unuftp.is Final Project 2014

EVALUATION OF APPROACHES TO ESTIMATE FISH CATCHES ON LAKE VICTORIA: A CASE STUDY USING CAS AND FRAME DATA FROM UGANDA

Nakiyende Herbert
National Fisheries Resources Research Institute
Plot 39/45 Nile Crescent, Jinja
P.O. Box 343, Jinja-Uganda
nakiyende@yahoo.ie

Supervisor:

Einar Hjorleifsson
Marine Research Institute
Skulagata 4, 121 Reykjavik, Iceland
einarhj@hafro.is

ABSTRACT

Since 2005, harmonized catch assessment surveys (CASs) have been implemented on Lake Victoria in the three riparian countries Uganda, Kenya, and Tanzania to monitor the commercial fish stocks and provide their management advice. The regionally harmonized standard operating procedures for CASs have not been wholly followed due to logistical difficulties. Yet the new approaches adopted have not been documented. This study investigated the alternative approaches used to estimate fish catches on the lake with the aim of determining the most reliable one for providing management advice and also the effect of current sampling routine on the precision of catch estimates provided. The study found the currently used lake-wide approach less reliable and more biased in providing catch estimates compared to the district-based approach. Noticeable differences were detected in catch estimates between different months of the year. The study recommends future analyses of CAS data collected on the lake to follow the district-based approach. Future CASs should also consider seasonal variations in the sampling design by providing for replication of sampling. The SOPs need updating to document the procedures that deviate from the original sampling design.

This paper should be cited as:

Nakiyende, H. 2018. *Evaluation of approaches to estimate fish catches on Lake Victoria: a case study using CAS and frame data from Uganda.* United Nations University Fisheries Training Programme, Iceland [final project].
http://www.unuftp.is/static/fellows/document/herbert14prf.pdf

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

The fisheries of Lake Victoria have for several decades contributed to the economic development in the riparian states; Uganda, Kenya, and Tanzania and to the livelihood of many people, especially those immediately dependent on the lake resources. The lake directly provide livelihoods for between 2–3 million people and indirectly to over 6 million people who depend on it in form of water, food and downstream activities. The lake fisheries contribute between 3–6% of the national GDPs in the three countries (WorldBank, 2009).

Key management advice for the fish stocks on major water bodies in the three East African countries is generated from fisheries monitoring studies including fishery independent gillnet, trawl and acoustic surveys, and the fishery dependent frame and catch assessment surveys (CASs). The information generated from latter surveys has delivered a set of indicators to aid decision-making in the context of policy and development planning and for the Lake Victoria fisheries. The estimated total annual catch that is obtained from these surveys is one of most important statistics used as input to many analytical stock assessment models.

All the fisheries studies undertaken on Lake Victoria should follow regionally harmonized Standard Operating Procedures (SOPs) for comparability of data used to inform uniform management decisions (MRAG, 2008). Since 2005, CASs have been conducted on Lake Victoria in the three countries (LVFO, 2014). Frame surveys on the other hand have been conducted on the lake every two years since 2000 (LVFO, 2012).

Successful management of fisheries resources world over depends in part on the reliability of management advice which in turn is determined by the methods used in collection and analysis of data (Sparre, 2000). The demand for reliable and effective management advice for Lake Victoria fisheries is increasing from national and regional fisheries management bodies. It is therefore crucial that available fishery dependent data are properly managed and that estimates from such data sets are accurately determined to provide the most reliable management advice for the fish stocks in the lake. This calls for a design of and adherence to standardized approach for collection and handling of fisheries data in the three countries.

1.2 Problem statement

As a result of logistical constraints, the implementation of CAS surveys has never adhered fully to the SOPs (LVFO, 2005). A key statistic that has rarely been collected but required in the SOPs algorithm is the number of boats that go fishing on the sampling day. In addition, the use of a specific database (EAFish) to store and process CAS data has not been possible due to technical complexities in using the database.

Due to the above, the data is recorded, stored and analyzed in MS Excel. The current formats of data recording and analysis are not fully standardized across different CAS surveys from different lakes. The raw data and analysis are intertwined and are hence difficult to review and reproduce.

Management of these spreadsheets is not coordinated nor centralized exposing valuable data to high risks of loss.

This study uses the CAS and frame survey data from the Ugandan part of Lake Victoria to explore alternative approaches to estimate catch landings on Lake Victoria in line with the procedures described in the SOPs. The study attempts to standardize and document alternative approaches with a view of updating the SOPs. Short reproducible codes are developed for CAS data analysis in R and MS Excel and improved data handling practices are proposed.

1.3 Significance of the study

Catch assessment surveys are the cornerstone in monitoring the fisheries in most water bodies in East Africa. The study seeks to improve the accuracy of estimates (quality of management advice) from CASs on Lake Victoria. The study creates simplified and reproducible analysis codes for CAS data with practical interim recommendations on good management practices for data stored and processed in MS Excel.

1.4 Scope of the study

The study investigates two approaches to analysis of CAS data to determine the most accurate one for providing catch estimates at a district and country level (Uganda). The study uses commercial catch data collected in 2005 and 2014 and frame data collected in 2012 from the Uganda portion of Lake Victoria.

1.5 Objectives of the study

The overall objective of the study was to explore and evaluate alternative approaches to Catch Assessment Survey data analysis and to develop standardized reproducible CAS data analysis. The specific tasks were to:

i. evaluate different approaches for estimating total catches by species,
ii. develop standardized and reproducible Excel and R scripts for the analysis,
iii. formulate good practices for handling and management of the data.

2 OVERVIEW OF CURRENT PROCEDURES

Monitoring of fish stocks on Lake Victoria started as early as 1928, when the first lake-wide fishing survey was implemented by the British colonial government (Graham, 1929; Cadwalladr and Stoneman, 1966). Collection of fisheries data on Lake Victoria in the past can be characterized by inconsistencies due to financial constraints (Cowx, 1996; Cowx *et. al*, 2003). In particular, data from fishery independent surveys have been limited and inadequate for the needed management advice. Fisheries resources monitoring programs on the lake gained importance in the last few decades due to the increasing need to sustain the lake fisheries resources based on sound management decisions. The establishment of the East Africa Freshwater Fisheries Research

Organization in the 1950s strengthened these efforts on the Lake but were neither harmonized nor coordinated. The first lake-wide catch assessment survey to be implemented on Lake Victoria followed a stratified random sampling design originally developed by G. Bazigos (Wetherall, 1972) and later by Muhoozi (2002). Similar attempts have continued under the coordination of East African Community (EAC) Lake Victoria Fisheries Organization (LVFO).

In 2005, the LVFO harmonized fisheries data collection around the lake including collection of information on fishing effort and fish catches through frame surveys and CASs and estimation of fish stock biomass and distribution through bottom trawl and acoustic surveys. Standard Operating Procedures (SOPs) detailing the data collection and analysis approaches were developed. A total of fourteen catch assessment and eight frame Surveys have been conducted on the lake since the harmonization of fisheries data collection in the three countries (LVFO, 2014; LVFO, 2012).

Fishery data collection has however never fully adhered to the CASs SOPs in part due to logistical constraints (LVFO, 2005a). The procedure describes that information on catches by species be estimated for each vessel-gear combination at each sampled landing sites. These statistics in addition to the frame survey are then supposed to be the basis for the estimates of catch of each species by district. The later then forms the basis for the estimates for each riparian country. This can be described as a "bottom-up approach". The current approach within each riparian country however follows more a "top-down approach" where the lake-wide catches by each riparian country is first calculated for each vessel-gear category. The sum by species (total landings estimates) are then split to estimate district landings based on the proportion of boats per district irrespective of the vessel-gear categories in each district. These procedures are not standardized nor are they documented.

The infrastructure for processing and management of fisheries data on Lake Victoria is also poorly developed and lacking in a number of aspects. The EAFish software, designed as a central database for the long-term storage and analysis of fisheries data from different fisheries studies in standard formats, is not utilized except for the frame surveys due to technical complexities in using the software (MRAG, 2008). This means an all too frequent correspondence with the designer for technical backstopping, even on minor issues, which is costly and unsustainable. The database has not been upgraded to link data sets from the different fisheries studies. This has resulted in the use of Microsoft Excel spreadsheets for storage and analysis of data.

The current formats of the spreadsheets are not standardized within or between the three countries. In this study, an attempt is made to standardize the current approaches in the analysis of data given the current reliance on Microsoft Excel. The focus is twofold; to improve the current algorithm in Microsoft Excel, with the aim of making the code more efficient, less prone to error and reproducible and to develop scripts in R that achieve the similar as well as the more complicated CAS analyses which are not easy to perform in Microsoft Excel. These could then be up scaled to standardize data analysis for the Lake Victoria CAS in the three countries. The study also compared the two approaches to analysis of CAS data with a view of determining which one is less biased and more accurate in providing district level catch estimates upon which district management advice is developed.

3 METHODS AND MATERIALS

3.1 Study area and scope

This study used the CASs for 2005– 2014 and the 2012 frame survey data sets collected from the Uganda portion of Lake Victoria. Between July 2005 and May 2014, fourteen Catch Assessment Surveys (CASs) were conducted at 54 in landing sites on the Uganda side of Lake Victoria (NaFIRRI, 2014) (Figure 1). These landing sites represent approximately 10% of all landing sites in the lake districts (Figure 2).

3.2 Data tidying and cleaning

The CAS and frame data sets were first cleaned in Excel with the aim of developing standard formats and codes for data entry and to make it easy to read and export into other analytical software such as R. CAS data collected in different survey periods were merged into a single Excel file for further cleaning and standardization. Additional cleaning was done in R while developing R scripts especially where landing site records of frame and catch data were not matching (Appendix 3). A copy of the cleaned data was made to serve as the analysis file for Excel while the raw data file was directly read into R for analysis. CAS data were analyzed separately in Excel and R to compare and validate results from the two analytical tools. Data used for R analysis were saved in different formats (.txt, .xls, .csv, .dat) that are readable in R but the final analysis file was in .txt format. The analysis followed procedures described in the LVFO harmonized standard operating procedures for collection of CAS data on Lake Victoria (LVFO, 2005).

