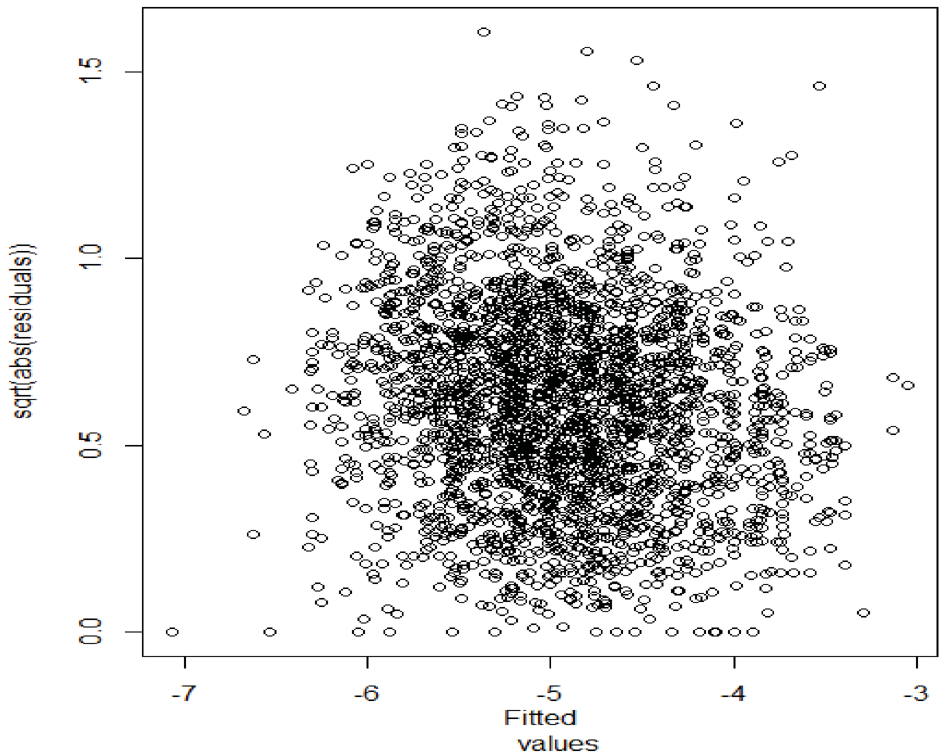


Figure 19: Plot of absolute standardized residuals

In Figure 19, a plot of absolute standardized residuals is also shown. This plot is similar to the residual plot but built on a different scale. It shows the square root of absolute residuals (where all values are positive) against the fitted values. It can be seen that the distribution of the variance is well behaved. It is sufficient to be unbiased and homoscedastic. Generally, the pattern looks good and if there was a problem, the points would be distributed inside a triangular shape with scatter of residuals increasing as the fitted value increases (see Crawley, 2015). A plot of the absolute residuals against fitted values gives a check on the adequacy of the assumed variance function. The null pattern shows no trend (see Nelder and McCullagh, 1989).

Natural logarithm of CPUE plotted against the fitted values is another common diagnostic plot used with Fisheries data.

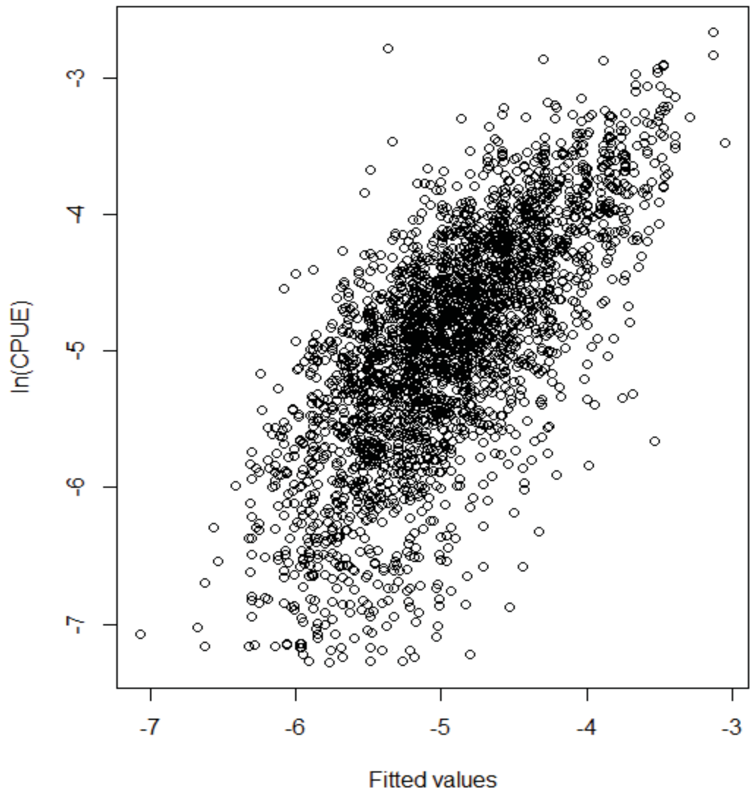


Figure 20: Fitted values and logarithmic CPUE

Finally, Figure 20 shows the fitted values on the x-axis and ln(CPUE) on the y-axis. It also indicates a strong linear relationship between the two variables.

3.5.4 Deviance Residuals of the Selected Model

Deviance residuals calculate the deviance associated for every data point in the model. It is a measure of goodness of fit of a generalized linear model. Higher numbers indicate worse fit. There are two forms of deviance. The null deviance shows how well the response variable is predicted by a model that only includes the intercept (grand mean). The output for the deviance residuals is the non-parametric description of the distribution.

In the selected model (3.2) for Fisheries in Fiji, the main effects are year, quarter, month, longitude and latitude aggregated by 1 degree resolution and vessel id. In the model, one interaction for the main effects is also included namely; $lat1 * long1$.The results from the GLM are given in the Table 11 below.

Table 11: Deviance analysis of residuals

Null deviance	601.36 on 2860 degrees of freedom
Residual deviance	290.56 on 2612 degrees of freedom
AIC	5603.5
Number of Fisher Scoring iterations	2

From the results, it can be seen that the deviance of the null model is 601.36 on 2860 degrees of freedom. By including the predicted variables year-quarter, month, latitude and longitude aggregated at 1 degree resolution, and vessel id, the deviance decreased to 290.56 on 2612 degrees of freedom which is a significant reduction in deviance. The residual deviance has decreased by 289.2 with a loss of 248 degrees of freedom. The Fisher's Scoring Algorithm, is a derivative of Newton's method for solving maximum likelihood problem numerically. The Fisher's Scoring Algorithm in Table 11 indicates that the solution was found in R in two iterations and this implies that the model converges with no problem. The residual deviance (290.56) is not larger than the residual degrees of freedom (2612) so there is no need to correct for overdispersion (see Crawely, 2015).

3.5.5 Model Summary

Appendix C shows the summary of the selected model. The summary of the selected model shows the estimate of coefficients of each explanatory variable as well as their standard error, *t* value, and significance. There are 66 coefficients not defined because of singularities and as such there are 66 coefficients with NA (missing value) in the table of the linear predictor. This indicates that there are no data in the dataframe to estimate the explanatory variable month for months 6, 9, and 12. As well as there are no data in the dataframe to estimate the interaction between latitude and longitude aggregated by 1 degree resolution for 63 interactions. In the summary of the model, the intercept was significant which indicated that the mean number of cells for year-quarter, catch-month, vessel id are greater than zero. This implies that all the categorical explanatory variables are non-negative.

3.5.6 Analysis of Deviance

After fitting the model, we can test the overall model fits regarding a subset of regression parameters.

Table 12: Presents an Analysis of Deviance factor

Factor	Df	Deviance Difference	Residual Df	Residual Deviance	% of deviance explained by each factor	Pr (> Chi)
NULL			2860	601.36		
as.factor (year_qtr)	44	152.175	2816	449.18	48.96	< 2.2e-16 ***
as.factor (catch_month)	8	18.594	2808	430.59	5.98	< 2.2e-16 ***
as.factor (lat1)	15	43.894	2793	386.70	14.12	< 2.2e-16 ***
as.factor (long1)	11	21.803	2782	364.89	7.02	< 2.2e-16 ***
as.factor (vessel_id)	68	50.015	2714	314.88	16.09	< 2.2e-16 ***
as.factor (lat1): as.factor (long1)	102	24.313	2612	290.56	7.82	1.757e-10 ***
Total explained deviance	248	310.794			10.87	

Df refers to the degrees of freedom. The coefficient of variable is significant if $p <$ 0.05. The estimates (coefficients of the predictors variables; year_qtr, catch_month, latitude and longitude aggregated by 1 degree resolution, vessel_id) are in units called logits.

The number of parameters of the selected model was 248. Every parameter that was included in the model resulted in large decrease of the AIC calculated similar to the results of de Lima et al. (2004). The lognormal model explained 10.9% of the variability. This was calculated by dividing the total of deviance difference by null residual degrees of freedom. The percentage of deviance explained by each factor was calculated by dividing deviance difference of each factor by the total explained deviance. The most important factor was year-quarter explaining 48.96% of the total variability followed by vessel id and latitude aggregated by one degree resolution which explained 16.09% and 14.12%, respectively. Catch-month represented 5.98%, and longitude aggregated by one degree resolution represented 7.02%. The interaction of latitude aggregated by one degree resolution with longitude aggregated by one degree resolution represented 7.82%. The R^2 for this model was 0.52 and all the explanatory variables are highly significant.

The analysis of variance, when mainly applied to orthogonal data with normal errors, is a highly useful technique for screening the effects of factors and their interactions. The null model only contains the intercept. If the deviance is used as a measure of discrepancy of a GLM, then each unit contributes a quantity d_i to that measure so that $\sum d_i = D$. If we define

$r_D = \text{sign}\ (y - u)\sqrt{d_i}$, we have a quantity that increases with $(y_i - u_i)$ and for which

$\sum r_D^2 = D$ (see Nelder and McCullagh, 1989).

3.5.7 Error Structure

Lognormal model was used where the logarithm of positive catch rate (log CPUE) is assumed to be normally distributed and an identity link function was used. A specialization of the linear model involves the assumption that the errors follow a normal distribution with constant variance δ^2 , mathematically shown as $\varepsilon \sim N(0, \delta^2)$ (see McCullagh and Nelder, 1989; Wang et al., 2008; Goni and Arrizabalaga, 2005). Since there were few sets with zero albacore catch or tuna catch, the lognormal distribution was used to fit CPUE data (see Hoyle et al., 2014). The residual patterns are not far from expected under the normal error distribution assumption indicating a reasonably good fit (see Gulati and Premchand, 2015).

3.5.8 CPUE Standardization

The nominal CPUE is seldom proportional to stock abundance once the time series and entire geographical range particularly for non-target species because several factors affect catch rates. Standardization of CPUE removes the effects of these factors. Standardized CPUE compromised of exponentiated year-quarter coefficients as used by many Fisheries journals including Gulati and Premchand (2015) and Langley (2006). To standardize the CPUE, a new data frame was set up in R software for each year quarter in the model. The values used for other predictor variables used in the CPUE standardization were kept at the most common level. The values from the first row were used in the CPUE standardization. This was tested by choosing other rows apart from the first row. The results were same when the standardized CPUE was normalized. The model prediction of CPUE was done on the log scale. The standardized CPUE was normalized by dividing with its mean. This is how the standardized CPUE are usually presented. Appendix D shows the R codes that were used to identify domestic vessels targeting albacore and using GLM to standardize CPUE.

Table 13 and Figure 21 show the nominal CPUE per 1000 hooks and standardized CPUE for the entire time series by year-quarter.