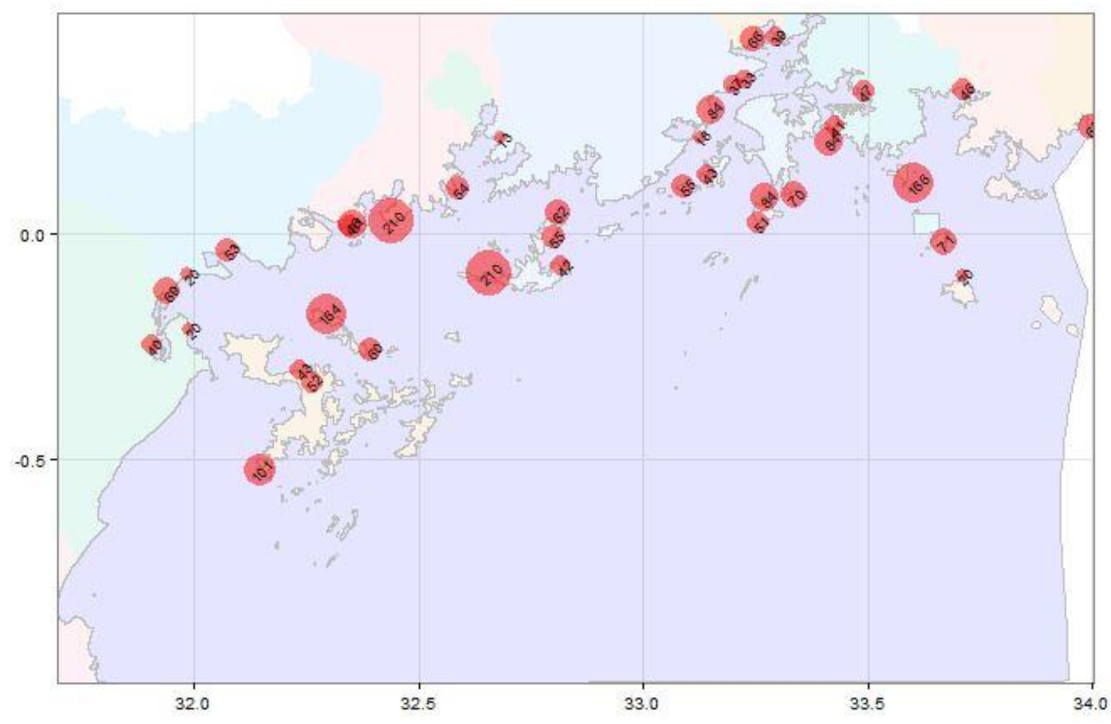


Figure 1: Catch Assessment Survey landing sites sampled between 2005 and 2014 and total number of boats as per the 2012 Frame Survey on the Uganda part of Lake Victoria.

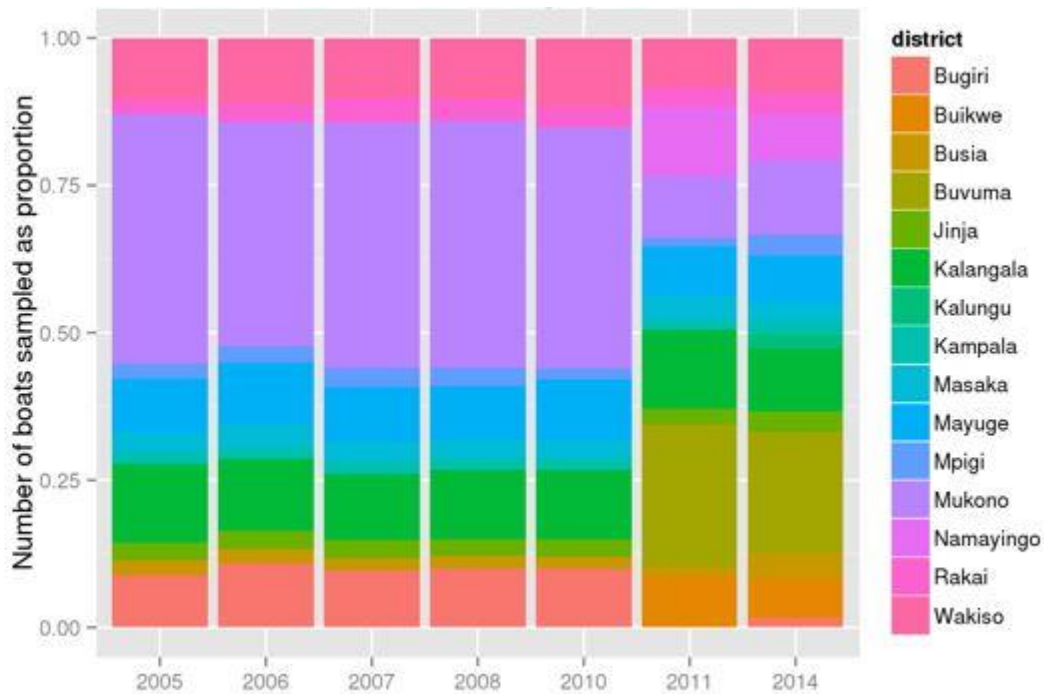


Figure 2: Proportion of boats sampled by district per year.

3.3 How data was collected

The CASs conducted in the Ugandan waters of Lake Victoria follow a design laid out in the approved standard operating procedures for catch assessment surveys on Lake Victoria (LVFO, 2005). This is a two-stage stratified sampling design. Within each district, a sample of primary sampling units (PSUs) i.e. the fish landing sites were first selected, and then, at each PSU, stratified samples of Secondary Sampling Units (SSUs) i.e. the vessel-gear types, are randomly selected by the field enumerator for sampling (GoU, 2003). A total of 56 PSUs have consistently been sampled in the Ugandan part of Lake Victoria since 2005. The landing sites were selected randomly with Probability Proportional to Size (PPS), where size is based on the number of vessels landing at the site. During the sampling period, the enumerators identified the numbers of all vessel-gear types at each landing site that landed or were expected to land during the sampling day and allocated sampling effort among the vessel-gear types in proportion to the number of vessels to be sampled. Sampling was done on four days in each sampled month, staggered to two consecutive days in the first and third or second and fourth weeks of the month. Regionally harmonized data forms were used to record field data. The enumerators were trained and provided with a field guide containing the data recording instructions to ensure effective data capture. Provision for close supervision of enumerators by sub-county fisheries officers and spot checks by district fisheries officers and officers from the national fisheries authorities, i.e. the National Fisheries Resources Research Institute (NaFIRRI), and the Department of Fisheries Resources (DFR) were made to ensure that data collection was done according to the laid down procedures and to eliminate fabricated records.

3.4 Data analysis and estimation of CAS based indicators

A. The approach defined in the CAS Standard Operating Procedures (LVFO, 2005), states that for each species and vessel-gear combination:

Calculate cpue (kg/boat/day) of the sampled boats in CAS survey at each sampled landing site.

Calculate the total catch per day at each sampled landing site (catch/day/site) by raising the cpue with total number of boats that landed during the CAS survey (Ncas).

Calculate monthly catch at each sampled landing sites (catch/month/site) by accounting for boat fishing activity (mean number of days fished in a month)

Calculate the total monthly landings by district (catch/month/district) by raising the above with the ratio of the total number of vessel-gear boats estimated in the FRAME survey (N) vs the total boats that landed during the CAS survey (Ncas).

The monthly catch by species in each district is then a simple summation of the catch by the vessel-gear combination and the lake wise monthly catch is then calculated as a summation of the catches by district.

Because of logistical difficulties, the total number of boats of each vessel-gear category that land during the CAS survey sampling days have not been estimated

B. The current procedure (lake-wide approach), states that for each species and vessel-gear combination:

Calculate mean cpue (kg/boat/day) of all the sampled boats in CAS survey on the Lake.

Calculate monthly catch (catch/boat/month) of the sampled boats by accounting for boat fishing activity (mean number of days fished in a sampled month).

Calculate the total monthly landings from the lake by raising the above with total number of boats (N) on the Lake in the FRAME survey (catch/month).

The monthly catch by species from the lake is then a simple summation of the catch of each vessel-gear combination. The estimated landings by district are then obtained by multiplying the total lake estimates with the proportion of boats in each district as determined from the frame survey. It should be noted that in this last step no account is taken of the vessel-gear combination, just the total number of boats.

C. An alternative approach, district-based calculation proposed by this study.

For each species and vessel-gear combination:

Calculate mean cpue (kg/boat/day) of all the sampled boats in a CAS survey within each district.

Calculate monthly catch (catch/boat/month) of the sampled boats by accounting for fishing activity (mean number of days fished in a sampled month).

Calculate the total monthly landings from the district by raising the above with total number of boats (Ndis) in the FRAME survey (catch/month/district).

The monthly catch by species in each **district** is then a simple summation of the catch by the vessel-gear combination and the lake wise monthly catch is then a summation of the catches by district.

3.4.1 Approach used in the study

In this study, the current lake-wide and the alternative district-based approach were compared and evaluated to determine the most appropriate for providing district and lake catch estimates. Analyses were performed in both Microsoft Excel and R to compare and validate the results. The study used CAS data collected in 2014 and Frame data collected in 2012 on the Uganda part of Lake Victoria. In addition, the study, used the 2005 CAS and 2012 Frame data to examine the effect of seasonality and frequency of sampling on the quality and accuracy of catch estimates. The fishing crafts were segregated into effort groups (vessel-gear categories) and mean fish catch rates (kg boat-1 day-1) were estimated for each effort group by species and district. For each effort group, the boat activity coefficient was estimated as the mean number of days boats in each effort group fished in a week divided by the number of days in a week multiplied by 30 days in a month. The catch (C) of each effort group was then estimated. The mean monthly estimates in each period were raised through 12 months to obtain annual catch estimates.

4 RESULTS

4.1 Data tidying and cleaning

Several mistakes were encountered with data stored in Microsoft Excel spreadsheets. Such inconsistencies included using different codes, names and connotations of locations; using different letter case where the names were consistent. In some cases, there were disparities between similar locations in the CAS and frame survey data sets making linkage of the two data sets difficult (Appendix 1). Other inaccuracies in the data included recording data in wrong columns e.g. the price of one species could be recorded in a column of weight or number of a different species or record of number in a column of weight for the same species. There were instances of recording several name codes for a single landing site by different data recorders. Some cases involved recording similar measurements in different units. The vessel length for example could be recorded in meters and feet within the same column. Some data from a single survey were entered in Microsoft Excel spreadsheets in varying formats and using different codes for the same variables. Besides, these data were scattered in different locations. Some data were recorded in such formats that was difficult to read into R and other software. Hence considerable time was spent in tidying and harmonizing the data sets before any analysis could be carried out. Some landing sites in the different districts share similar names and these are treated as a single variable in Microsoft Excel and R if only a column for landing sites is analyzed.

4.2 Analysis of CAS data in Microsoft Excel and R

Simple reproducible codes were developed in Microsoft Excel for analysis of CAS data. The new standardized codes were tested using the 2014 CAS and 2012 frame surveys to estimate the CAS indicators including CPUE and catches by species, district and for the lake. In Microsoft Excel a joint vessel gear category key in both the CAS and frame data was generated. The use of keys, such as above in pivot table analysis allow for an automatic linking of data from different tables (e.g. CAS and frame surveys) using the VLOOKUP function (Excel, 2010). Equivalent R scripts were developed to analyze the data. The main functions used in these scripts were the ddply, a function that performs tasks similar to the Microsoft Excel pivot table and the join function, an equivalent of the VLOOKUP function in Excel.

4.3 Overview of Catch Assessment Survey sampling design

A total of 52,000 boats have been sampled in 14 catch assessment surveys implemented on the Uganda side of Lake Victoria since the start of harmonized CASs in the three countries in 2005 (Table 1). The number of samples per survey has varied over time, and surveys have become less frequent with time. On average, 37,00 boats have been sampled per survey.

The CAS samples were obtained from stratified landing sites comprised of small sites (<100 boats), the medium (100 –200 boats) and large sites (>200 boats) (Figure 1). Generally, a larger proportion of boats were sampled at the smaller sites (Figure 3). The proportional sample size (number of boats) at the small sites ranged between 0.1 and 0.8 while that of the medium sites ranged from 0.1to 0.2 and 0.05 – 0.15 for the large sites (Figure 3).

Table 1: Sample size recorded as number of boats sampled in Uganda per survey per year**.**

						Month						
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**11**	**12**	**Total**
2005							3375	2948	3563	3715		13601
2006			2845					3976			3472	10293
2007			3792					3878				7670
2008		3738									3773	7511
2010				3769								3769
2011					4938							4938
2014					4285							4285

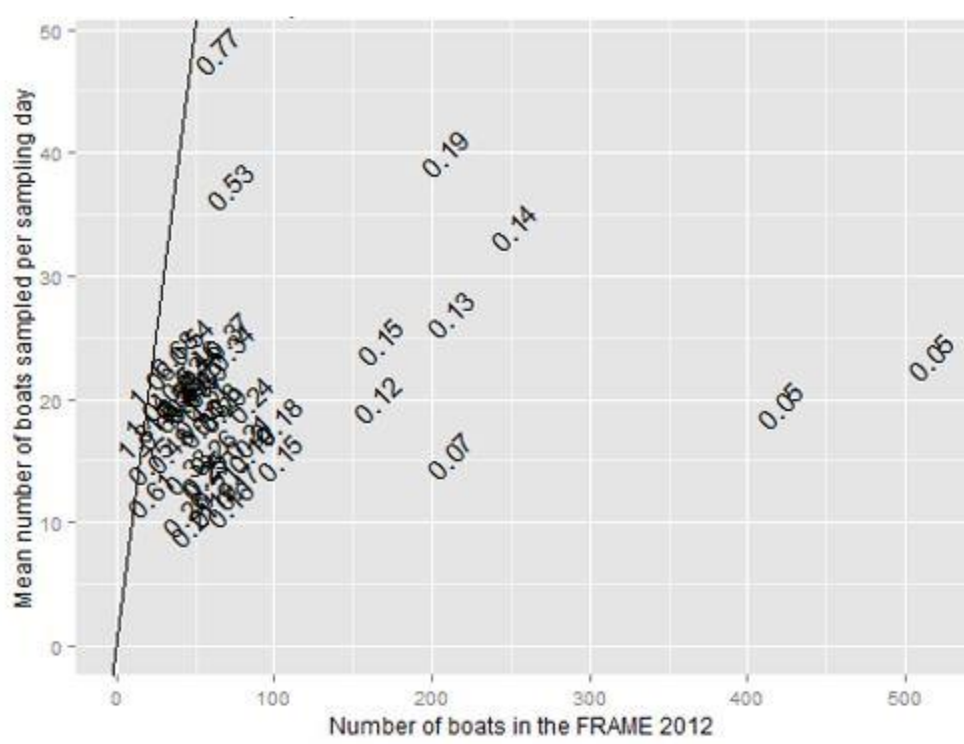


Figure 3: Number of boats sampled in the catch assessment survey 2014 relative to the total number of boats recorded in the 2012 frame survey at the sampled landing sites (numbers represent the proportion of boats sampled per landing site strata).

4.4 Approaches to CAS data analysis

Results of the bottom-up (lake-wide) and top-down (district) based approaches to estimation of Uganda's catch landings showed substantial differences by species, district and Lake. (Tables 2 & 3). Overall, the lake-wide based approach gave substantially higher estimates of catch in 2014 for the lake (12%) and district (20–78%) than the district-based approach, except for the five districts of Bugiri, Kalangala, Kalungu, Mukono and Rakai (Table 2). The lake-wide approach also gave higher catch estimates for all the species recorded in the 2014 CAS (Table 3).

4.5 The main fisheries in the 2014 CAS on Lake Victoria, Uganda

Nile perch (*Lates niloticus*), tilapia (*Oreochromis spp*), and a silver cyprinid (*Rastrineobola argentea*) are the main commercial fisheries on the Uganda side of Lake Victoria. The two approaches to catch estimation showed substantial differences for all the three commercial species in the fifteen districts. Overall, the lake-wide approach resulted into higher estimates of catches for each species in most districts (Figures 4, 5 & 6).

Table 2: The lake-wide and district-based approaches to estimation of total landings (t) in each district in 2014.

	Lake-wide Approach	District Approach	% Difference
District	**Catch (t)**	**Catch (t)**	
Bugiri	867.1	254.1	70.7
Buikwe	14,326.9	20,623.7	(44.0)
Busia	1,907.7	728.1	61.8
Buvuma	53,800.5	42,900.2	20.3
Jinja	3,102.4	746.7	75.9
Kalangala	44,512.6	61,119.0	(37.3)
Kalungu	1,734.3	2,002.2	(15.4)
Kampala	1,416.3	400.5	71.7
Masaka	10,636.8	1,941.3	81.7
Mayuge	26,090.9	11,535.2	55.8
Mpigi	7,091.2	1,758.2	75.2
Mukono	36,939.7	50,438.7	(36.5)
Namayingo	38,201.8	26,814.8	29.8
Rakai	5,405.1	7,881.2	(45.8)
Wakiso	23,489.6	8,718.7	62.9
Total	**269,522.9**	**237,862.5**	**11.7**

Table 3: The lake-wide and district-based approaches to estimation of catch landings (t) by species in 2014.

Species	Lake-wide Approach	District Approach	% Difference
Silver fish/Dagga	165,907.3	147,162.7	11.3
Nile perch	67,496.6	66,500.8	1.5
Tilapia spp	21,119.0	12,990.0	38.5
Haplochromines	5,179.9	3,928.6	24.2
Bagrus spp	290.1	246.1	15.2
African lung fish	1,671.2	1,159.2	30.6
Clarias spp	1,201.4	1,137.3	5.3
Other species	6,657.5	4,737.8	28.8
Total	**269,522.9**	**237,862.5**	**11.7**

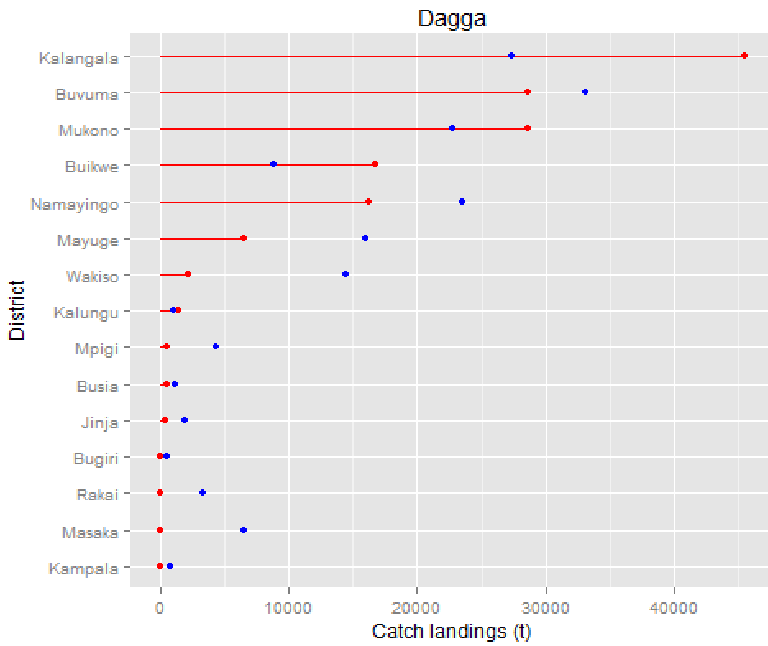


Figure 4: Variation between the district (red) and lake-wide based approaches to estimates of catch landings of Dagga by district.

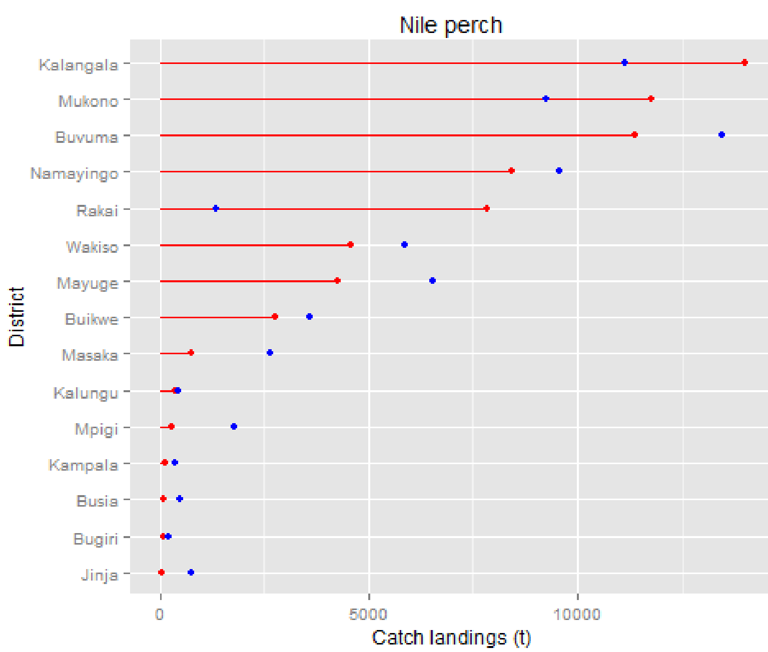


Figure 5: Variation between the district (red) and lake-wide based approaches to estimates of catch landings of Nile perch by district.

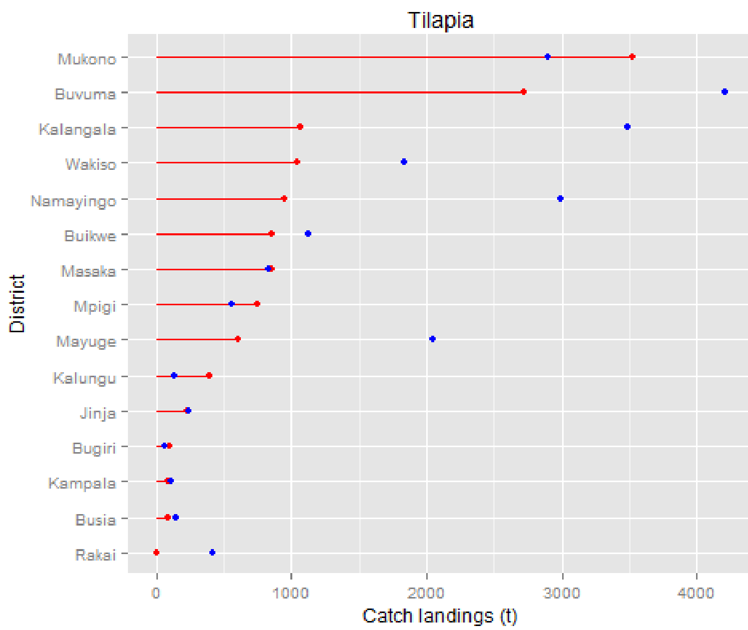


Figure 6: Variation between the district (red) and lake-wide (blue) based approaches to estimates of catch landings of Tilapia by district.