Table 13: Nominal and standardized CPUE

Year - Quarter	Catch (albacore number)	Effort (Hook - Set)	Nominal Albacore CPUE per 1000 hooks	Standardized CPUE
2003	108	17025	0.11	0.45
2003.0.25	232	28750	0.23	1.66
2003.0.5	1492	108958	1.49	0.97
2004	559	94200	0.56	1.29
2004.0.25	286	16625	0.29	1.11
2004.0.5	762	58065	0.76	1.07
2004.0.75	148	10070	0.15	0.42
2005	828	98128	0.83	1.11
2005.0.25	2679	176571	2.68	0.98
2005.0.5	5676	344036	5.68	1.33
2005.0.75	1830	161724	1.83	0.87
2006	925	86379	0.93	0.32
2006.0.25	6660	354693	6.66	1.07
2006.0.5	4618	308034	4.62	0.62
2006.0.75	1256	83686	1.26	0.70
2007	1397	121964	1.40	0.50
2007.0.25	1642	174201	1.64	1.17
2007.0.5	1944	140268	1.94	0.61
2007.0.75	1702	130539	1.70	1.01
2008	644	65143	0.64	0.64
2008.0.25	2320	262434	2.32	0.98
2008.0.5	1022	106242	1.02	0.50
2008.0.75	2909	272808	2.91	0.63
2009	552	51724	0.55	1.63
2009.0.25	2426	235940	2.43	0.83
2009.0.5	1089	104535	1.09	1.62
2009.0.75	706	74784	0.71	1.12
2010	332	51179	0.33	1.14
2010.0.25	678	115693	0.68	2.62
2010.0.5	259	37912	0.26	1.48
2011	46	8877	0.05	0.91
2011.0.25	391	63690	0.39	0.92
2011.0.75	80	16326	0.08	1.45
2012	220	44604	0.22	0.71
2012.0.25	1026	79155	1.03	0.91
2012.0.5	286	42030	0.29	0.87
2012.0.75	97	26173	0.10	0.32
2013	674	154189	0.67	1.29
2013.0.25	3555	364934	3.56	0.62
2013.0.5	2161	340221	2.16	0.40
2013.0.75	1186	167415	1.19	0.43
2014	3218	370770	3.22	0.77
2014.0.25	1677	167388	1.68	0.82
2014.0.5	5835	720749	5.84	3.18
2014.0.75	6625	833966	6.63	0.93

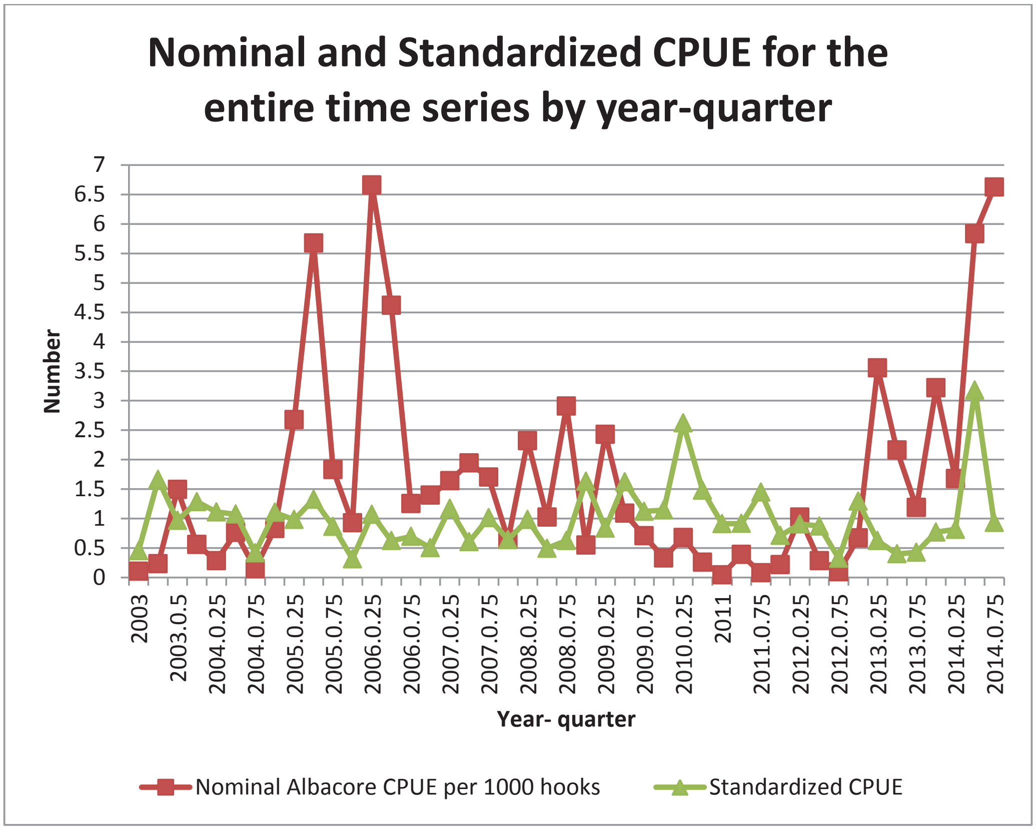


Figure 21: Plot of Nominal CPUE per 1000 hooks and Standardized CPUE

Table 13 and Figure 21 reveals that standardized CPUE was lower than nominal CPUE per 1000 hooks for 25 quarters in the time series. There were 20 quarters with higher standardized CPUE than nominal CPUE per 1000 hooks. Generally, there were more fluctuations in nominal CPUE than the standardized CPUE. The nominal CPUE per 1000 hooks had sudden increases in quarter 3, 2005, quarter 2, 2006, quarters 3 and 4, 2014. Maximum nominal CPUE was in quarter 4, 2014 and the minimum was in quarter 1, 2011. The nominal CPUE per 1000 hooks increased towards the end of time series in the study.

The standardized CPUE of 0.45 albacore in quarter 1, 2003 increased to 1.66 tuna in quarter 2. Generally there was a decreasing trend seen till quarter 4, 2004 to 0.42 albacore. An increasing trend was seen till quarter 3, 2005 with 1.33 albacore. Then a decreasing trend was seen till quarter 1, 2006 with 0.32 albacore. An increase was seen in quarter 2, 2006 with 1.07 albacore. There was a decreasing trend seen till quarter 1, 2007 with 0.50 albacore. There were fluctuations in standardized CPUE seen from quarter 2, 2007 till quarter 2, 2008. The standardized CPUE fluctuated between 1.17 to 0.61 albacore during this period. There was decrease in quarter 3, 2008 with 0.50 albacore. The standardized CPUE increased till quarter 1, 2009 with

1.63 albacore. There was a decrease in quarter 2, 2009 to 0.83 albacore. The standardized CPUE increased in quarter 3, 2009 to 1.62 albacore. A decrease to 1.12 albacore was seen in quarter 4, 2009. There was slight increase in quarter 1, 2010 to 1.14 albacore. In quarter 2, 2010 there was a big increase to 2.62 albacore. Generally, a decreasing trend was seen till quarter 4, 2012 to 0.32 albacore. There was an increase in quarter 1, 2013 to 1.29 while there was a decrease till quarter 3, 2013 to 0.40. There was an increase in standardized CPUE till quarter 3, 2014 reaching a maximum of 3.18 albacore in the entire time series. There was significant decline in standardized CPUE seen in quarter 4, 2014 to 0.93.

The standardized CPUE had sudden increases in quarter 2, 2010 and quarter 3, 2014. Maximum standardized CPUE was in quarter 3, 2014 indicating abundance and the minimum was in quarter 1, 2006. The standardized CPUE generally, fluctuated during the time series with decline from 2010 to 2013 indicating over-exploitation. In quarter 3, 2014 there was a sudden increase in standardized CPUE indicating sustainable management of albacore species. However, there was a decline seen in the standardized CPUE in quarter 4, 2014 indicating over- exploitation. The standardized CPUE plot indicated that the standardized trend was less variable than nominal, similar to the findings of Gulati and Premchand (2015). The standardized CPUE plot is more stable than the nominal CPUE for albacore. This indicated that standardization removed some variability attributable to the explanatory variables. Table 14 and Figure 22 show the nominal CPUE per 1000 hooks and standardized CPUE for the entire time series by year.

Table 14: Nominal and standardized CPUE by year

Year	Catch (albacore number)	Effort (Hook - Set)	Nominal Albacore CPUE per 1000 hooks	Standardized CPUE
2003	108	17025	0.61	1.03
2004	559	94200	0.44	0.97
2005	828	98128	2.75	1.07
2006	925	86379	3.36	0.68
2007	1397	121964	1.67	0.82
2008	644	65143	1.72	0.69
2009	552	51724	1.19	1.30
2010	332	51179	0.42	1.75
2011	46	8877	0.17	1.09
2012	220	44604	0.41	0.70
2013	674	154189	1.89	0.69
2014	3218	370770	4.34	1.42

Figure 22 shows the nominal albacore CPUE per 1000 hooks and standardized CPUE by year throughout the entire time series used in this study.

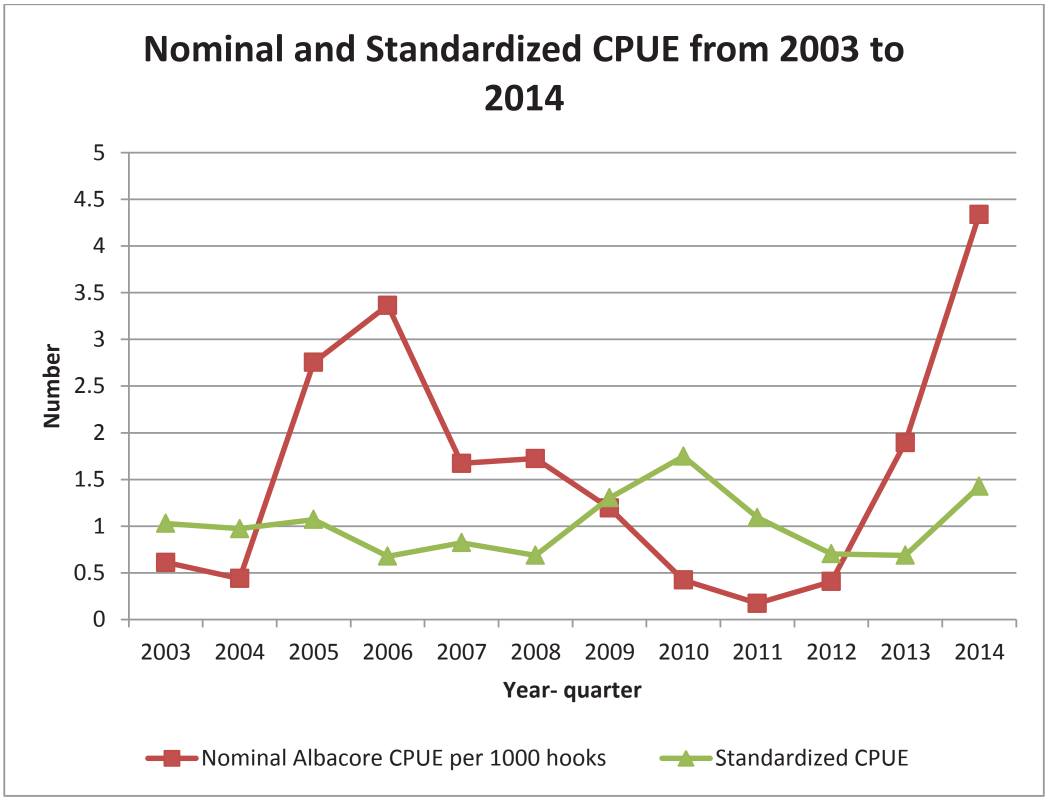


Figure 22: Plot of Nominal CPUE per 1000 hooks and Standardized CPUE aggregated by year

Table 14 and Figure 22 reveal that the nominal CPUE shows more variability than the standardized CPUE. The nominal CPUE decreased from 0.61 albacore in 2003 to 0.44 albacore in 2004. There was an increase to 3.36 albacore till 2016. Generally, a decline was seen till 2011 to 0.17 albacore. Since then the nominal CPUE increased till 2014 to 4.34 albacore. The standardized CPUE showed fluctuation from 2003 to 2008 of around 1.07 to 0.68 albacore. There was an increase seen in 2010 to 1.75 albacore. There was a significant decrease till 2013 to 0.69 albacore. In 2014, the standardized CPUE increased to 1.42 albacore. Generally, the standardized CPUE trend for albacore fishery was similar to the study by Molina et al. (2011) for the albacore fishery in the South Pacific, except for years 2006, 2007, and 2010.

3.6 Conclusion

This section dealt with using generalized linear model to standardize CPUE. Outliers were removed before the different GLMs were tested. Diagnostic tests were performed on the selected model. Upon investigation it was found that the selected GLM model conformed to normality and can be used as an indicator of indexing for abundance of albacore tuna in Fiji.

Chapter 4

Conclusion

Fisheries studies frequently use CPUE as an indicator for the level of stock (see Haputhantri et al., 2011). According to Maunder, et al. (2006), CPUE based analysis provide useful information about fisheries changes, discussion points, and indicate areas of future research. The standardized CPUE have been used as an index of relative abundance for albacore tuna (see Molina et al., 2011).

Due to some logistical issues there was considerable delay in getting the fisheries data. There were also challenges in mastering the techniques to standardize the CPUE of albacore tuna.

There were enormous data sets provided by the Fiji Fisheries Department which required considerable time to summarize into a single table to be used in R software for analysis. Gibson (2004) discussed that while the assessment of tuna was an issue, Fiji had been very efficient in data collection.

In this thesis, we used hierarchical and partitioning clustering techniques were used to identify the domestic fleets that targeted albacore in Fiji's EEZ. This ensured that only the effort to catch albacore was considered to avoid bias. A standardized CPUE index was developed for longline albacore fishery in Fiji using GLM.

Generally, the clusters targeting albacore were mostly on Fiji's eastern longitude and the variability in latitude decreased for the albacore fisheries from 2003 to 2014.

The CPUE is a standard tool used by biologists to determine developments in fish stocks and used by economists as an indicator for the efficiency of the fishing operation (see Hoof and Salz, 2001). Standardized CPUE estimates changes in fish abundance and provides fish stock health to the fisheries managers.

The changes in CPUE imply changes to albacore tuna's abundance in Fiji. A decreasing CPUE indicates overexploitation while an unchanging CPUE indicate sustainable harvesting. Generally, there were no major fluctuations seen in the standardized CPUE for the entire time series. In the case of nominal CPUE per 1000 hooks, there were spikes seen in quarter 3, 2015, quarter 2, 2006 and quarter 3, 2014. Generally, the standardized CPUE was lower than the nominal CPUE per 1000 hooks. The standardized CPUE oscillated between 0.32 to 3.18 albacore tuna caught. After fluctuating and declining trends, there was an increase in standardized CPUE in quarter 3, 2014. The increase in standardized CPUE for albacore indicates that healthy albacore stock was maintained. In quarter 4, 2014 although nominal CPUE increased, the standardized CPUE decreased indicating over-exploitation. Generally, the standardized CPUE indicates lack of consistency in sustainability management of albacore in Fiji. There has been sustainable harvesting and over-exploitation seen when considering the entire time series. Considering the decline in standardized CPUE seen in quarter 4, 2014 there may be a need to set total allowable catch for albacore and monitor the limit on the number of licenses issued to fishing vessels.

Regression diagnostic statistics suggested an acceptable performance of GLM and an agreement with the error structure assumed. The accuracy of the catch and effort information affects the calculation to monitor fish stock.

The standardized CPUE indices were used to infer the trends in the variation of the albacore abundance during the time series. CPUE provides useful information for the proper sustainable management of albacore stock in Fiji and combating illegal, unregulated and unreported fishing with strictest adherence to national, regional and international instruments.

In this study the environmental variables and vessel information, skippers experience, changes in crews or vessels were not readily available so they were not included in the analysis. The inclusion of these variables would have improved the estimates of the coefficients.

Indices of abundance play an important role in fisheries management and are commonly used to tune stock assessment models. In tuna fisheries, an index of relative abundance (CPUE) is used for management or for fitting single species stock assessment models (see Molina et al., 2011).

SEAPODYM is a physical – biological – fisheries model with spatial structure which has been developed in the South Pacific. With the inclusion of environmental variables, this index will be suitable for use in SEAPODYM model for albacore and is expected to provide further improvements to the estimates generated by SEAPODYM. Since the whole region was used, this would allow the usage of

SEAPODYM to predict the albacore migration patterns and zooming at Fiji's EEZ (see Molina et al., 2011).

The results allowed us to evaluate which are the factors of the fishery that affect the CPUE of albacore tuna. The selected methods can influence the resulting abundance index, the stock status and ultimately management advice (see Hoyle et al., 2014).

Similar CPUE standardization technique could be applied to other species of tuna including bigeye and yellowfin. This technique could be extended to other regional countries as well. Based on the nature of the data for other species, there could be variations in the model selected to standardize CPUE.

Bibliography

[1] Ayramo S. and Karkkainen, T. (2006). Introduction to partitioning based clustering methods with a robust example, *Reports of the Department of Mathematical Information Technology,*no. C.1/2006, 1-34.

[2] Bailey, T.C. and Gatrell, A.C. (1995). Interactive Spatial Data Analysis. *Longman Scientific and Technical*.

[3] Bentley, N., Kendrick, T. H., Starr, P. J. and Breen, P. A. (2012). Influence Plots and Metrics: Tools for better understanding Fisheries Catch-Per-Unit-Effort Standardizations. *ICES Journal of Marine Science*, 69(1), 84-88.

[4] Bernasconi, J. F., Perier, M. R. and Di Gia'como, E. E. (2015). Standardized Catch Rate of Cockfish, Callorhinchus Callorynchus, in a bottom trawl fishery of Patagonia: Is it possible its use as a predictor of abundance trend. *Brazilian Journal of Oceanography*, 63 (2), 14.

[5] Bigelow, K. and Hoyle, S. (2008). Standardized CPUE for Distant-Water Fleets Targeting South Pacific Albacore. *Western and Central Pacific Fisheries Commission*, no. WCPFC SC4 ME-WP-3, 1-23.

[6] Bigelow, K. and Hoyle, S. (2009). Standardized CPUE for Distant-Water Fleets Targeting South Pacific Albacore. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC5-2009/ SA-WP-5, 1-12.

[7] Bigelow, K. and Hoyle, S. (2012). Standardized CPUE for South Pacific Albacore. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC8-2012/ IP-14, 1-12.

[8] Bigelow, K. (2006). Incorporation of other Oceanographic Factors into CPUE Standardizations. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC2-2006/ME WP-2, 1-16.

[9] Bromhead, D., Hoyle, S., Williams, A., Wang, S. B. and Chang, S. K. (2009). Factors influencing the size of Albacore Tuna sampled from the South Pacific Albacore Longline Fisheries. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC5-2005/SA-IP-05,1-67.

[10] Chang, S.K., Yuan, T. L. and Hoyle, S. (2011). Standardizations of Taiwanese Distant-Water Longline CPUE up to 2010 for Yellowfin and Bigeye Tunas in Region 6 of WCPO. *Western and Central Pacific Fisheries Commission*, no. WCPFC- SC7-2011/ SA- IP-11, 1-15.

[11] Coelho, R., Nikolic, N., Evano, H., Santos, M. N. and Bourjea, J. (2014). Reunion Island Pelagic Longline Fishery Characterization and Standardization of Albacore Catch Rates. *IOTC*, 5 (12), 36.

[12] Cohen, S. S. (1988) .Practical Statistics. J.W. Arrowsmith Ltd., Great Britain, 86-90.

[13] Crawley, M. J. (2005). Statistics an Introduction using R, John Wiley & Sons Ltd.

[14] Crawley, M. J. (2015). Statistics an Introduction using R (2nd ed.), John Wiley & Sons Ltd., New York.

[15] de Vries, A. and Meys, J. (2015). R for Dummies. John Wiley and Sons, Inc., New York.

[16] Everitt, B. (1974). Cluster Analysis. Heinemann Educational Books.

[17] Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013). Regression Models, Methods and Applications. Springer, New York , 301-309.

[18] Gabr, M. H. and El-Haweet, A. E. (2012). Pelagic Longline Fishery for Albacore (Thunnus Alalunga) in the Mediterranean Sea off Egypt. *Turkish journal of Fisheries and Aquatic Sciences*, *12* (4), 7.

[19] Gibson, D. (2014). Tuna Data Setback. *The Fiji Times*, 22nd April, 2014.

[20] Glazer, J. P. and Butterworth, D. S. (2016). GLM-based standardization of the catch per unit effort series for South African west coast hake, focussing on adjustments for targeting other species. *South African Journal of Marine Science*, 24, 323-339.

[21] Goni, N. and Arrizabalaga, H. (2005). Analysis of Juvenile North Atlantic Albacore (Thunnus alalunga) Catch Per Unit Effort by Surface Gears in Relation to Environmental Variables. *ICES Journal of Marine Science*, 64, 1475-1482.

[22] Gulati, D. K. and Premchand. (2015). Standardization of Distant Water Tuna Longline Hooking Rate for Yellowfin Tuna (Thunnus albacares) from Fishery Survey of India Fleet (1981-2012). *IOTC*, 17-24.

[23] Haputhantri, S. S. K., Moreau, J. and Lek, S. (2011). Exploring Gillnet Catch Efficiency of Sardines in the Coastal Waters of Sri Lanka by Means of Three Statistical Techniques: A Comparison of Linear and Nonlinear Modelling Techniques. *Journal of Applied Statistics*, *36* (2), 13.

[24] Hazin, F. H. V., Hazin, H.G., Travassos, P. and Oliveira, I.D.M. (2007). Standardized Catch Per Unit Effort of White Marlin, *Tetrapturus Albidus*, and Blue

Marlin, *Makaira Nigricans*, Caught by Brazilian Tuna Longline Fleet. *ICCAT* , 60 (5), 11.

[25] He, X.,Bigelow, K. A. and Boggs, C. H. (1997). Cluster Analysis of Longline Sets and Fishing Strategies within the Hawaii-Based Fishery.*Elsevier,*147-158.

[26] Hilborn, R, and Walters, C.J. (2001). Quantitative Fisheries Stock Assessment Choice, Dynamics and Uncertainity. *Kluwer Academic Publishers,* SWG-10-JM-06, 1-25.

[27] Hoof, L. V. and Salz, P. (2001). Applying CPUE as a Management Tool. Salerno: *EAFE* , 1.

[28] Hoyle, S. (2009). CPUE Standardization for Bigeye and Yellowfin Tuna in the Western and Central Pacific Ocean. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC5-2009/SA-WP-1, 1-45.

[29] Hoyle, S., Davies, N. and Chang, S. K. (2013). Analysis of Swordfish Catch Per Unit Effort Data for Japanese and Chinese Taipei Longline Fleets in the Southwest Pacific Ocean. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC9-2013/SA-IP-03, 1-26.

[30] Hoyle, S., Langley, A. D. and Campbell, R. A. (2014). Recommended Approaches for Standardizing CPUE Data from Pelagic Fisheries.*Western and Central Pacific Fisheries Commission*, no. WCPFC-SC10-2014/SA-IP-10, 1-21.

[31] Joseph, J. (2003). Managing Fishing Capacity of the World Tuna Fleet. *Fisheries and Aquaculture Department*, Rome, http://www.fao.org/fishery/fishtech/1010/en.

[32] Kell, L., Palma, C. and Prince, E. (2011), Standardization of Blue Marlin CPUE taking into account Habitat Compression. *ICCAT*, 66(4), 1738 - 1759.

[33] Kolody, D., Herrera, M. and Million,J. (2010). Exploration of Indian Ocean Bigeye Tuna Stock Assessment Sensitivities 1952-2008 using Stock Synthesis. *IOTC,*no. WPTT-04, 1-93.

[34] Langfelder, P., Zhang B. and Horvath S. (2008). Defining Clusters form a Hierarchical Cluster Tree: The Dynamic Tree Cut Package for R. *bioinformatics*, 5(24):719-720.

[35] Langley, A. D. (2006). Spatial and Temporal trends in Yellowfin and Bigeye Longline CPUE for the Japanese Fleet in WCPO. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC2-2006/ME IP-1, 1-19.

[36] Lee, L. K., Chang, F. C., Chen, C. Y., Wang, W. J. and Yeh, S. Y. (2012).

Standardized CPUE of Indian Albacore (Thunnus Alalunga) based on Taiwanese Longline Catch and Effort Statistics dating from 1980 to 2011. *IOTC*, 18, 14.

[37] Lee, S., Kim, Z. G., Lee, M. K., Ku, J. E., Park, H. W. and lee, D. W. (2014). CPUE Standardization of Bigeye Tuna caught by Korean Tuna Longline Fishery in the Indian Ocean. *IOTC*, 1-10.

[38] Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P. and Harley, S. J. (2006). Interpreting Catch Per Unit Effort Data to Assess the Status of Individual Stocks and Communities. I*CES Journal of Marine Science*, 63, 1373 - 1385.

[39] Maunder, M. N. and Punt, A. E. (2004). Standardizing Catch and Effort Data: A Review of Recent Approaches. 141-159.

[40] Maunder, M. N. and Starr, P. J. (2003). Fitting Fisheries Models to Standardized CPUE abundance Indices. *Elsevier*, 63, 43-50.

[41] Mbachu, H. I., Nduka,E.C. and Nja, M. E. (2012). Designing a Pseudo R-Squared Goodness-of-Fit Measure in Generalized Linear Models. *Journal of Mathematics Research* , 4(2), 148:154.