4.6 Catch composition by vessel and gear type

A total of eight fish species or species groups (tilapia spp=TL, *Protopterus aethiopicus*=PA, Nile perch=LN, haplochromine spp=HA, Dagaa=DA, Clarias spp=CG, Bagrus spp=BD and other unidentified species lumped=OT) were recorded in the commercial fisheries during the 2014 CAS on the Uganda side of Lake Victoria (Figures 7 & 8). These species groups were landed by four main vessel categories; Ssesse flat at one end (SF), Ssesse pointed at both ends (SP), parachute boats (PA) and other unidentified vessel groups (OT). Similarly, the species groups were targeted by different gear types; beach seines (BS), cast nets (CN), gill nets (GN), hand lines (HL), long-lines (LL), scoop nets (SN), small seines (SS), traps (TR), and categories (OT) not desribed.

All the species groups showed variations in catch landed in the different vessel groups (Figure 7) and gear types (Figure 8). The SF is the main vessel type used in the fisheries exploitable (Figure 7), followed by the PA. Gill-nets on the other hand were the most frequently used gear in the 2014 CAS (Figure 8). The main fishing gear for the dagaa and haplochromine fisheries were small seines. Longlines were mainly used in the harvest of the large species groups Nile perch, Clarias spp, Bagrus spp and *Protopterus aethiopicus*.

4.7 Differences in monthly estimates

In 2005, four separate CAS surveys were conducted in the months of July, August, September and November. An analysis where each survey was treated as a sole data source available to estimate total annual landings show that there can be substantial difference in the estimates (Figure 9). The catch landings of dagaa for example in August were two-fold that of November (Figure 9), suggesting a possibility inaccuracy of the current catch estimates based on catch landings from only one sampling (month). The data were however too limited to detect if this difference is due to seasonality in the fisheries of different species or if this is just a reflection of variability in the data.

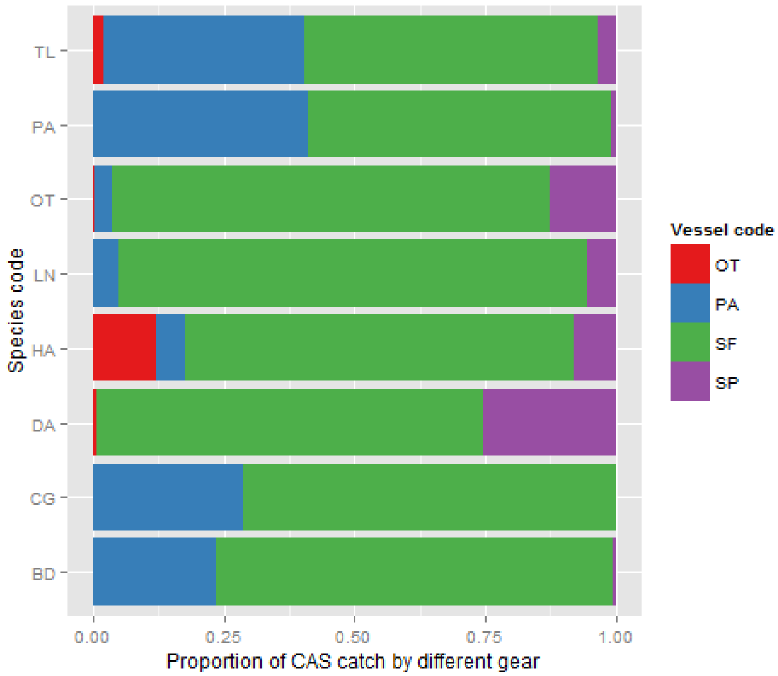


Figure 7: Proportion of catch landed by species in the fishing vessel categories operated on the Uganda part of Lake Victoria in the 2014 CAS.

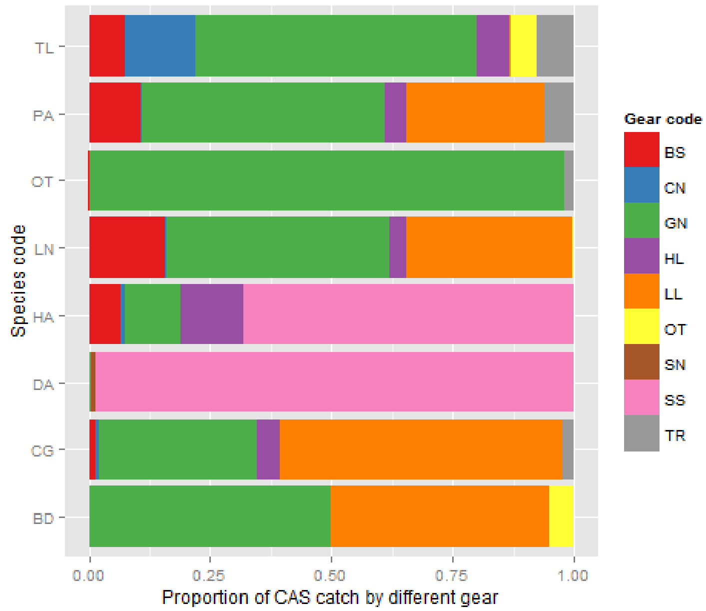


Figure 8: Proportion of catch landed by species in the different fishing gears used on the Uganda side of Lake Victoria in 2014 CAS.

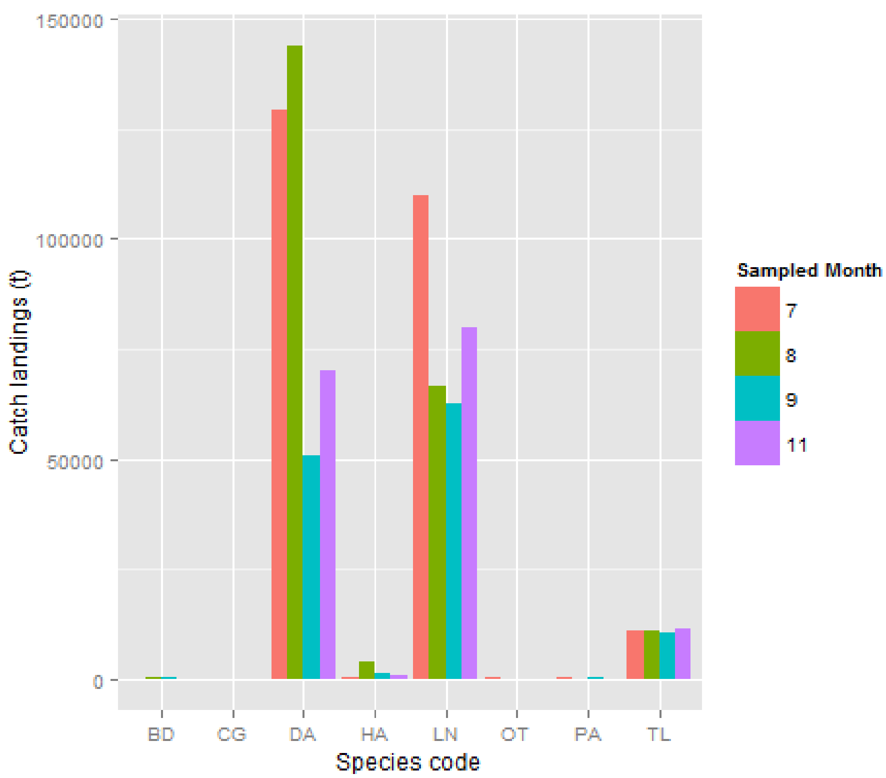


Figure 9: Estimation of annual catch (t) in Uganda by species based on CAS surveys in the different months in 2005.

5 DISCUSSION

5.1 Recording, storage and management of CAS and Frame data

Using different notations and codes to record similar or identical data, entering data into a wrong column and recording similar data in different units of measure is misleading. It is practically impossible to use such data by scientists who never participated in its collection or those not familiar with the lake system from where data is collected. Such data is too demanding in terms of further tidying and harmonizing before a meaningful analysis is performed. R is case sensitive and any slight change in spelling, letter-case or code is treated as a new attribute. Excel is not case sensitive but responsive to spelling. Analysis of poorly recorded data results in misleading, inaccurate and biased estimates of parameters. Data recorded in different units of measure similarly gives biased and unrealistic estimates of parameters.

Related or similar data stored in scattered spreadsheets and directories is prone to misuse and vulnerable to loss. It is evident that obtaining fisheries data in developing countries remains a major challenge due to financial limitation for which reason there are inconsistencies in data collection (Cowx *et. al*, 2003). Standardized recording of data by maintaining codes, names and data recording formats is needed to avoid data inconsistencies in future research. All personnel handling fisheries data need to be sensitized on the necessity for good data recording and storage. As an interim solution, i.e. until the practice of entering and storing the raw data in a standardized database becomes the routine, a single directory in Microsoft Excel with sub-folders should be used to stores similar and related data. These may be separated by water body, data type, year of data collection and survey reference. There is need for some personnel entrusted with the overall storage and management of stock assessment data.

5.2 Standardized analysis of CAS data in Microsoft Excel and R.

Until this study, it has been difficult to follow and reproduce analysis of CAS data performed in Microsoft Excel. No attempts had been made by the CAS regional working group to analyze the Lake Victoria CAS data in R. CAS analyses have been based on producing long pivot tables and have involved a lot of copy and paste because the vessel and gear code columns have been treated separately in the pivot table analysis. The vessel gear codes in the frame and CAS files were manually matched which slows the analysis process and increases the risk of human introduced errors. The whole process used so far can be characterized by slow, hard to reproduce and human error prone procedures.

The approach developed in this study allows all analysis to be completed in a short and easy to follow pivot table, allows automatic match of variables like the vessel gear codes in frame and CAS spreadsheets even when stored in different locations. The study utilizes two Microsoft Excel functions, “paste" to join the vessel gear codes into a single parameter and the VLOOKUP to automatically match the vessel gear codes in frame and CAS (Excel, 2010). Analysis procedures by this study thus reduce the time spent on analysis while producing more accurate error free results. The procedures developed in this study would contribute to timely provision of management advice through a speedy, error free and efficient analysis of CAS information in a familiar and user-friendly environment of Microsoft Excel. To further improve the efficiency and timely analysis of fisheries data, standardized R scripts (Micheal, 2007) were developed for analysis of similar data. R codes serve as a written record of what was done, making it easy to reproduce thus reducing the time spent on analysis. Besides, the visual outputs (graphs and maps) are of superior quality than those of Microsoft Excel.

5.3 Catch Assessment Survey sampling design on Lake Victoria

Although the overall sampling design for CASs (LVFO, 2005) is representative covering the entire lake in a stratified fashion comprised of small landing sites (<100 boats), medium landing sites (100 - 200 boats), and large landing sites (>200 boats) (figure 1), results of the 2014 CAS show unrepresentative sampling in the different landing sites. Higher proportions of boats are sampled in the small landing sites (up to 0.77) than in the medium (0.2) and large (0.05) landing sites. Frame data have shown the medium to large size landing sites to constitute more vessel gear categories than the smaller landing sites. And the vessel gear categories form the basis for estimation of catch landings. Collection of fewer samples from the medium and large landing sites may affect some vessel gear categories in terms of being under represented or even missed out and this in turn could lead to biased estimation of fish catches on the lake. As much as the proportion of boats to be sampled at each CAS landing site depends on the total number of boats by gear active on the sampling day, persistent occurrence of this pattern should be a concern to the stock assessment scientists and hence need to be investigated to rule out biased sampling that could lead to inaccurate estimation of catches and wrong management advice. The study recommends a design that offers representative sampling of all the primary sampling units. Efforts need to be directed towards having more boats sampled at the medium and large landing sites.