[42] McKechnie, S., Hoyle, S. and Harley, S. (2013). Longline CPUE series that account for changes in the spatial extent of fisheries. WCPFC-SC9-2013/SA-IP-05.

[43] McKechnie, S., Harley, S., Chang, S.K., Liu, H.I. and Yuan, T.L. (2014). Analyses of Longline Catch Per Unit Effort Data for Bigeye and Yellowfin Tunas. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC10-2014/SA-IP-03, 1-52.

[44] Meneses, J. H., Andrade, H. A., Fredou, F. L. and Travassos, P. (2001). Standardized CPUE for Albacore (Thunnus Alalunga) from the Brazillian Longline Fishery in the South Atlantic, from 1991 through 2001. *ICCAT*, 56 (4), 8.

[45] Moffitt, P. E. (2008). Method and Apparatus for Long Line and Recreational Bait Fishing Patent application number: 20080202013, http://www.patentsencyclopedia.com/app/20080202013#ixzz4a1vKs3Ap.

[46] Molina, J. J., K Bigelow, S Hoyle, S Nicol, and K Briand. (2011). Developing a Spatially adjusted CPUE for the Albacore Fishery in the South Pacific. *Western and Central Pacific Fisheries Commission*, no. WCPFC-SC7-2011/EB-WP-IP-05, 1-31.

[47] Mourato, B. L., Arfelli, C. A., Amorim, A. F., Hazin, H. G., Carvalho, F. C. and Hazin, F. H. V. (2011). Spatio- Temporal Distribution and Target Species in a Longline Fishery off the Southern Coast of Brazil. *Brazilian Journal of*

*Oceanography***,** 59 (2), 10.

[48] Nelder, J. A. and McCullagh, P. (1989).*Generalized Linear Models* (Second Edition), Chapman and Hall/CRC, USA.

[49] Nishida, T. and Chen, D. (2004). Incorporating Spatial Autocorrelation into the General Linear Model with an Application to the Yellowfin Tuna Longline CPUE Data. *Fish Res.,* 265-274.

[50] Okamoto, H., Miyabe, N. and Matsumoto T. (2001). GLM Analyses for Standardization of Japanese Longline CPUE for Bigeye Tuna in the Indian Ocean applying Environmental Factors. *IOT*C**,** 4, 32.

[51] Okamoto, H. Miyabe, N. and Matsumoto T. (2011). GLM Analyses for Standardization of Japanese Longline CPUE for Bigeye Tuna in the Indian Ocean applying Environmental Factors. *IOTC*, 4, 32.

[52] Pons, M. and Domingo, A. (2014). Update of Standardized CPUE of Albacore Tuna, *Thunnus Alalunga*, Caught by Uruguayan Longliners in the Southern Atlantic Ocean (1983 - 2012). *ICCAT* , 70 (3), 14.

[53] Romesburg, H. C. (1984). Cluster Analysis for Researches. Lifetime Learning Publications, California.

[54] Rousseeuw, P. J., Struyf, A. and Hubert, T. (1992) Clustering in an Object-Oriented Environment. Department of Mathematics and Computer Science, U.I.A., Universiteitsplein 1, B-2610 Antwerp, Belgium

[55] SPC (2010). A Community-Based Ecosystem Approach to Fisheries Management: Guidelines for Pacific Island Countries. Secretariat of the Pacific Community, Noumea, New Caledonia.

[56] Teo, S. L. H. and Block, B. A. (2010). Comparative Influence of Ocean Conditions on Yellowfin and Atlantic Bluefin Tuna Catch from Longlines in the Gulf of Mexico. *PLoS ONE***,** 5 (5), 11.

[57] Travassos, P., Hazin, H., Hazin, F., Mourato, B. and Carvalho, F. (2009). Standardization of a CPUE Series of Yellowfin Tuna, *Thunnus Albacares*, caught by Brazilian Longliners in the Southwestern Atlantic Ocean. *ICCAT*, 64 (3), 11.

[58] Wang, S. P., Semba, Y. and Nishida, T. (2008). CPUE Standardization of Swordfish (Xiphias gladius) caught by Taiwanese Longline Fishery in the Indian Ocean. *IOTC*, WPB-10, 1-18.

[59] Yeh, W. M. and Chang, S.T. (2013). CPUE Standardizations for Yellowfin Tuna Caught by Taiwanese Longline Fishery in the Indian Ocean using Generalized Linear Model. *IOTC*, *38, 13*.

Appendix A: Summary of Key Findings from relevant CPUE Analyses Conducted in the Western and Central Pacific Fisheries Commission

Broad Research Question	Geographic Area	Years (Data Coverage)	Data Source	Statistical Methods/ Techniques	Conclusions	Reference
To develop a spatial-temporal structured CPUE based on Fisheries data targeting albacore in the South Pacific.	South Pacific	1960 – 2010	Data was used for longline fishery for countries including Japan, Korea, Taiwan, American Samoa, Cook Islands, Fiji, New Caledonia, French Polynesia, Tonga, Vanuatu, and Samoa. Main species were albacore, yellowfin and bigeye. Canneries were Pagopago, American Samoa and Levuka, Fiji. Fleets used were domestic and distant water.	Cluster analysis was done to identify vessels targeting South Pacific albacore and GLMs were used to standardize CPUE. Analysis was done at 5*5 degree spatial resolution	SEAPODYM offers the opportunity of including the adjusted CPUE data to integrate biological and ecological knowledge of tuna species and their responses to fishing pressure.	"An adjusted CPUE index with spatial structure was developed for albacore (Thunnus alalunga) in the South Pacific based on operational logsheet data from distant-waters and domestic fleets" (see Molina et al., 2011).
To examine the	Western and	1952-2005	Catch and effort data	To analyse the spatial	There was	The Japanese

homogeneity of CPUE trends by undertaking an analysis of the time-series of CPUE data from the Japanese longline fleet at the finest level of spatial resolution (5 degree resolution).	Central Pacific		for the Japanese longline fleet was used for yellowfin and bigeye tuna in longline fishery. Data was aggregated by year, month and spatial resolution.	and temporal CPUE trends for yellowfin and bigeye tuna for the six model regions identified in the study. The analysis for each region involved the calculation of the nominal CPUE (number of fish per 100 hooks set) GLM model was used in core regions to include additional variables such as gear configuration and associated catch of the other species.	considerable spatial heterogeneity in the CPUE trend for both species within each of the main regions and the repeat of GLM model showed no substantive difference in the resulting trends in CPUE indices.	longline fleet data for the catch and effort for the yellowfin and bigeye stock assessments were used in a GLM to derive a standardized effort series for the key longline fisheries in each region (see Langley, 2006).
To develop standardized CPUE indices of distant-water fleets targeting albacore in the South Pacific Ocean.	South Pacific	1960 - 2007	Data on catch and effort were compiled from individual distant water vessels landing in Pagopago, American Samoa and Levuka, Fiji. Vessel flags were from Japan, Korea and Taiwan.	GLMs were used and vessels were analysed that consistently participated in albacore fishery. Poisson distribution was also used with the dependent variable as catch and	The study refined the assessment indices of CPUE by defining albacore targeting fleets and applying standardization methods especially with the use of	The approach used to approximate the GLM was similar to approaches used for yellowfin and bigeye in the WCPO by Langley et al. in

				effort (hooks) as an offset. There were a total of 12 GLMs conducted for the combinations of three fleets and four regions.	vessel effects.	2005 (see Bigelow and Hoyle, 2008).
To study the gillnet catch efficiency of sardine fisheries and develop three predictive models using linear and nonlinear techniques.	Sri Lanka	1999 – 2002	Data was obtained from the commercial fishery data collection programme by the National Aquatic Resources Research and development Ageny (NARA) during May 1999 – August 2002.	Analysed using 3 statistical techniques multiple linear regression (MLR), GAM and RTM using S-PLUS 2000 software.	Nonlinear techniques fit fisheries better and may lead to improved result than linear as many relationships found in fisheries are nonlinear.	Generalized additive models (GAMs) and regression tree models (RTMs) are relatively two new methodologies developed in the field of statistics in the recent past now are applied in quantitative fisheries modelling (see Haputhantri et al., 2009).
To produce standardized CPUE index per	South Pacific	1960 – 2011	The data on effort and catches in number of fish by species were	Analysis was done using clustering techniques and GLM.	The standardization methodology used in this paper	Research in standardizing CPUE indices was

region for vessels targeting South Pacific albacore on a specific trip.			compiled from individual vessels submitting logsheet data of longline activity in the South Pacific.		constructed one CPUE index per region. Vessel effects were considered claiming they better capture fishing performance between vessels rather than subsetting by flag (fleet).	developed in 2008 by Bigelow and Hoyle for distant-water fleets targeting south Pacific albacore by analysing operational level data of vessels landing at Pago Pago, American Samoa and Levuka, Fiji Canneries (see Bigelow and Hoyle, 2012).
To provide updated standardized CPUE indices for the 2013 stock assessment of swordfish in the WCPO based upon aggregated distant water fishing nation longline data.	South West Pacific	1952- 2012	Catch and Effort data from Japanese and Chinese Taipei longline fleets aggregated by year, month and spatial cell were used.	Analysed for the swordfish was done using the GLM.	The study updated the CPUE standardization analysis and recommended using the Japanese index for the Western Central region and the Taiwanese index for the Eastern Central region as indices of	The stock assessment of swordfish (Xiphiasgladius) was done in 2008 and this paper updated the CPUE standardization that was presented by Langley in 2006 (see Hoyle et al., 2013).

					abundance	
To analyse the indices for the 2009 stock assessment including modifications made to the GLM approach.	WCPO	1952 to 2007	Data on bigeye and yellowfin tuna was available from the Japanese distant water longline fleet	Analysis was done using GLM with changes done from the models used in previous years with regards to hooks and target species indicator.	There has been a higher rate of effort creep for yellowfin tuna than for bigeye in region 3, which was the only region with comprehensive datasets.	The authors have referred to the work of Langley, 2003;Bigelow et al., 2004;Langley et al., 2005); that the indices for previous assessments have been prepared using generalized linear modelling and habitat- based standardization of data from the Japanese longline fleet (see Hoyle, 2009).
To evaluate factors of vertical distribution of hooks and the vertical distribution of species catchability based on depth and	Central North Pacific Ocean	1952 – 2007	Single time-depth recorders (TDRs) were used by Fishery observers of the National Marine Fisheries Service (NMFS) to obtain actual longline fishing depths.	GLM was developed to explain the vertical distribution in catchability by depth and habitat.	Bigeye tuna capture occurred near high temperature gradients and 50 to 100 metres above the maximum micronekton biomass and catchability for	The authors have referred to the work of Maunder and Punt, 2004 stating that standardizing catch rates for factors other than abundance, such

habitat (Oceanography) in CPUE standardization.					bigeye in the Hawaii- based factory is the highest at the bottom of the thermocline.	as historical changes in catchability is one of the most commonly applied techniques in fisheries science (see Bigelow, 2006).
To identify factors driving the size trends of albacore.	South Pacific	1962-2007	Length data on aggregate was collected from distant water fishing fleets and logsheet level data was collected from distant water longline vessels offloading catches in Pago Pago and Levuka Canneries with the analysis done using upper/ lower percentile trends and characterization of fleet movement over time.	GLM and GAM	The factors (length frequency, data collection sampling bias, selectivity changes and fishing impacts temporal and spatial trends) that were investigated were found to be affecting the size of albacore tuna and there was a suggestion for areas to be investigated in the future.	GLM based analysis of size trends were done by Langley and Hoyle, in 2008 which had similar trends in mean length of albacore (see Bromhead et al., 2009).
To develop standardized CPUE series for	Region 6 of Western and Central	1964- 2010	Logbook data of Taiwanese distant- water longline were	Delta log- normal model with Regression Tree	The final lognormal index in Region 6 was recommended	In 2011, Chang has demonstrated that delta-

yellowfin and bigeye tuna in the WCPO using a delta- lognormal approach.	Pacific		obtained from the Overseas Fisheries Development Council from 1964 to 2010.	analysis was used for CPUE standardisations.	to be used for stock assessment of yellowfin tuna and hpb< 13.5 or latitude south of 25°S was preferred for stock assessment of bigeye tuna and regression tree analysis on albacore catch rate showed that latitude was the significant classifying factor.	lognormal model was more suitable for CPUE standardisation on Taiwanese longline data with multi-target species (see Chang et al., 2011).
To describe a developing pelagic longline fishery targeting albacore species in the Egyptian waters of the Mediterranean Sea. The species selectivity, CPUE, length frequency and	Mediterranea n Sea off Egypt	June – July 2010	Pelagic longline fishery with albacore was the main fishery. Data was collected from 18 fishing trips surveying operations in the albacore fishing season.	The equation describing the length-weight relationship was transformed into linear form and the regression analysis (the ordinary least squares method) was done.	The study claimed that Egyptian fishermen use highly selective pelagic longline to catch albacore and the length data ranged from 54 to 106cm with CPUE ranging from 7-22 fish/ hooks.	The authors investigated the developing pelagic longline fishery targeting albacore species in the Egyptian waters. (see Gabr and El- Haweet, 2012).

length- weight relationship of albacore were investigated.						
The specific research problems were to analyse the fishery effort distribution, the spatial and seasonal variability and catch-at-size distribution of albacore and standardizing CPUE index.	Reunion Island	1992 - 2013	Pelagic longline fishery for albacore. The datasets used were divided into four different periods from 1992 to 2013 and captured including voluntary logbook program, logbook mandatory program and traditional logbooks and by the use of electronic logbooks and Vessel Monitoring System (VMS).	CPUE was standardized using GLMs. Four different modelling approaches were used, including Tweedie, lognormal, Negative Binomial and Delta method.	The standardized CPUE was presented from 1994 to 2013 with the exception from 2001 to 2004 for future stock assessments	The use of explanatory variable; regions in CPUE standardization have been used previously by Kolody et al. in 2010. There has been reference made to the work of other authors who have worked with the different types of models and distribution in dealing with zero catches (see Coelho et al., 2014).
To standardize CPUE series of yellowfin tuna caught by	Around equatorial and South-Western	1986 - 2007	Logsheet data used was for longline sets by the Brazilian tuna longline fleets including	GLMs were used to standardize CPUE as well as error distributions of	The authors claimed that the Poisson and Tweedie	The authors referred to the work of Maunder and Punt, 2004

Brazilian longliners	Atlantic Ocean from 0° to 60° W longitude and from 7° N to 50° S of latitude		national and chartered vessels.	Tweedie and Poisson were used and comparison was made between the two to determine the best model for CPUE standardization.	distributions gave very close results which were satisfactory to standardize the CPUE. The Tweedie model was slightly better than the Poisson distribution. The fluctuations in CPUE were likely related to changes in fishing strategy and change of targeting strategy in North East Brazil than in the South – South East.	that a better way to standardize the CPUE would be to eliminate the factors not related to stock abundance (see Travassos et al., 2009).
To examine ways to minimize bycatch while maximizing the catch of target species at a specific effort level and minimizing	Gulf of Mexico	1992 - 2005	The study used catch and effort data from the commercial longline vessels from 1992 to 2005 and the other data were collected by the laboratory during six scientific longline cruises from 1998 to	GAMs were preferred over GLMs. According to the authors GAMs provided improved prediction skills but the model results were difficult to interpret and analyse	The VMS could be used to redirect pelagic longline vessels from areas of high risk areas to lower risk of Bluefin bycatch and higher probability of yellowfin catch.	The previous work focused on a variety of aspects including management and mitigation of the bycatch of non-target species, reducing bycatch

mitigation costs.			2002. Target species were yellowfin and Bluefin tuna and the fishery involved pelagic longline.	for comparative purposes. To meet the aim of the study GLMs were used which provided good fits to the data.	The authors claimed the results of their study can be incorporated into fisheries management plans and the bycatch of Bluefin could be reduced and improve the CPUE of yellowfin tuna.	from pelagic longlines targeting tuna and swordfish, gear changes and spatio temporal management of fishing effort to reduce bycatch of non-target species, eddy kinetic energy, cyclonic eddies, positive vorticity and association with cooler sea surface temperatures (see Teo and Block, 2010).
To interpret CPUE data to assess the status of individual stocks and communities.	Tropical Pacific Ocean	1952 - 2002	Different fisheries and species were used to demonstrate the efficiency of different methods.	The authors discussed the problem associated with raw Catch Per Unit Effort (CPUE) data, management of single stocks, three methods that can overcome problem with CPUE	The authors suggested that analysis solely based on CPUE data can be misinterpreted. CPUE based analyses play an important role in	According to Maunder and Punt, 2004, CPUE standardisation is one of the most commonly applied analyses in fishery to remove factors associated with

				that include integrated stock assessment models, management strategy evaluation and adaptive management.	cases where full stock assessments are not possible, for example, when catch time-series are not available.	abundance so that CPUE is proportional to abundance. According to Maunder and Starr, 2003 , integrated stock assessment is considered to be one of the main approaches used in modern fisheries stock assessment (see Maunder et al., 2006).
To categorize longline sets made by the Sao Paulo longline fleet using cluster analysis as a way to characterize the target species to generate standardized CPUE series	South eastern coast of Brazil (Southern Atlantic Ocean)	1998 - 2006	15 categories of species were used. The logbook data was used from the longliners in Sao Paulo.	Cluster analysis and CPUE standardization	In relation to the longline fleets in Sao Paulo, blue shark's importance in this fishery has been growing over time especially in the offshore areas and in the first and fourth quarters of the year as a result	According to HE et al., 1997; one of the most important factors affecting CPUE is the targeting strategy (see Mourato et al.,2011).

which is the essential information for most stock assessments.					of changes in the spatio temporal distribution of fishing effort and variation of the proportion of each cluster by sub area.	
To standardize CPUE of albacore tuna caught using GLM to assess abundance index for the Korean tuna longline fishery in the Indian Ocean	Indian Ocean where the study area was divided into 2 parts.	1977 - 2013	Operational data of Korean tuna longline fishery was used and the main species was albacore tuna.	GLM was used to standardize albacore tuna CPUE for the whole area and core area.	The authors compared standardized CPUEs among the whole and core area model where all the CPUEs showed an increase in 2007 and different trends thereafter among each model.	The authors have also referred to the work of Lee et al, 2014 in dividing the number of hooks between floats (hbf) in 3 classes (see Lee et al., 2014).
To standardize the Indian albacore abundance indices based on Taiwanese data from 1980 to 2011 using GLMs.	Indian Ocean	1980 – 2011	Logbook data of Taiwanese longline vessels were provided by Overseas Fisheries Development Council (OFDC) of Taiwan and the main species was Indian albacore.	Cluster analysis was carried out to define the subareas. GLM models for yearly and quarterly standardization with normal error structure were formulated.	The authors claimed that the use of ANOVA tables pointed that in both yearly and quarterly series model, factor sub area played an important role in explanation of its orthogonal variation	The authors have referred to the work of Robson, 1966; Gavaris, 1980; Kimura, 1981 in using GLM with normal error structure to standardise yearly and quarterly

					to the total.	CPUE series for Indian albacore (see Lee et al., 2012).
To obtain an updated standardized CPUE series for yellowfin tuna caught by Taiwanese longline fishery in the Indian Ocean.	Indian Ocean area was aggregated by whole and tropical Indian Ocean. the tropical region was divided into 2 areas.	1980 - 2012	The daily catch and effort data was used from the logbooks of Taiwanese longline fishery which were provided by OFDC. Data was also used on the number of hooks between floats (nhbf) which were available from 1995 and the main species was yellowfin tuna.	GLMs were developed to standardize the CPUE.	The authors claimed that the standardized CPUE series showed similar trend with the nominal CPUE except before 1986.	The authors have referred to the work of Yeh et al.2012 where the analytical methods which were used in previous studies were applied to the available data in 2012 (see Yeh and Chang, 2013).
To standardize CPUE of White and Blue Marlins caught by the commercial longliners operating from Brazil using the GLM.	Brazil The total fishing ground was divided into two areas; to the north and south of 15° S based on the spatial	1980 - 2004	There were 6 species used in the cluster analysis in the fishery Brazilian longline fleets were used. The depth of the fishing ground was obtained from the National Geophysical Data Center. Data on Sea Surface Temperature (SST)	Target species were identified using cluster analysis. The relative abundance indices for both species were estimated using a delta-lognormal GLM and other statistical techniques (Binomial Error	Environment factors such as SST, depth of fishing ground and moon phase showed low influence on both blue and white marlin CPUE. According to the authors, the nominal and	The authors have referred to the work of several other authors where GLM has been referred as the most common statistical tools used for CPUE standardization and less frequent

	distribution of CPUE.		were obtained from the Physical Oceanography Distributed Active Archive Center of Jet Propulsion Laboratory/ NASA, Geophysical Fluid Dynamics Lab/ Ocean data IRI/ ARCS/ Ocean assimilation.	Distribution and Gaussian Error Distribution).	standardized yearly CPUE did not differ much during the whole period. The yearly trend of blue marlin was more stable when compared to white marlin which had strong fluctuations.	tool has been referred to GAMs (see Hazin et al., 2007).
To update the standardized catch rate of albacore tuna captured by the Uruguayan tuna longline fleet upto 2012.	South Western Atlantic Ocean	1981 – 2013	Data was analysed from the logbooks of the Uruguayan tuna longline fleet.	The standardized CPUE of albacore was calculated using generalized linear mixed models (GLMm).	The authors claimed that standardized CPUE series showed a slightly decrease in relative abundance from 1983 to 2005 but remained constant till the end of the time series.	Various authors in their studies discussed Uruguayan tuna longline fleets targeting strategy over a period of time (see Pons and Domingo, 2014).
To improve the standardization of Japanese longline CPUE with recommendations to provide standardized CPUEs for the	Indian Ocean and Tropical Areas	1946-2000	Data of Japanese longliners were analysed for bigeye tuna. The SST data was taken from the Subarctic Gyre Experiment compiled by the Climate and	There were 12 GLMs developed and run to standardize CPUE.	The results of the study suggested that the catch at age information did not reflect enough the change in abundance of each age, may be	The authors have referred to the work of Okamoto et al. (2001) that the use of environmental factors in the GLM model

stock assessment methods in the Indian Ocean by applying environment factors			Marine Department of Japan Meteorological Agency. The data on monthly Southern Oscillation Index (SOI) was downloaded from the National Oceanic and atmospheric Administration website.		because of shortage of time and space coverage of size data. A high peak in the 1977 and 1978 for the age 4-5 group was seen for both areas.	might be tested to improve the standardization (see Okamoto et al., 2001).
To standardize distant water tuna longline hooking rate for Yellow fin tuna(Thunnus albacares) from Fishery Survey of India (FSI) fleet	Indian Ocean	1981 - 2012	Main species was yellowfin tuna for the distant water fleets from the Fishery Survey of India (FSI).	GLM was used to standardize the nominal yellow fin tuna catch rate of longline fishery operating in Indian Ocean and to evaluate the impacts of temporal, spatial and fisheries operational variables on the catch rate.	The authors claimed that the frequency distribution of standardized residuals for all variables combined effects was approximately normally distributed. All the variables except catch rates of sailfish, skipjack and marlin were significant in the GLM analysis which accounted for 26% of the	According to Hilborn and Walters, 2001; Nishida and Chen, 2004 the data of CPUE, is often used as a relative index of fisheries stock abundance (see Gulati and Premchand, 2015).

					variance in nominal CPUE. The main factor that affected CPUE variability was year followed by latitudes and the hooking rate decreased to a very low in recent years.	

Appendix B: Data Description

No.	Variable	Description
1	set_id	This links to longline logsheet set record. The number of sets within a trip has unique ids. The data type is integer.
2	alb_n	This refers to the number of albacore tuna caught. This was derived from the sp_code in the file I_setcatch.
3	avg_alb_len	This refers to the average length in centimeters of albacore tuna caught in a set. This was derived from the variable length from the file I_setcatch.
4	avg_alb_wt_est	This refers to the average weight estimate of albacore tuna caught in a set. This is an estimate of weight if no measuring devices have been used. The estimation can be based on length-weight relationship or other methods. This was derived from the variable wt_est from the file I_setcatch.
5	bet_n	This refers to the number of bigeye tuna caught. This was derived from the sp_code in the file I_setcatch.
6	avg_bet_len	This refers to the average length in centimeters of bigeye tuna caught in a set. This was derived from the variable len from the file I_setcatch.
7	avg_bet_wt_est	This refers to the average weight estimate of bigeye tuna caught in a set. This is an estimate of weight if no measuring devices have been used. The estimation can be based on length-weight relationship or other methods. This was derived from the variable wt_est from the file I_setcatch.