Until up to 2010, Mukono district had the highest number of boats sampled per year but was overtaken by Buvuma district in the later years (Figure 2). Prior to the 2011 CAS surveys, the three districts Mukono, Buvuma and Bwike were under one district Mukono. Since 2011, there have been changes in district administrative boundaries with new districts created from the existing (Figure 2). Mukono was split up into three administrative units thus the decrease in the sample size for Mukono district. These changes however affect the location of CAS landing sites as these may shift to new districts and their original names may also change. The study proposes a unique identification number for each of the CAS landing site (primary sampling unit) so that even when such changes occur, landing sites can easily be traced.

5.4 Inconsistency in CAS sampling frequency

The variations in sampling frequency between sampled years (Table 1) could have an effect on the credibility of catch estimates from CASs mainly due to differences in sample size and due to seasonality differences. The Implementation of a Fisheries Management Plan (IFMP) project funded CASs between 2005 and 2008 but since the end of the project, CASs have been plagued by break-ups in sampling with some years missing out completely on CAS data collected while those where sampling has been possible, only a single survey has been implemented. Effects of these differences in sampling were evidently detected in the catch estimates between the sampled months in 2005 (Figure 9). Some months registered significantly higher catches for the same species than others. This variation is a clear indication that calculating annual landings based on catch estimates from only one sampling (month) is unrealistic. To improve the accuracy of estimates from CASs, a sampling design that provides for repeated sampling taking into account seasonal variations of the fisheries is recommended.

5.5 Differences in catch estimates between district and lake-wide based approaches

The discrepancy in catch estimates from the two approaches to CAS data analysis (Table 2 & 3; Figure 4, 5, & 6) should be of concern to fish stock assessment scientists dealing with the Lake Victoria CAS data. The CAS reports are required to provide information on catches by species, district and then the lake catches. The lake-wide based approach splits the lake catch among districts based on the proportion of boats for each district in the frame, disregarding the vessel gear categories in each district. The district approach on the other hand takes into account the vessel gear categories in each district while estimating district and lake catches. It is clear in the Frame survey report (LVFO, 2012) that different districts have different vessel-gear compositions. Associated to these differences, are variations in species composition and quantities landed in the different districts. Some districts lack some fisheries completely. Apportioning the lake catch to districts without accounting for the above differences gives biased estimates of catches and wrong fisheries management or development advice. For instance, over estimating dagaa catches in a district where such a fishery does not exist may make management develop infrastructure for dagga fishery in the wrong place. The current lake-wide based approach to CAS data analysis is not the best practice. The adoption of the more realistic district-based approach in CAS data analysis is proposed.

6 CONCLUSION

The current lake-wide based approach to CAS data analysis is not appropriate. Given the available data, the district-based approach provides more accurate estimates of species, district and lake catch landings and thus is more reliable for generating management advice. The current CAS sampling design in SOPs is not wholly followed. There is need to update the SOPs, particularly to document the current procedures that deviate from the original design. If fully utilized, Microsoft Excel remains a very useful tool for reproducible analysis of CAS data. R is however a more powerful analysis software that should be considered in future analysis of these data.

7 RECOMMENDATIONS

7.1 Sampling design, data collection and analysis

1. Future CASs need to take into account variations in the catch by months. Further investigations need to be undertaken to detect the effect of seasonality and sampling frequency on the accuracy of estimates from CASs as a basis to design a suitable sampling routine for CASs.
2. District-based approach of CAS analysis is more realistic than the lake-wide based approach in providing catch estimates and it fits well in the current data collection design. It is recommended that this approach be considered in the future analysis of CAS data in the region.

3. Future CASs should ensure representative sampling of all the primary sampling units. More boats need to be sampled at the medium and large size primary sampling units (landing sites) than currently sampled.
4. Scientists dealing with fisheries data should ensure that all the analysis performed are easy to follow and reproduce by other scientists. R is most suitable to achieve this, but full utilization of the available Microsoft Excel functions could help.

7.2 Data handling, storage and management

1. There is need to develop a checklist of spellings of names used and attributes measured in any given survey and these must be adhered to by all the research team members.
2. Before undertaking any field survey, there should be a preparatory meeting of all members to be involved in the survey to agree on the survey design, re-emphasize the need for uniformity and consistence in data capture and any other issues relevant to the survey. This should be a routine activity before any survey is implemented.
3. Uniform and standard forms and formats should be used for both field data collection and data entry and should be adhered to by all members involved in the survey to avoid any divergence that may result into sampling errors and bias in estimates.
4. Prior to data entry and analysis, there should be a meeting by all staff involved in data processing to agree and standardize formats to be used. But for comparability, standard formats should be used in similar surveys and related data and any changes should be noted in the reports from such surveys.
5. There should be a data cleaning session to harmonize names both in frame and catch assessment surveys before analyses are performed.
6. There is need for consistence in notations used for sampling locations. Where a location is known by two names, either one or both should be maintained for consistence.
7. A unique identification number should be assigned to each CAS landing site to overcome problems associated with changes in administrative boundaries and those where two landing sites in different districts share a similar name.
8. All the three research institutions in Uganda, Kenya and Tanzania and the Lake Victoria Fisheries Organization have well established websites. However, most of the useful technical reports from fisheries studies are not accessible on these sites and information from such reports is difficult to cite. Efforts should be made to upload this valuable information for easy access and future reference.
9. There is need for regular update of the fisheries survey SOPs to document all the procedures especially when there are deviations from the original designs.
10. There is need to improve the handling, storage and management of data in Microsoft Excel. In absence of a database, an organized data directory in Microsoft Excel (Appendix 4) where data is stored by water body, data type and survey period, with a clear separation between raw and analysis files is suggested in this study.
11. Data handling and management should be included on the UNU-FTP introductory course.

ACKNOWLEDGEMENTS

First, I extend special thanks to my dear wife Nakasiita Flavia Gloria for her continued support and encouragement and to our lovely sons Nakiyende Mark, Nakiyende Mathew and Nakiyende Marvin for enduring the lonely six months I spent away from them. I am grateful to the profound love and guidance from my parents Mr and Mrs Mudiima Sande George William.

I thank the United Nations University of Fisheries Training Program staff Dr. Tumi Tomasson, Mr. Thor H. Asgeirsson, Ms. Sigridur Kr. Ingvarsdottir and Ms. Mary Frances Davidson for giving me a chance to participate in this training and for the guidance provided during the course. I sincerely appreciate the dedicated guidance from my supervisor Einar Hjorleifsson towards the production of this thesis.

I thank the director of the National Fisheries Resources Research Institute (NaFIRRI), Dr. Balirwa John and my institute supervisors Dr. Taabu Anthony Munyaho and Dr. Mbabazi Dismas for endorsing me to participate in this training. I thank Mr. Bassa Samuel, Mr. Muhumza Elias and Ms. Kagoya Esther of NaFIRRI for the support rendered throughout the training.

LIST OF REFERENCES

Cadwalladr, D. A. and Stoneman, J. (1966). *A review of the fisheries of Uganda waters of Lake Albert (East Africa, 1928–1965/66, with catch data mainly from 1953). East Africa Commission Services Organization, EAFFRO Supp. Publ. No. 3: 19p.*

Cowx, I. (1996). The integration of fish stock assessment into fisheries management. In: Cowx, I.G. (Ed.), stock assessment in inland fisheries, pp. 495–506.

Cowx, I., and Knaap, M. Van der. (2003). Improving fishery catch statistics for Lake Victoria. *Aquatic Ecosystem Health and Management* 6(3):299–310.

Excel (2010). Formulas and functions. *Microsoft Excel 2010*. [December 2014] <www.usd.edu/academics/center-for-teaching-and-learning/upload/Excel-3-Formulas-and Functions.pdf>

GoU. (2003). *Statutory instruments 2003 No. 35. The fish (BEACH MANAGEMENT) Rules, 2003.* Entebbe: The Uganda Gazzette.

Graham, M. (1929). *The Victoria Nyanza and its fisheries. A report on the fishing survey of Lake Victoria 1927-1928, 225 pp.* London: Crown Agents for the Colonies.

LVFO. (2005). *Standard operating procedures for catch assessment surveys. 42pp.* Jinja: Lake Victoria Fisheries Organization.

LVFO. (2012). *Regional status report on Lake Victoria bi-ennial frame surveys between 2000 and 2012 for Kenya, Tanzania and Uganda.* Jinja: Lake Victoria Fisheries Organization.

LVFO. (2014). *Regional Catch Assessment survey synthesis report, June 2005 to April 2014.* Jinja: Lake Victoria Fisheries Organization.

Micheal, J. (2007). *The R Book.* Imperial college London at Silwood Park, UK: John Willy & Sons, Ltd.

MRAG. (2008). *Provision of technical services for the implementation of a fisheries management plan for Lake Victoria project No. 8 ACP ROR 29. Final report, July 2004 – August 2008.* Jinja: Lake Victoria Fisheries Organization.

Muhoozi, L. (2002). *Exploitation and management of the artisanal fisheries in the Ugandan waters of Lake Victoria.* PhD thesis, University of Hull U.K.

NaFIRRI. (2014). *National report of the catch assessment surveys conducted on the Uganda part of Lake Victoria in May 2014.* Jinja.

Sparre, P. (2000). *Manual on sample-based data collection for fisheries. Examples from Vietnam.*

Wetherall, J. (1972). *Catch Assessment Survey (CAS) of Lake Victoria. EAFFRO Occasional Paper 14: 57 pp.*

WorldBank. (2009). *LVEMP II project appraisal document. Report No.45313-AFR, 197 pp.*

APPENDICES

Appendix 1: Using different spellings of names of the same place or location.