8	yft_n	This refers to the number of yellowfin tuna caught. This was derived from the sp_code in the file I_setcatch.
9	avg_yft_len	This refers to the average length in centimeters of yellowfin tuna caught in a set. This was derived from the variable len from the file I_setcatch.
10	avg_yft_wt_est	This refers to the average weight estimate of yellowfin tuna caught in a set. This is an estimate of weight if no measuring devices have been used. The estimation can be based on length-weight relationship or other methods. This was derived from the variable wt_est from the file I_setcatch.
11	tuna_n	This refers to the sum of albacore, bigeye and yellowfin tuna caught.
12	avg_tuna_len	This refers to the average length in centimeters of albacore, bigeye and yellowfin tuna caught in a set.
13	avg_tuna_wt_est	This refers to the average weight estimate of albacore, bigeye and yellowfin tuna caught in a set.
14	other_catch	This refers to other species being caught apart from albacore, bigeye and yellowfin tuna.
15	avg_other_len	This refers to the average length in centimeters of other species excluding tuna.
16	avg_other_wt_est	This refers to the average weight estimate of other species excluding tuna.
17	total_catch	This refers to the sum of tuna and other species caught.
18	avg_tot_len	This refers to the average length in centimeters of tuna and other species caught.

19	avg_tot_wt_est	This refers to the average weight estimate of tuna and other species caught.
20	hk_bt_flt	This refers to the number of hooks between floats.
21	hook_set	This refers to the total number of hooks in a set.
22	target_tun_yn	This gives an indication on which tuna is targeted.
23	catch_quarter	This refers to the quarter in which catch occurred. This was derived from the variable catch_dtime which shows the date and time of catch.
24	catch_month	This refers to the month in which catch occurred. This was derived from the variable catch_dtime which shows the date and time of catch.
25	ret_year	This refers to the year in which catch occurred. This was derived from the variable catch_dtime which shows the date and time of catch.
26	year_qtr	This is the combination of the catch year and quarter effects.
27	Lat	This refers to latitude of the catch in decimal format.
28	long	This refers to longitude of the catch in decimal format.
29	lat1	This refers to the latitude aggregated at 1 degree resolution.
30	long1	This refers to the longitude aggregated at 1 degree resolution.

31	vessel_id	This refers to the unique identifier that is assigned to each vessel taking part in the longline fishing activity. A vessel can make more than 1 fishing trip in the time series.
32	set_number	This value was used to calculate the number of sets in a trip that was used to calculate the weight.
33	weight	Each set is weighted by $\frac{1}{\sqrt{number\ of\ sets\ by\ trip}}$ as the individual sets within a trip are often highly correlated.
34	Nominal.CPUE	This is calculated by albacore catch divided by hook set. Nominal CPUE does not take into account other factors such as targeting change, vessels and so on that affect catch.
35	obstrip_id	This is the unique id that is assigned to the fishing trips. There are more than 1set within a trip. This is an integer.

Appendix C: Summary of the Selected Model

Factor	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	-5.535476	1.11195	-4.978	0.000000684***
as.factor(year_qtr)2003.0.25	0.4247435	0.539288	0.788	0.431002
as.factor(year_qtr)2003.0.5	0.2391326	0.655534	0.365	0.715297
as.factor(year_qtr)2004	-0.2602761	0.646649	-0.402	0.687349
as.factor(year_qtr)2004.0.25	-0.024975	0.698226	-0.036	0.971469
as.factor(year_qtr)2004.0.5	0.931829	0.694666	1.341	0.179905
as.factor(year_qtr)2004.0.75	1.5989122	0.348529	4.588	0.0000047***
as.factor(year_qtr)2005	0.2592103	0.630855	0.411	0.681189
as.factor(year_qtr)2005.0.25	0.9248177	0.638224	1.449	0.147444
as.factor(year_qtr)2005.0.5	0.558703	0.627588	0.89	0.373419
as.factor(year_qtr)2005.0.75	0.5770376	0.643653	0.897	0.370066
as.factor(year_qtr)2006	-0.3487869	0.635227	-0.549	0.583001
as.factor(year_qtr)2006.0.25	0.9511335	0.636663	1.494	0.135313
as.factor(year_qtr)2006.0.5	0.4082421	0.640184	0.638	0.523728
as.factor(year_qtr)2006.0.75	1.4056171	0.646115	2.175	0.029683*
as.factor(year_qtr)2007	0.693513	0.636844	1.089	0.276261
as.factor(year_qtr)2007.0.25	0.5483448	0.6498	0.844	0.398821
as.factor(year_qtr)2007.0.5	0.5138433	0.650224	0.79	0.42945
as.factor(year_qtr)2007.0.75	0.8318845	0.653522	1.273	0.203158
as.factor(year_qtr)2008	-0.4311648	0.644151	-0.669	0.503329
as.factor(year_qtr)2008.0.25	0.5422614	0.641503	0.845	0.398022
as.factor(year_qtr)2008.0.5	0.425304	0.644534	0.66	0.5094
as.factor(year_qtr)2008.0.75	0.7250431	0.638116	1.136	0.255967
as.factor(year_qtr)2009	0.3505439	0.642094	0.546	0.585155
as.factor(year_qtr)2009.0.25	0.3539106	0.637087	0.556	0.578591
as.factor(year_qtr)2009.0.5	0.3700701	0.647251	0.572	0.567536
as.factor(year_qtr)2009.0.75	0.8105393	0.645947	1.255	0.209661
as.factor(year_qtr)2010	0.1010095	0.633466	0.159	0.873323
as.factor(year_qtr)2010.0.25	0.3455662	0.642693	0.538	0.59084
as.factor(year_qtr)2010.0.5	0.3056979	0.646378	0.473	0.636296
as.factor(year_qtr)2011	-0.6873353	0.658617	-1.044	0.296765
as.factor(year_qtr)2011.0.25	0.6987165	0.648801	1.077	0.281608
as.factor(year_qtr)2011.0.75	-0.0300399	0.665487	-0.045	0.963999
as.factor(year_qtr)2012	-0.4756323	0.641956	-0.741	0.458814
as.factor(year_qtr)2012.0.25	0.3011613	0.656872	0.458	0.646647
as.factor(year_qtr)2012.0.5	-0.4037717	0.665615	-0.607	0.544159
as.factor(year_qtr)2012.0.75	0.1819987	0.666419	0.273	0.784798
as.factor(year_qtr)2013	-0.7063085	0.631006	-1.119	0.263099
as.factor(year_qtr)2013.0.25	0.5103671	0.637426	0.801	0.423396
as.factor(year_qtr)2013.0.5	-0.0287952	0.640805	-0.045	0.964162

as.factor(year_qtr)2013.0.75	0.0817642	0.642893	0.127	0.898806
as.factor(year_qtr)2014	-0.2444402	0.629829	-0.388	0.697969
as.factor(year_qtr)2014.0.25	0.5989913	0.642254	0.933	0.351093
as.factor(year_qtr)2014.0.5	-0.0571751	0.641418	-0.089	0.928979
as.factor(year_qtr)2014.0.75	0.4536021	0.638565	0.71	0.477553
as.factor(catch_month)2	0.2800108	0.100518	2.786	0.00538**
as.factor(catch_month)3	0.1959483	0.121302	1.615	0.106351
as.factor(catch_month)4	-0.2089956	0.07326	-2.853	0.004368**
as.factor(catch_month)5	-0.1550452	0.068545	-2.262	0.023783*
as.factor(catch_month)6	NA	NA	NA	NA
as.factor(catch_month)7	0.5274396	0.069214	7.62	0.0000000000000351***
as.factor(catch_month)8	0.3359491	0.067241	4.996	0.000000623***
as.factor(catch_month)9	NA	NA	NA	NA
as.factor(catch_month)10	-0.3732496	0.083356	-4.478	0.00000787***
as.factor(catch_month)11	-0.0091162	0.081333	-0.112	0.910765
as.factor(catch_month)12	NA	NA	NA	NA
as.factor(lat1)-24	0.7328576	0.999942	0.733	0.463685
as.factor(lat1)-23	0.4934514	1.116177	0.442	0.65846
as.factor(lat1)-22	-0.0443762	0.941789	-0.047	0.962422
as.factor(lat1)-21	0.2508673	0.92631	0.271	0.786548
as.factor(lat1)-20	0.2441432	0.920561	0.265	0.790868
as.factor(lat1)-19	-0.1237869	0.924739	-0.134	0.893522
as.factor(lat1)-18	-0.0505863	0.930641	-0.054	0.956655
as.factor(lat1)-17	0.0133424	0.976547	0.014	0.9891
as.factor(lat1)-16	0.2292756	0.912945	0.251	0.801727
as.factor(lat1)-15	0.0830575	0.89819	0.092	0.92633
as.factor(lat1)-14	0.3649382	0.881566	0.414	0.678933
as.factor(lat1)-13	-0.0259625	0.887074	-0.029	0.976653
as.factor(lat1)-12	-0.3137741	0.519566	-0.604	0.545952
as.factor(lat1)-11	0.188662	0.416473	0.453	0.650587
as.factor(lat1)-10	-0.4313095	0.62899	-0.686	0.492952
as.factor(long1)-179	-0.5056743	0.581969	-0.869	0.38498
as.factor(long1)-178	-0.6034395	0.672653	-0.897	0.369746
as.factor(long1)-177	1.0052343	0.438122	2.294	0.021846*
as.factor(long1)173	0.9332168	1.064293	0.877	0.380653
as.factor(long1)174	1.073122	0.955707	1.123	0.261601
as.factor(long1)175	0.5464535	0.956301	0.571	0.567761
as.factor(long1)176	0.3446165	0.825931	0.417	0.676532
as.factor(long1)177	0.9152896	0.784368	1.167	0.243352
as.factor(long1)178	0.5195686	0.275421	1.886	0.059345.
as.factor(long1)179	0.2578077	0.158698	1.625	0.104386
as.factor(long1)180	0.627564	0.654067	0.959	0.337406
as.factor(vessel_id)20154	0.1001611	0.136828	0.732	0.46422
as.factor(vessel_id)21863	0.2019723	0.203182	0.994	0.320291
as.factor(vessel_id)21882	-0.3214712	0.303067	-1.061	0.288913

as.factor(vessel_id)21884	-0.6600848	0.267428	-2.468	0.01364*
as.factor(vessel_id)21888	0.1268805	0.223885	0.567	0.570952
as.factor(vessel_id)21889	0.3009972	0.125075	2.407	0.016174*
as.factor(vessel_id)22479	-0.1600709	0.170914	-0.937	0.349072
as.factor(vessel_id)22485	-0.0542087	0.176767	-0.307	0.759121
as.factor(vessel_id)22488	-0.0989431	0.186021	-0.532	0.594846
as.factor(vessel_id)22492	0.4777224	0.118553	4.03	0.0000575***
as.factor(vessel_id)22510	0.2038916	0.114661	1.778	0.075486.
as.factor(vessel_id)22511	-0.4269336	0.352101	-1.213	0.225419
as.factor(vessel_id)22512	0.1716103	0.293426	0.585	0.558699
as.factor(vessel_id)22515	-0.3668281	0.211837	-1.732	0.083453.
as.factor(vessel_id)22517	-0.14287	0.135748	-1.052	0.292684
as.factor(vessel_id)22520	0.2044081	0.232739	0.878	0.379876
as.factor(vessel_id)22522	-0.1341281	0.14474	-0.927	0.354177
as.factor(vessel_id)22538	-0.4521896	0.184216	-2.455	0.014166*
as.factor(vessel_id)22540	-0.855537	0.224822	-3.805	0.000145***
as.factor(vessel_id)22554	0.6939814	0.130539	5.316	0.000000115***
as.factor(vessel_id)22767	0.2333712	0.192559	1.212	0.225643
as.factor(vessel_id)22768	0.5017711	0.279117	1.798	0.072338.
as.factor(vessel_id)23295	-0.0839899	0.253669	-0.331	0.740595
as.factor(vessel_id)23299	-0.6108668	0.248595	-2.457	0.014064*
as.factor(vessel_id)23303	0.4136751	0.407716	1.015	0.310382
as.factor(vessel_id)23305	0.6735894	0.151908	4.434	0.00000962***
as.factor(vessel_id)23306	0.3232788	0.174985	1.847	0.064792.
as.factor(vessel_id)23310	0.1530514	0.210949	0.726	0.468186
as.factor(vessel_id)23318	-0.5190489	0.594626	-0.873	0.382798
as.factor(vessel_id)23499	0.8257654	0.242038	3.412	0.000655***
as.factor(vessel_id)23502	0.3222912	0.216046	1.492	0.135879
as.factor(vessel_id)23503	0.6111829	0.204779	2.985	0.002866**
as.factor(vessel_id)23554	0.7191773	0.206132	3.489	0.000493***
as.factor(vessel_id)23556	-0.243219	0.237112	-1.026	0.305101
as.factor(vessel_id)23709	-0.5876636	0.112896	-5.205	0.000000209***
as.factor(vessel_id)24137	-0.2142479	0.218763	-0.979	0.327492
as.factor(vessel_id)24774	0.1139237	0.221834	0.514	0.607607
as.factor(vessel_id)26083	0.1183738	0.097027	1.22	0.222571
as.factor(vessel_id)26401	0.3807637	0.117357	3.244	0.001191**
as.factor(vessel_id)26732	0.0152422	0.136647	0.112	0.911194
as.factor(vessel_id)26756	0.0723693	0.109652	0.66	0.509319
as.factor(vessel_id)28128	-0.293297	0.260722	-1.125	0.260717
as.factor(vessel_id)28131	0.5738543	0.18592	3.087	0.002046**
as.factor(vessel_id)28134	0.119489	0.138014	0.866	0.386694
as.factor(vessel_id)28320	-0.0746304	0.159484	-0.468	0.63986
as.factor(vessel_id)28436	0.2624565	0.198326	1.323	0.185832
as.factor(vessel_id)28437	-0.1883536	0.154239	-1.221	0.222129
as.factor(vessel_id)28987	0.0691679	0.166137	0.416	0.677203