Name in Frame	Varied names in CAS		
Zinga	Zzinga	Zingga	
Luuku	Luuku/Nabisukiro	Nabisukiro	Luuku-Nabisukiro
Nakiga	Nakigga	Nakkiga	
Kaaza	Khaaza	Kazha	
Kaggulube	Kagulube	Kagulubbe	
PortBell	Port-Bell	Port Bell	

Appendix 2: Annual catch (t) by district and by species calculated using the lake-wide approach for the CAS data collected in 2014 on Lake Victoria, Uganda.

	catch (t)									
District	**NP**	**TL**	**DA**	**HA**	**BD**	**PA**	**CG**	**OT**	**Total**	**%**
Bugiri	217.2	67.9	533.8	16.7	0.9	5.4	3.9	21.4	867.1	0.3
Buikwe	3587.9	1122.6	8819.1	275.3	15.4	88.8	63.9	353.9	14326.9	5.3
Busia	477.7	149.5	1174.3	36.7	2.1	11.8	8.5	47.1	1907.7	0.7
Buvuma	13473.3	4215.6	33117.4	1034.0	57.9	333.6	239.8	1328.9	53800.5	20.0
Jinja	776.9	243.1	1909.7	59.6	3.3	19.2	13.8	76.6	3102.4	1.2
Kalangala	11147.3	3487.9	27400.1	855.5	47.9	276.0	198.4	1099.5	44512.6	16.5
Kalungu	434.3	135.9	1067.5	33.3	1.9	10.8	7.7	42.8	1734.3	0.6
Kampala	354.7	111.0	871.8	27.2	1.5	8.8	6.3	35.0	1416.3	0.5
Masaka	2663.8	833.5	6547.6	204.4	11.4	66.0	47.4	262.7	10636.8	3.9
Mayuge	6533.9	2044.4	16060.5	501.4	28.1	161.8	116.3	644.5	26090.9	9.7
Mpigi	1775.8	555.6	4365.0	136.3	7.6	44.0	31.6	175.2	7091.2	2.6
Mukono	9250.8	2894.5	22738.6	709.9	39.8	229.0	164.7	912.5	36939.7	13.7
Namayingo	9566.9	2993.4	23515.5	734.2	41.1	236.9	170.3	943.6	38201.8	14.2
Rakai	1353.6	423.5	3327.2	103.9	5.8	33.5	24.1	133.5	5405.1	2.0
Wakiso	5882.5	1840.6	14459.2	451.4	25.3	145.7	104.7	580.2	23489.6	8.7
Total	**67496.6**	**21119.0**	**165907.3**	**5179.9**	**290.1**	**1671.2**	**1201.4**	**6657.5**	**269522.9**	**100**
%	**25.0**	**7.8**	**61.6**	**1.9**	**0.1**	**0.6**	**0.4**	**2.5**	**100**	

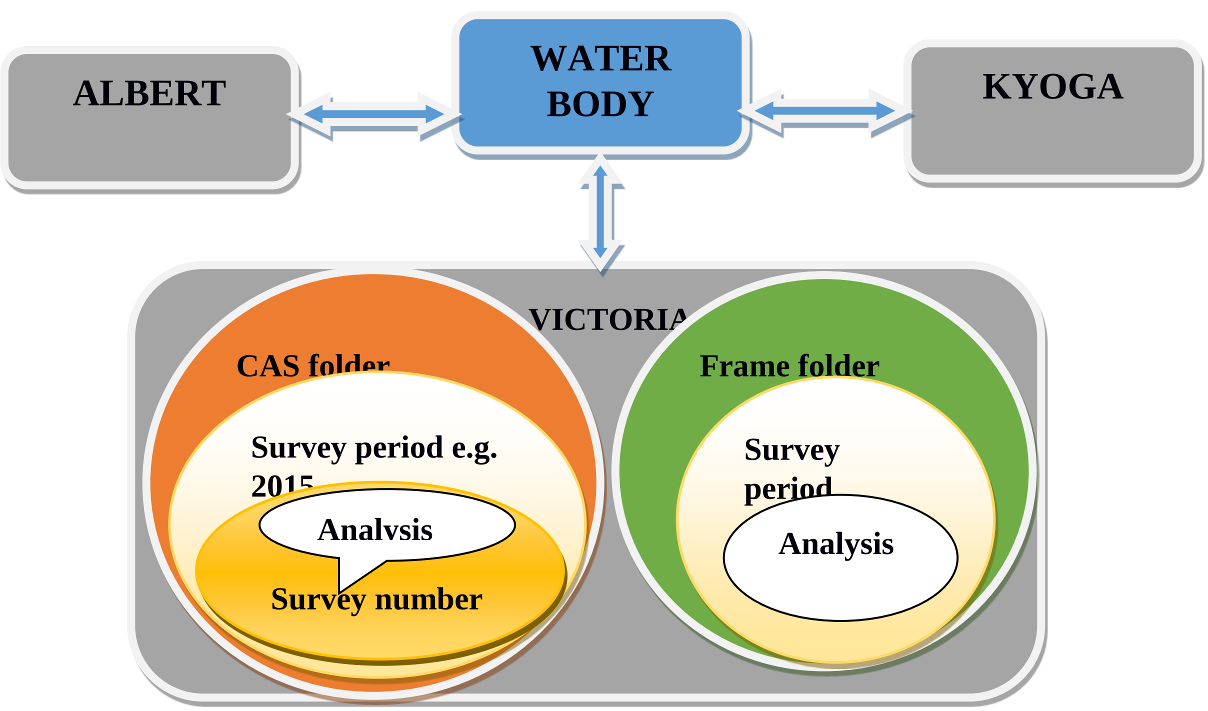
NP=Nile perch, TL=Tilapia, DA=Dagga, HA=Haplochromines, BD=Bagrus spp, PA=Protopterus, CG=Claria spp, OT=Unidentified species combined

Appendix 3: Annual catch (t) by district and by species calculated using the district approach for the CAS data collected in 2014 on Lake Victoria, Uganda.

	Catch (t)									
District	**NP**	**TL**	**DA**	**HA**	**BD**	**PA**	**CG**	**OT**	**Total**	**%**
Bugiri	48.5	94.5	65.8	35.4	0.0	9.4	0.7	0.0	254.1	0.1
Buikwe	2746.3	858.0	16652.4	342.1	4.0	8.5	3.1	9.3	20623.7	8.7
Busia	93.7	78.9	270.5	3.3	0.0	3.6	8.3	269.7	728.1	0.3
Buvuma	11349.4	2606.4	28662.5	0.0	36.7	51.3	46.8	147.1	42900.2	18.0
Jinja	58.7	190.9	372.0	70.8	0.0	48.1	5.4	0.8	746.7	0.3
Kalangala	14025.1	1069.3	45535.2	349.9	17.5	98.1	19.9	4.0	61119.0	25.7
Kalungu	109.2	323.8	1484.3	12.9	0.0	65.1	7.0	0.0	2002.2	0.8
Kampala	141.2	59.2	0.0	0.0	0.0	85.3	109.3	5.4	400.5	0.2
Masaka	770.6	839.2	0.0	13.7	0.0	316.2	1.7	0.0	1941.3	0.8
Mayuge	4242.9	605.7	6582.1	43.4	1.6	39.8	16.2	3.5	11535.2	4.8
Mpigi	293.2	735.8	574.9	15.4	1.9	7.9	129.1	0.0	1758.2	0.7
Mukono	11782.5	3522.9	28607.7	2618.9	92.5	256.1	22.9	3535.4	50438.7	21.2
Namayingo	8415.6	954.2	16203.3	422.8	88.5	5.9	0.0	724.5	26814.8	11.3
Rakai	7848.5	4.2	0.0	0.0	3.4	0.0	25.2	0.0	7881.2	3.3
Wakiso	4575.3	1047.3	2152.2	0.0	0.0	164.0	741.8	38.0	8718.7	3.7
Total	**66500.8**	**12990.0**	**147162.7**	**3928.6**	**246.1**	**1159.2**	**1137.3**	**4737.8**	**237862.5**	**100.0**
%	**28.0**	**5.5**	**61.9**	**1.7**	**0.1**	**0.5**	**0.5**	**2.0**	**100**	

NP=Nile perch, TL=Tilapia, DA=Dagga, HA=Haplochromines, BD=Bagrus spp, PA=Protopterus, CG=Claria spp, OT= Unidentified species combined.

Appendix 4: Proposed interim data storage and management design in Microsoft Excel.



Appendix 5: Composition of fishing crafts by vessel gear category and sampling proportion in CAS on the Ugandan part of Lake Victoria (Frame survey 2012, CAS 2014).

Vessel Gear (VG) type	Number of boats	Boats sampled in CAS	% sampled
Catamaran-Small Seines	8	5	62.5
Foot Fisher-Cast Net	12	3	25.0
Foot Fisher-Gill Net	77	6	7.8
Foot Fisher-Hand Line	235	40	17.0
Foot Fisher-Trap	16	1	6.3
Parachute-Beach Seine	135	5	3.7
Parachute-Cast Net	536	109	20.3
Parachute-Gill Net	1447	419	29.0
Parachute-Hand Line	1187	115	9.7
Parachute-Long Line	197	74	37.6
Parachute-Other gears	1032	4	0.4
Parachute-Trap	247	89	36.0
Rafts-Hand Line	75	6	8.0
Ssesse Flat-Beach Seine	964	244	25.3
Ssesse Flat-Cast Net	588	38	6.5
Ssesse Flat-Gill Net	7822	1598	20.4
Ssesse Flat-Hand Line	1067	132	12.4
Ssesse Flat-Long Line	5239	588	11.2
Ssesse Flat-Other gears	824	21	2.5
Ssesse Flat-Scoop Net	548	4	0.7
Ssesse Flat-Small Seine	1927	246	12.8
Ssesse Flat-Trap	66	33	50.0
Ssesse Pointed-Beach Seine	132	41	31.1
Ssesse Pointed-Cast Net	101	13	12.9
Ssesse Pointed-Gill Net	385	121	31.4
Ssesse Pointed-Hand Line	151	37	24.5
Ssesse Pointed-Long Lines	802	69	8.6
Ssesse Pointed-Other gears	172	6	3.5
Ssesse Pointed-Small Seine	1731	213	12.3
Ssesse Pointed-Trap	23	5	21.7
Total	**27746**	**4285**	**15.4**

Appendix 6: R scripts used in the analysis of CAS data (only Uganda data).

Required packages for the analysis

```
require(ggplot2)
require(ggmap)
require(gdata)
require(reshape2)
require(plyr)
require(lubridate)
require(stringr)
require(Hmisc)
```

GIS data

```
attach("data/geo.rda")
attach("data/cas.gis.rda")
```

Reading frame survey data (.txt format) into R

```
frame.raw <- read.table("data_raw/FS_Data_2012.txt", header=TRUE, sep="\t",  skip=1,
stringsAsFactors=FALSE)
```

Reading CAS survey data (.txt format) into R

```
cas.raw <- read.table("data_raw/CAS_Victoria_2005-2014_cleaned_Herbert.txt", header=TRUE, sep="\t",
skip=1, stringsAsFactors=FALSE)
```

Preliminary cleaning and defining some key attributes

```
cas.raw$date <- mdy(cas.raw$date)
cas.raw$year <- year(cas.raw$date)
cas.raw$month <- month(cas.raw$date)
cas.raw <- cas.raw[!is.na(cas.raw$year),]

i <- cas.raw$gCode %in% c("BES","BOS")
cas.raw$gCode[i] <- "BS"

cas.raw$wDA <- cas.raw$nDA * cas.raw$wDA
cas.raw$wHA <- cas.raw$nHA * cas.raw$wHA
cas.raw$id <- 1:nrow(cas.raw)
cas.raw$pLN <- as.numeric(cas.raw$pLN)
```

Additional data cleaning and match of CAS and frame survey files

```
frame <- ddply(frame.raw, c("district","landing"), summarise, N=length(district))
names(frame) <- c("district","landing","N")
cas <- ddply(cas.raw[cas.raw$year %in% 2014,], c("district","landing"), summarise,
        n=length(district))
cas <- join(cas,frame)

## Joining by: district, landing

print("CAS 2014 landings sites with no match in 2012 FRAME survey")

## [1] "CAS 2014 landings sites with no match in 2012 FRAME survey"

cas[is.na(cas$N),]
```

```
##     district    landing   n  N
## 7     Buvuma     Bukaali  53 NA
## 11    Buvuma    Kiruguma  81 NA
## 12    Buvuma        Lufu  80 NA
## 14    Buvuma      Nyenda  80 NA
## 15    Buvuma    Wabuziba  55 NA
## 16    Buvuma       Ziiru  94 NA
## 17    Buvuma      Zzinga  64 NA
## 23 Kalangala  Kyagalanyi  73 NA
## 24 Kalangala      Mweena  83 NA
## 25 Kalangala Nabisuukiro  87 NA
## 29    Masaka     Nakigga  80 NA
## 30    Mayuge      Khaaza  99 NA
## 39    Mukono       Kiimi 108 NA
## 40    Mukono    Kinagaba  46 NA
## 45 Namayingo    Bumeru.A  52 NA
## 46 Namayingo  Butanira.B  71 NA
## 47 Namayingo      Golofa  78 NA
## 48 Namayingo        Hama  88 NA
## 50     Rakai Kasensero.A  80 NA
## 51     Rakai Kasensero.B  80 NA
## 56    Wakiso      Kitufu  68 NA
```

Further cleaning

```
i <- frame.raw$district %in% "Buvuma" & frame.raw$landing %in% "Kiriguma"
frame.raw$landing[i] <- "Kiruguma"
i <- frame.raw$district %in% "Buvuma" & frame.raw$landing %in% "Luufu"
frame.raw$landing[i] <- "Lufu"
i <- frame.raw$district %in% "Buvuma" & frame.raw$landing %in% "Malijja"
frame.raw$landing[i] <- "Wabuziba"

# In CAS we have only Ziiru - in FRAME we have three separate Ziiru (see below)
i <- frame.raw$district %in% "Buvuma" &
  frame.raw$landing %in% c("Ziiru Bushgayi","Ziiru Kibulwe","Ziiru Muto")
frame.raw$landing[i] <- "Ziiru"
i <- frame.raw$district %in% "Buvuma" & frame.raw$landing %in% "Zinga"
frame.raw$landing[i] <- "Zzinga"

i <- frame.raw$district %in% "Kalangala" & frame.raw$landing %in% "Luku/Nabusukira"
frame.raw$landing[i] <- "Nabisuukiro"
i <- frame.raw$district %in% "Kalangala" & frame.raw$landing %in% "Mwena"
frame.raw$landing[i] <- "Mweena"
i <- frame.raw$district %in% "Kalangala" & frame.raw$landing %in% 'Nakatiba'
frame.raw$landing[i] <- "Kyagalanyi"

i <- frame.raw$district %in% "Masaka" & frame.raw$landing %in% "Nakiga"
frame.raw$landing[i] <- "Nakigga"
i <- frame.raw$district %in% "Mayuge" & frame.raw$landing %in% "Kaaza"
frame.raw$landing[i] <- "Khaaza"

i <- frame.raw$district %in% "Mukono" & frame.raw$landing %in% "Kimmi"
frame.raw$landing[i] <- "Kiimi"
i <- frame.raw$district %in% "Mukono" & frame.raw$landing %in% "Kinaggaba"
frame.raw$landing[i] <- "Kinagaba"
```

```
i <- frame.raw$district %in% "Namayingo" & frame.raw$landing %in% 'Butanira "B"'
frame.raw$landing[i] <- "Butanira.B"
i <- frame.raw$district %in% "Namayingo" & str_sub(frame.raw$landing,1,6) %in% "Bumeru"
frame.raw$landing[i] <- "Bumeru.A"
i <- frame.raw$district %in% "Namayingo" & frame.raw$landing %in% "Gorofa"
frame.raw$landing[i] <- "Golofa"

## Here we join two landing sites in the CAS data, because there is no A & B split in the FRAME data
i <- cas.raw$district %in% "Rakai" & cas.raw$landing %in% c("Kasensero.A","Kasensero.B")
cas.raw$landing[i] <- "Kasensero"

i <- cas.raw$landing %in% "Hama"
cas.raw$landing[i] <- "Siamulala"

i <- frame.raw$district %in% "Wakiso" & frame.raw$landing %in% "Kituufu"
frame.raw$landing[i] <- "Kitufu"

# lets check where there is still a mismatch
frame <- ddply(frame.raw, c("district","landing"), summarise, N=length(district))
names(frame) <- c("district","landing","N")
cas <- ddply(cas.raw[cas.raw$year %in% 2014,], c("district","landing"), summarise,
         n=length(district))
cas <- join(cas,frame)

## Joining by: district, landing

print("Postcleaning: CAS 2014 landings sites with no match in 2012 FRAME survey")

## [1] "Postcleaning: CAS 2014 landings sites with no match in 2012 FRAME survey"

cas[is.na(cas$N),]
```

district	landing	n	N
Buvuma	Bukaali	53	NA
Buvuma	Nyenda	80	NA

Here, we considered the total number of boats sampled in CAS to be the total number of boats at the two landing site s in frame survey

Estimation of total landings

Reformatting the CAS data

The CAS data are stored as a wide table where each line corresponds to one boat sampled in a particluar landing site on a particular date. The recording of the catch (and price) of each species landed by a particlar boat is then stored in separate columns:

```
# Work only on one year
i <- cas.raw$year %in% 2014

# Only select columns of relavance
cn <- c("id","date","district","landing","vCode","gCode","wLN","wTL","wDA","wHA","wBD","wPA","wCG","w
OT")
cas_wide <- cas.raw[i,cn]
head(cas_wide)
```

To make the code for the calculation as simple as possible the wide table is converted into a long table, where each line corresponds to a boat landing for each species. The variable species is hence stored in a single column and the variable weight (and price) of landings are in two separate columns:

```
cas_long <- melt(cas_wide,c("id","date","district","landing","vCode","gCode"),
          variable.name = "sCode", value.name = "w")
head(cas_long)

     id        date      district       landing    vCode            gCode   sCode    w
1 47801  2014-05-27   Mayuge        Nakirimira    SP             SS         wLN     NA
2 47802  2014-05-27   Mayuge        Nakirimira    SP              SS        wLN     NA
3 47803  2014-05-27   Mayuge        Nakirimira    SF              BS         wLN    5.5
4 47804  2014-05-27   Mayuge        Nakirimira    SF             BS          wLN    21.0
5 47805  2014-05-27   Mayuge        Nakirimira    SF             BS         wLN     5.0
6 47806  2014-05-27   Mayuge        Nakirimira    SF             BS         wLN     5.0
```

Add the price as a column:

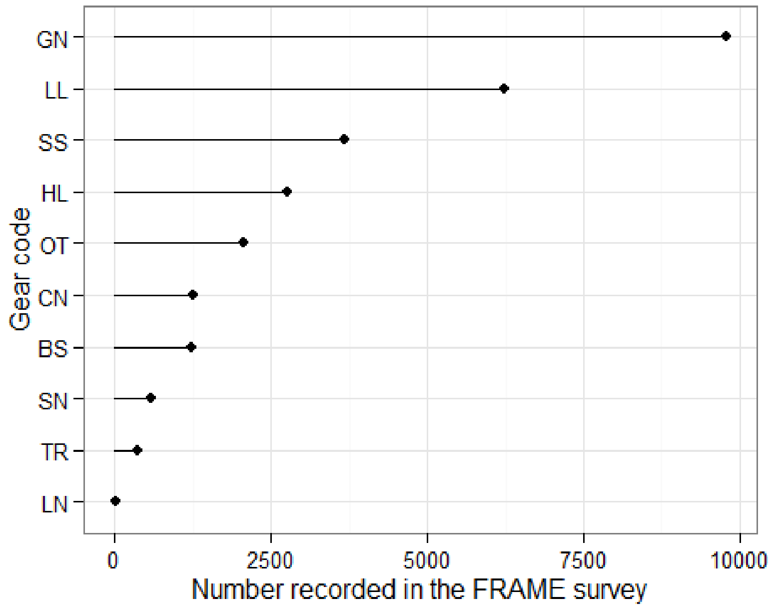
```
i <- cas.raw$year %in% 2014
cn <- c("id","date","district","landing","vCode","gCode","pLN","pTL","pDA","pHA","pBD", "pPA","pCG","pOT")
tmp <- cas.raw[i,cn]

tmp <- melt(tmp,c("id","date","district","landing","vCode","gCode"),
         variable.name = "sCode", value.name="p")

cas_long$p <- tmp$p

# Get rid of "w" in front of the species name
cas_long$sCode <- str_sub(cas_long$sCode,2,3)
head(cas_long)

        id       date       district        landing   vCode gCode sCode    w    p
1 47801 2014-05-27   Mayuge       Nakirimira    SP    SS       LN   NA   NA
2 47802 2014-05-27   Mayuge       Nakirimira    SP    SS       LN   NA   NA
3 47803 2014-05-27   Mayuge       Nakirimira    SF    BS       LN  5.5 3000
4 47804 2014-05-27   Mayuge       Nakirimira    SF    BS       LN 21.0 3000
5 47805 2014-05-27   Mayuge       Nakirimira    SF    BS        LN  5.0 3000
6 47806 2014-05-27   Mayuge       Nakirimira    SF    BS        LN  5.0 3000
```

Where the weight is "NA (not available" that means zero:

```
i <- is.na(cas_long$w)
cas_long$w[i] <- 0
```

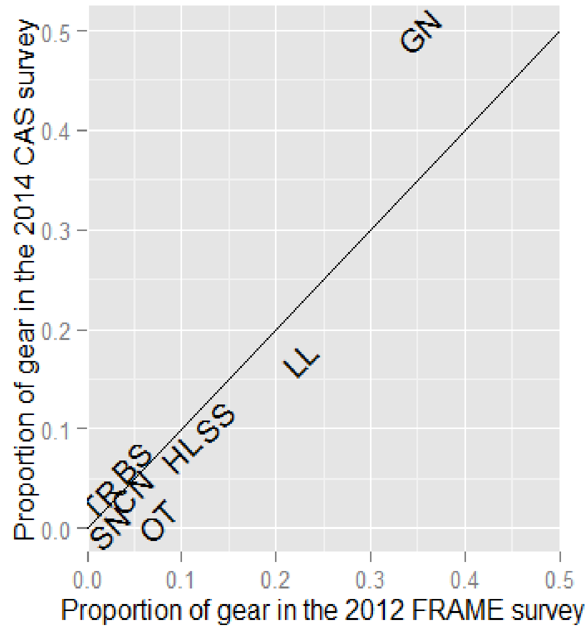
General observations

```
d <- ddply(frame.raw,c("gCode"),summarise,N=length(gCode))
d$P <- d$N/sum(d$N)

ggplot(d,aes(reorder(gCode,N),N)) + theme_bw() + geom_point() + geom_linerange(aes(ymin=0,ymax=N)) +
  labs(x="Gear code",y="Number recorded in the FRAME survey") + coord_flip()
```