as.factor(vessel_id)29000	0.1738981	0.226364	0.768	0.442425
as.factor(vessel_id)51066	0.136216	0.089113	1.529	0.12649
as.factor(vessel_id)51117	0.3963172	0.122758	3.228	0.00126**
as.factor(vessel_id)57031	0.0075604	0.244351	0.031	0.975319
as.factor(vessel_id)57188	-0.1019588	0.266251	-0.383	0.701793
as.factor(vessel_id)57300	0.0758579	0.124391	0.61	0.542023
as.factor(vessel_id)59120	0.589248	0.20042	2.94	0.00331**
as.factor(vessel_id)59121	-0.0029964	0.158129	-0.019	0.984883
as.factor(vessel_id)59832	0.161168	0.114064	1.413	0.157787
as.factor(vessel_id)90823	-0.1739785	0.201079	-0.865	0.386996
as.factor(vessel_id)90825	-0.2518141	0.123658	-2.036	0.041813*
as.factor(vessel_id)91795	0.3140882	0.103537	3.034	0.00244**
as.factor(vessel_id)92208	-0.1201154	0.11727	-1.024	0.305805
as.factor(vessel_id)92287	-0.2329618	0.108892	-2.139	0.032497*
as.factor(vessel_id)92308	-0.5263182	0.135296	-3.89	0.000103***
as.factor(vessel_id)92499	-0.1375905	0.119297	-1.153	0.248874
as.factor(vessel_id)92512	-0.033355	0.09275	-0.36	0.71916
as.factor(vessel_id)92703	0.0120751	0.187727	0.064	0.948718
as.factor(vessel_id)92704	-0.2192482	0.14299	-1.533	0.125319
as.factor(vessel_id)99097	-0.5610155	0.242367	-2.315	0.020704*
as.factor(lat1)-24:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-23:as.factor(long1)-179	-0.3374332	0.946843	-0.356	0.721587
as.factor(lat1)-22:as.factor(long1)-179	-0.4355685	0.664697	-0.655	0.51234
as.factor(lat1)-21:as.factor(long1)-179	-1.9578527	0.818134	-2.393	0.016778*
as.factor(lat1)-20:as.factor(long1)-179	0.4317265	0.625378	0.69	0.490039
as.factor(lat1)-19:as.factor(long1)-179	-0.3079929	0.653188	-0.472	0.637307
as.factor(lat1)-18:as.factor(long1)-179	0.4554057	0.640027	0.712	0.476812
as.factor(lat1)-17:as.factor(long1)-179	0.2458226	0.69372	0.354	0.723102
as.factor(lat1)-16:as.factor(long1)-179	0.4744881	0.597424	0.794	0.427138
as.factor(lat1)-15:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-14:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-13:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-12:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-11:as.factor(long1)-179	NA	NA	NA	NA
as.factor(lat1)-10:as.factor(long1)-179	NA	NA	NA	NA

as.factor(lat1)-24:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-23:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-22:as.factor(long1)-178	0.6609105	0.841076	0.786	0.432061
as.factor(lat1)-21:as.factor(long1)-178	0.602106	0.719594	0.837	0.402821
as.factor(lat1)-20:as.factor(long1)-178	0.4343625	0.700666	0.62	0.535359
as.factor(lat1)-19:as.factor(long1)-178	0.7393612	0.706771	1.046	0.295606
as.factor(lat1)-18:as.factor(long1)-178	0.591666	0.715107	0.827	0.408097
as.factor(lat1)-17:as.factor(long1)-178	0.6269339	0.770739	0.813	0.416052
as.factor(lat1)-16:as.factor(long1)-178	0.8279176	0.717932	1.153	0.248935
as.factor(lat1)-15:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-14:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-13:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-12:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-11:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-10:as.factor(long1)-178	NA	NA	NA	NA
as.factor(lat1)-24:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-23:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-22:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-21:as.factor(long1)-177	-1.2915063	0.83563	-1.546	0.122335
as.factor(lat1)-20:as.factor(long1)-177	-0.9586329	0.585517	-1.637	0.101701
as.factor(lat1)-19:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-18:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-17:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-16:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-15:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-14:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-13:as.factor(long1)-177	NA	NA	NA	NA

as.factor(lat1)-12:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-11:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-10:as.factor(long1)-177	NA	NA	NA	NA
as.factor(lat1)-24:as.factor(long1)173	NA	NA	NA	NA
as.factor(lat1)-23:as.factor(long1)173	NA	NA	NA	NA
as.factor(lat1)-22:as.factor(long1)173	0.2228648	1.367115	0.163	0.870517
as.factor(lat1)-21:as.factor(long1)173	-0.6861996	1.100453	-0.624	0.532971
as.factor(lat1)-20:as.factor(long1)173	-0.6281632	1.095548	-0.573	0.566438

Appendix D: R Codes

```
#Reading data from the file df2
df2=read.csv("C:/Users/Sandeep  Singh/Desktop/df2.csv",header=T)
names(df2)

set_total=tapply(df2$set_number,df2$ret_year,sum)
set_total

test4=cbind(tapply(df2$bet_n,df2$obstrip_id,sum),tapply(df2$yft_n,df2$obstrip_id,
sum),tapply(df2$alb_n,df2$obstrip_id,sum),
        tapply(df2$tuna_n,df2$obstrip_id,sum))

# To avoid "atomic vector error" when calculating the species composition:
test4=as.data.frame(test4)

# Adding the column names:
colnames(test4)<-c("bet_n","yft_n","alb_n","tuna_n")
head(test4)
test4

#Calculating the species composition
test4$bet_perc<-test4$bet_n/test4$tuna_n
test4$yft_perc<-test4$yft_n/test4$tuna_n
test4$alb_perc<-test4$alb_n/test4$tuna_n

head(test4)

# building an equivalent trip_id with the same length as size composition:
obstrip_id<-tapply(df2$obstrip_id,df2$obstrip_id,mean)

# To add the trip_id information to the data set test4:
test4<-cbind(test4,obstrip_id)
head(test4)
```

```
names(test4)

#To add ret_year information,column 8 in test4 and column 35 in df2 is the obstrip_id and column 25 is the ret_year in df2
test4$ret_year<-df2[match(test4[,8],df2[,34]),24]
test4_backup<-test4

# Displays the first 6 rows of the file test 4
head(test4)

# Displays the first 6 rows of the file test 4_backup
head(test4_backup)

# Erasing extra info not needed in cluster calculations
test4$alb_n<-NULL
test4$bet_n<-NULL
test4$yft_n<-NULL
test4$tuna_n<-NULL

# Placing the tuna_n in the last column
test4$tuna_n<-test4_backup$tuna_n
head(test4)

# checking
test4[1,]
summary(test4)


# First step for the cluster analysis:
install.packages("amap")
library(amap)
library(cluster)

# Puting the values of test4 into df_clust
```

```
df_clust<-test4
names(df_clust)

# Displays the first 6 rows of the file df_clust
head(df_clust)

# Include data that are different from zero
newjunk3<-df_clust[df_clust$tuna_n!=0,]
dim(newjunk3)
summary(newjunk3$ret_year)
summary(newjunk3)
dim(newjunk3$ret_year)

# Here we select for the first period of time from 2003 to 2005 inclusive:
newjunk5prev<-newjunk3[newjunk3$ret_year>=2003 &
newjunk3$ret_year<=2005,]
dim(newjunk5prev)

#This line carries out the cluster analysis using a subroutine from the amap package:
fitprev<-hcluster(newjunk5prev[4:6],method="euclidean")

#This line provides the plot of the clusters so we can appreciate the number of
different
#targeting:
plot(fitprev,labels=FALSE,hang =-1,main="Dendrogram  2003-2005")
groups<-cutree(fitprev,k=3)
rect.hclust(fitprev,k=3,border="blue")
dev.off()

# Here we select for the second period of time from 2006 to 2008 inclusive:
newjunk6prev<-newjunk3[newjunk3$ret_year>=2006 &
newjunk3$ret_year<=2008,]
dim(newjunk6prev)
```

#This line carries out the cluster analysis using a subroutine from the amap package:
fitprev1<-hcluster(newjunk6prev[4:6],method="euclidean")

#This line provides the plot of the clusters so we can appreciate the number of different
#targeting:
plot(fitprev1,labels=FALSE,hang =-1,main="Dendrogram 2006-2008")
groups<-cutree(fitprev,k=3)
rect.hclust(fitprev1,k=3,border="blue")
dev.off()

# Here we select for the third period of time from 2009 to 2011 inclusive:
newjunk16prev<-newjunk3[newjunk3$ret_year>=2009 & newjunk3$ret_year<=2011,]
dim(newjunk16prev)

#This line carries out the cluster analysis using a subroutine from the amap package:
fitprev2<-hcluster(newjunk6prev[4:6],method="euclidean")

#This line provides the plot of the clusters so we can appreciate the number of different
#targeting:
plot(fitprev2,labels=FALSE,hang =-1,main="Dendrogram 2009-2011")
groups<-cutree(fitprev,k=3)
rect.hclust(fitprev2,k=3,border="blue")
dev.off()

# Here we select for the fourth period of time from 2012 to 2014:
newjunk17prev<-newjunk3[newjunk3$ret_year>=2012 & newjunk3$ret_year<=2014,]
dim(newjunk17prev)

#This line carries out the cluster analysis using a subroutine from the amap package:
fitprev3<-hcluster(newjunk6prev[4:6],method="euclidean")

#This line provides the plot of the clusters so we can appreciate the number of different
#targeting:
plot(fitprev3,labels=FALSE,hang =-1,main="Dendrogram 2012-2014")
groups<-cutree(fitprev,k=3)
rect.hclust(fitprev3,k=3,border="blue")
dev.off()

#The second step in building the cluster analysis is
#the use of the clara subroutine for partitioning the data sets into the appropriate number of
#clusters as determined by dendrograms.

#use of clara subroutine in building cluster analysis

#For the period 2002-2004
includes the data necessary to make clara analysis avoiding memory waste
junk8prev<-
newjunk5prev[,c("bet_perc","yft_perc","alb_perc","ret_year","obstrip_id")]
dim(junk8prev)

#uses clara with three groups, result from the dendrograms
claraprev<-clara(junk8prev[1:3],3) # used 3 groups
junkprev<-aggregate(claraprev$data, by=list(claraprev$clustering),FUN=mean)
junkprev

# Creates a column claraprev$clustering in junk2prev file
junk2prev <- table(claraprev$clustering)
junk2prev

#Pasting the cluster information
newjunk5prev <- cbind(junk8prev,claraprev$clustering)
dim(newjunk5prev)

```
names(newjunk5prev)

#Changing the name for clustering
names(newjunk5prev)[6] <- "clustering"
names(newjunk5prev)
dim(newjunk5prev)

#For the period 2006-2008
# includes the data necessary to make clara analysis avoiding memory waste
junk9prev<-
newjunk6prev[,c("bet_perc","yft_perc","alb_perc","ret_year","obstrip_id")]
dim(junk9prev)

#uses clara with three groups, result from the dendrograms
claraprev1<-clara(junk9prev[1:3],3)  # used 3groups
junkprev1<-aggregate(claraprev1$data, by=list(claraprev1$clustering),FUN=mean)
junkprev1
junk3prev <- table(claraprev1$clustering)
junk3prev

#Pasting the cluster information
newjunk6prev <- cbind(junk9prev,claraprev1$clustering)
dim(newjunk6prev)
names(newjunk6prev)

#Changing the name for clustering
names(newjunk6prev)[6] <- "clustering"
names(newjunk6prev)

#For the period 2009-2011
# includes the data necessary to make clara analysis avoiding memory waste
junk22prev<-
newjunk16prev[,c("bet_perc","yft_perc","alb_perc","ret_year","obstrip_id")]
dim(junk22prev)
```

```
#uses clara with three groups, result from the dendrograms
claraprev2<-clara(junk22prev[1:3],3)  # used 3 groups
junkprev2<-aggregate(claraprev2$data, by=list(claraprev2$clustering),FUN=mean)
junkprev2
junk4prev <- table(claraprev2$clustering)
junk4prev


#Pasting the cluster information
newjunk16prev <- cbind(junk22prev,claraprev2$clustering)
dim(newjunk16prev)
names(newjunk16prev)

#Changing the name for clustering
names(newjunk16prev)[6] <- "clustering"
names(newjunk16prev)

#For the period 2012-2014
# includes the data necessary to make clara analysis avoiding memory waste
junk10prev<-
newjunk17prev[,c("bet_perc","yft_perc","alb_perc","ret_year","obstrip_id")]
dim(junk10prev)

#uses clara with three groups, result from the dendrograms
claraprev3<-clara(junk10prev[1:3],3)  # used 3 groups
junkprev11<-aggregate(claraprev3$data, by=list(claraprev3$clustering),FUN=mean)
junkprev11
junk6prev <- table(claraprev3$clustering)
junk6prev

#Pasting the cluster information
newjunk17prev <- cbind(junk10prev,claraprev3$clustering)
dim(newjunk17prev)
```

```
names(newjunk17prev)

#Changing the name for clustering
names(newjunk17prev)[6] <- "clustering"
names(newjunk17prev)

# Exclude those clusters not targeting albacore
# For the period 2003-2005 Cluster 1 did not target albacore
#From literature review if the sum of yellowfin and albacore catch was less than 80%
then the fishery did not target albacore

dim(newjunk5prev)
junk5g<-newjunk5prev[newjunk5prev$clustering =="2",]
dim(junk5g)

dim(newjunk5prev)
junk5h<-newjunk5prev[newjunk5prev$clustering =="3",]
dim(junk5h)

# For the period 2006-2008 Cluster 3 did not target albacore

dim(newjunk6prev)
junk5e<-newjunk6prev[newjunk6prev$clustering =="1",]
dim(junk5e)

dim(newjunk6prev)
junk5f<-newjunk6prev[newjunk6prev$clustering =="2",]
dim(junk5f)

# Exclude those clusters not targeting albacore
# For the period 2009-2011 Cluster 1 did not target albacore

dim(newjunk16prev)
junk5i<-newjunk16prev[newjunk16prev$clustering =="2",]
```

```
dim(junk5i)

dim(newjunk16prev)
junk5j<-newjunk16prev[newjunk16prev$clustering =="3",]
dim(junk5j)

# Exclude those clusters not targeting albacore
# For the period 2012-2014 Cluster 3 did not target albacore

dim(newjunk17prev)
junk5k<-newjunk17prev[newjunk17prev$clustering =="1",]
dim(junk5k)

dim(newjunk17prev)
junk5l<-newjunk17prev[newjunk17prev$clustering =="2",]
dim(junk5l)

# Now paste the junk files that target albacore

junk3<- rbind(junk5e,junk5f,junk5g,junk5h,junk5i,junk5j,junk5k,junk5l)
names(junk3)
dim(junk3)

# For back up a csv file a RData was created:
write.csv(junk3,file="junk3.csv")
save(junk3,file="junk3.RData")

#Next step is incorporating the cluster information to the file df2:
names(df2)
df2$cluster<-junk3[match(df2[,34],junk3[,5]),6]
dim(df2)
summary(df2$cluster)

#Writing and saving the results of cluster analysis as file cpueSP2
```

```
write.csv(df2,file="cpueSP2.csv")
save(df2,file="cpueSP2.RData")

#Reading from the file cpueSP2
cpueSP2=read.csv("C:/Users/Sandeep  Singh/Documents/cpueSP2.csv",header=T)

#Building the input file for the GLM:
#Selecting the data for the GLM
cpueSP2                                    <-cpueSP2[,c("ret_year","catch_quarter",
"alb_n","tuna_n","hook_set","obstrip_id","vessel_id","lat","long","weight","lat1",
"long1", "cluster", "catch_month", "year_qtr")]
dim(cpueSP2)
summary(cpueSP2$cluster)
save(cpueSP2, file="cpueSP2.RData")

#Using CPUE1 = log((catch + 0.5)/effort)
cpueSP2$CPUE1<- log((cpueSP2$alb_n+0.5)/cpueSP2$hook_set)

#Boxplot to remove outliers
boxplot(cpueSP2$CPUE1)
summary(cpueSP2$CPUE1)
cpueSP4<-  subset(cpueSP2,  cpueSP2$CPUE1<=-2.645  &  cpueSP2$CPUE1  >=  -
7.285)

#Writing the results of boxplot as file cpueSP4
write.csv(cpueSP4,file="cpueSP4.csv")

#Reading from the file cpueSP4
cpueSP4=read.csv("C:/Users/Sandeep  Singh/Documents/cpueSP4.csv",header=T)
names(cpueSP4)
dim(cpueSP4)
#GLM Tested
model_test_tot2 <- glm((CPUE1) ~
                  as.factor(ret_year), data=cpueSP4, weights=cpueSP4$weight)
```

```
model_test_tot2
model_test_tot3 <- glm((CPUE1) ~
            as.factor(catch_month), data=cpueSP4,weights=cpueSP4$weight)
model_test_tot3
model_test_tot4 <- glm((CPUE1) ~
            as.factor(year_qtr), data=cpueSP4, weights=cpueSP4$weight)
model_test_tot4
model_test_tot5 <- glm((CPUE1) ~
            as.factor(vessel_id), data=cpueSP4, weights=cpueSP4$weight)
model_test_tot5
model_test_tot6 <- glm((CPUE1) ~
            as.factor(catch_month)*poly(lat,degree=3),
data=cpueSP4,weights=cpueSP4$weight)
model_test_tot6
model_test_tot7 <- glm((CPUE1) ~
            as.factor(lat1)*as.factor(long1),                    data=cpueSP4,
weights=cpueSP4$weight)
model_test_tot7
model_test_tot8 <- glm((CPUE1) ~
            as.factor(year_qtr)+as.factor(catch_month),          data=cpueSP4,
weights=cpueSP4$weight)
model_test_tot8
model_test_tot9 <- glm((CPUE1) ~
            as.factor(year_qtr)+as.factor(catch_month)+    as.factor(vessel_id),
data=cpueSP4, weights=cpueSP4$weight)
model_test_tot9

model_test_tot10 <- glm((CPUE1) ~
            as.factor(year_qtr)+as.factor(catch_month)+
as.factor(lat1)*as.factor(long1), data=cpueSP4, weights=cpueSP4$weight)
model_test_tot10

model_test_tot1 <- glm((CPUE1) ~
```

```
as.factor(year_qtr)+as.factor(catch_month)+
as.factor(lat1)*as.factor(long1)+as.factor(vessel_id),                    data=cpueSP4,
weights=cpueSP4$weight)
model_test_tot1
model_test_tot11 <- glm((CPUE1) ~
                    as.factor(year_qtr)+as.factor(catch_month)*poly(lat,degree=3),
data=cpueSP4,weights=cpueSP4$weight)
model_test_tot11
model_test_tot12 <- glm((CPUE1) ~

as.factor(year_qtr)+as.factor(catch_month)*poly(lat,degree=3)+as.factor(lat1)*as.fac
tor(long1), data=cpueSP4,weights=cpueSP4$weight)
model_test_tot12
model_test_tot13 <- glm((CPUE1) ~
                    as.factor(year_qtr)+as.factor(catch_month)*poly(lat,degree=3)+
as.factor(vessel_id), data=cpueSP4,weights=cpueSP4$weight)
model_test_tot13

model_test_tot14 <- glm((CPUE1) ~

as.factor(year_qtr)+as.factor(catch_month)*poly(lat,degree=3)+as.factor(lat1)*as.fac
tor(long1)+as.factor(vessel_id), data=cpueSP4,weights=cpueSP4$weight)

model_test_tot14
model_test_tot15 <- glm((CPUE1) ~

as.factor(year_qtr)+as.factor(lat1)*as.factor(long1)+as.factor(vessel_id),
data=cpueSP4,weights=cpueSP4$weight)
model_test_tot15
model_test_tot16 <- glm((CPUE1) ~
                    as.factor(year_qtr)+as.factor(catch_month)+     as.factor(vessel_id),
data=cpueSP4,weights=cpueSP4$weight)
model_test_tot16
```

```
#Step to compute the AIC for each term in turn
model55 <- step(model_test_tot)

#Model summary and diagnostics
AIC(model_test_tot1, model_test_tot2, model_test_tot3, model_test_tot4, model_test_tot5, model_test_tot6,
model_test_tot7,model_test_tot8,model_test_tot9,model_test_tot10,model_test_tot11
,model_test_tot12,model_test_tot13,model_test_tot14,model_test_tot15,model_test_t
ot16)

#Diagnostic plots
par(mfrow= c(2,2))
plot(model_test_tot1)

#Year and month effects
par(mfrow= c(2,2))
termplot(model_test_tot1, cpueSP4, "Year", se= TRUE)
return

#Plotting residual
hist(model_test_tot1$residuals, breaks=20, xlab="Residuals",ylim=c(0,500),
main=NULL, col="light blue")

#Plotting residuals on a histogram
hist(model_test_tot1$residuals, breaks=20,xlim=c(-2,2),xlab="Residuals",col="light
blue", main=NULL,freq=FALSE)
# add normal curve
LL <- function(mu, sigma) {
  R = dnorm(model_test_tot1$residuals, mu, sigma)
  -sum(log(R))
}

library(stats4)
mle(LL, start = list(mu = 1, sigma=1))
```

```
curve(dnorm(x,mean = 0.001991768, sd = 0.582065940  ),from=-2, to = 2,add=T,
col="red", lwd=2)

#Plotting fitted values with residuals
plot(model_test_tot1$fitted.values,model_test_tot1$residuals,xlab="Fitted
values",ylab="Residuals",
    main=NULL)
abline(0,0)

#Plotting fitted values with absolute square root of residuals
plot(model_test_tot1$fitted.values,sqrt(abs(model_test_tot1$residuals)), xlab="Fitted
    values",ylab="sqrt(abs(residuals))",main=NULL)

#Plotting fitted values and ln CPUE
plot(model_test_tot1$fitted.values,model_test_tot1$y,xlab="Fitted
values",ylab="ln(CPUE)", main=NULL)

#Summary statistics of the selected model
summary_model_test_tot1 <- summary (model_test_tot1)
summary_model_test_tot1

#QQ Plot
qqnorm(model_test_tot1$residuals)
qqline(model_test_tot1$residuals, datax = FALSE)

#ANOVA table
anova(model_test_tot1, test="Chisq")

#Saving the file
save(summary_model_test_tot1, file="summary_model_test_tot1.RData")


#Calculating pseudo R squared
rsquared <- 1- (model_test_tot1$deviance/model_test_tot1$null.deviance)
```

```
# Constructing a standardised CPUE index from the selected model
# Setting up a new data frame first with each year quarter in the model
# Choosing a value for all the other factors in the model
# Chosen the first value in all cases.
#This should not matter - choosing another one also gives the same result when normalized
new.dat1 <- data.frame(year_qtr=as.factor(unique(cpueSP4$year_qtr)), catch_month=as.factor(cpueSP4$catch_month[1]), lat1=as.factor(cpueSP4$lat1[1]),
        long1=as.factor(cpueSP4$long1[1]),
vessel_id=as.factor(cpueSP4$vessel_id[1]))

# The model prediction of CPUE on the log scale
new.dat1$log.pred <- predict(model_test_tot1, newdata=new.dat1, type="response")

#Standardizing the CPUE
new.dat1$cpue.pred <- exp(new.dat1$log.pred)
## This is the standardised CPUE normalized (divided by its mean) which is how they are usually presented
new.dat1$cpue.pred.normalised <- new.dat1$cpue.pred/mean(new.dat1$cpue.pred)
#The end
```