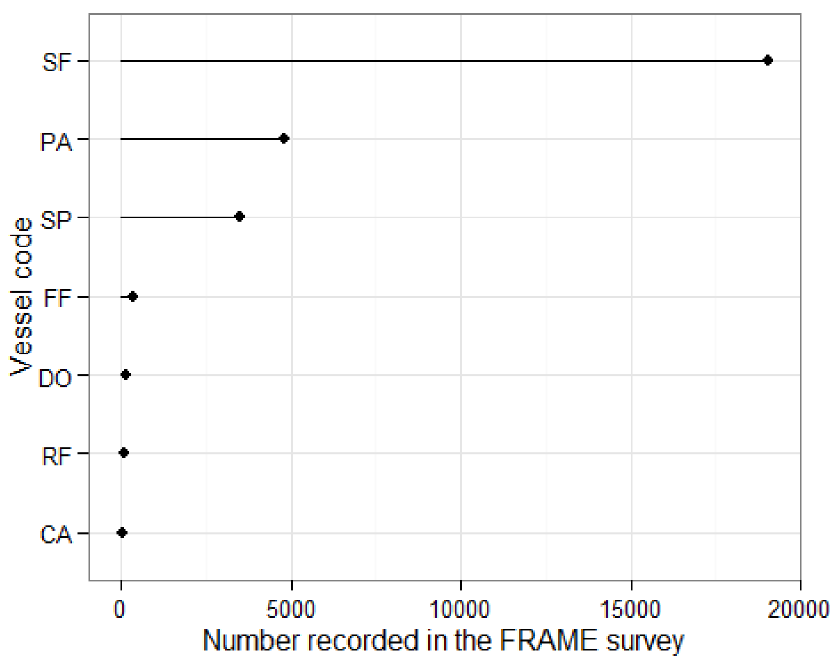
The figure shows the most prevalent gear in the 2012 FRAME survey, with GN, LL, SS and HL being the four most common ones. Of note is that in the FRAME survey we have substantial recording (7%) of gear as "OT" meaning other gear.

```
d2 <- ddply(cas_long,c("gCode"),summarise,n=length(gCode))
d2$p <- d2$n/sum(d2$n)


x <- join(d,d2,type="full")

## Joining by: gCode

x <- x[order(-x$P),]
ggplot(x,aes(P,p,label=gCode)) + geom_text(angle=45) + geom_abline(intercept=0,slope=1) + labs(x="Proportion of gear in the 2012 FRAME survey", y="Proportion of gear in the 2014 CAS survey") +
  coord_equal(xlim=(c(0,max(x$p,na.rm=TRUE))))
```



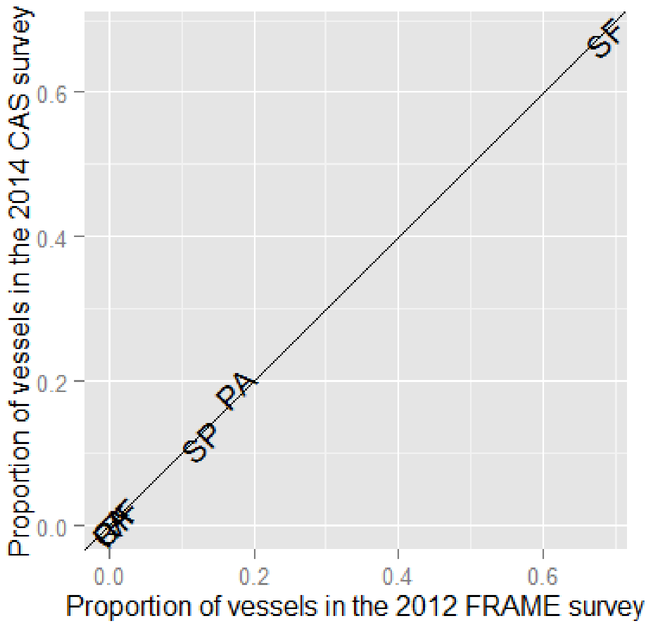
This figure shows that the percentage of GN in the 2014 CAS survey is much higher than in the 2012 FRAME survey while a lot of other important gear show lower percentage in the CAS survey. Note that gear "OT" is very low in the CAS survey.

```
d <- ddply(frame.raw,c("vCode"),summarise,N=length(vCode))
d$P <- d$N/sum(d$N)
ggplot(d,aes(reorder(vCode,N),N)) + theme_bw() + geom_point() + geom_linerange(aes(ymin=0,ymax=N)) +
  labs(x="Vessel code",y="Number recorded in the FRAME survey") + coord_flip()
```



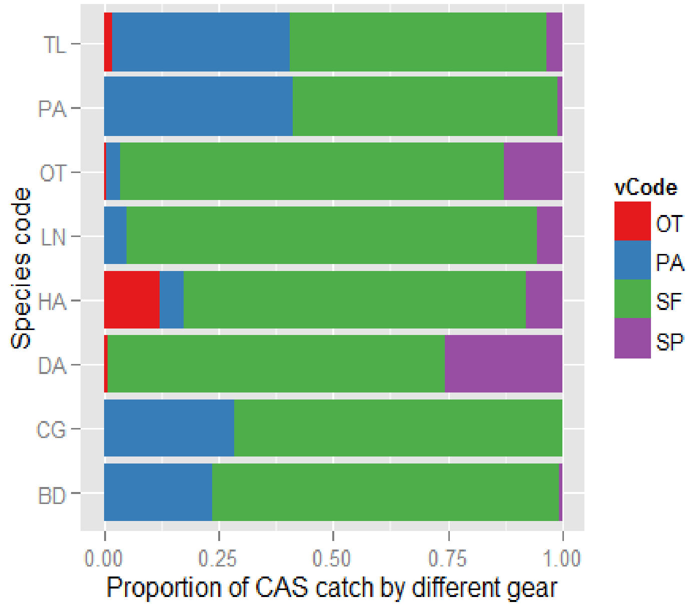
This figures shows that there three main vessel types, SF, PA and SP in the FRAME survey.

```
d2 <- ddply(cas_long,c("vCode"),summarise,n=length(vCode))
d2$p <- d2$n/sum(d2$n)

x <- join(d,d2,type = "full")

## Joining by: vCode

x <- x[order(-x$P),]
ggplot(x,aes(P,p,label=vCode)) + geom_text(angle=45) + geom_abline(intercept=0,slope=1) +
  labs(x="Proportion of vessels in the 2012 FRAME survey", y="Proportion of vessels in the 2014 CAS survey") +
  coord_equal()
```



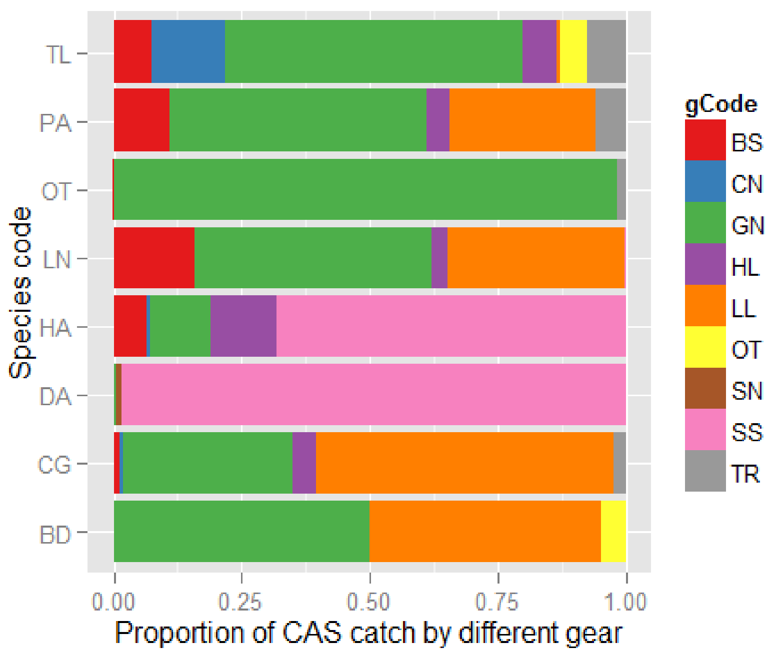
The figure shows the proportion of vessel categories sampled in the CAS survey relative to the proportion in the FRAME survey.

```
cas <- cas_long
i <- cas$vCode %in% c("SF","PA","SP")
cas$vCode[!i] <- "OT"

d <- ddply(cas,c("sCode","vCode"),summarise,w=sum(w))
d <- ddply(d,c("sCode"),transform,p=w/sum(w))
ggplot(d,aes(sCode,weight=w,fill=vCode)) + geom_bar(position="fill") + scale_fill_brewer(palette="Set1") +
  labs(x="Species code",y="Proportion of CAS catch by different gear") + coord_flip()
```



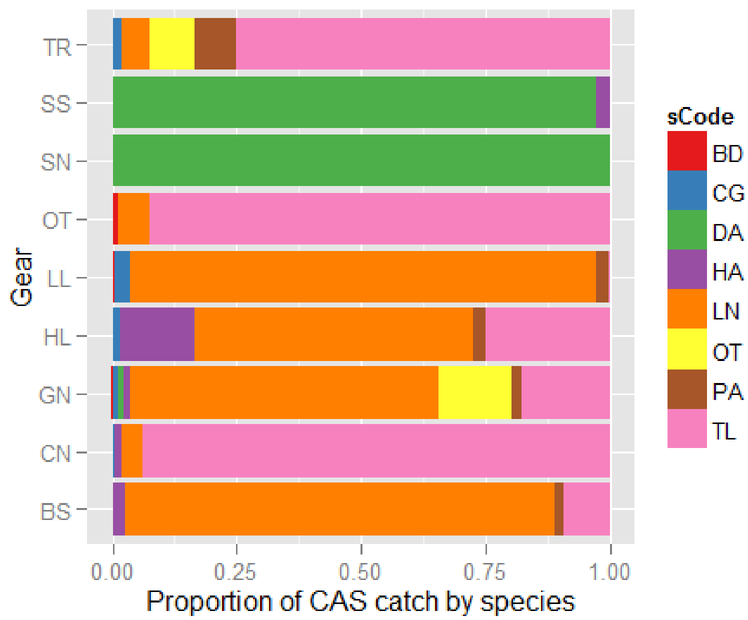
This figure shows that vessels other than PA, SF and SP are only contributing to the catches of HA (about 12% of the catches).

```
# What is the prevalent gear used to catch species
d <- ddply(cas,c("sCode","gCode"),summarise,w=sum(w))
d <- ddply(d,c("sCode"),transform,p=w/sum(w))
ggplot(d,aes(sCode,weight=w,fill=gCode)) + geom_bar(position="fill") + scale_fill_brewer(palette="Set1") +
  labs(x="Species code",y="Proportion of CAS catch by different gear") + coord_flip()
```



Here we see that the primary gear to catch dagaa (DA) and haplochromines (HA) is small seine (SS). The DA fisheries can be really thought of as singe gear fishery. Gill nets are important for *Protopterus aethiopicus* (PA), Nile perch (LN) and tilapias (TL). Cast nets (CN) is of relatively little importance except for the TL.

```
ggplot(d,aes(gCode,weight=w,fill=sCode)) + geom_bar(position="fill") + scale_fill_brewer(palette="Set1") +
  labs(x="Gear",y="Proportion of CAS catch by species") + coord_flip()
```



Here we see the proportion of species cought in different gears. We see that SS and SN (although only 4 samples in the cas survey) are primarily catching DA (not surprisingly). LL, BS and CN can also be described as single species fishery, the former two largly catching LN and CN catching primarily TL.

Approaches to estimate catch landings on Lake Victoria (Uganda data only)

A. Whole lake ("top-bottom") approach

In the "whole lake" approach the sum of landings in weigth and price by vessel type, gear type and species of all the boats in the CAS survey are calculated as:

```
group_variables <- c("vCode","gCode","sCode")
cas <- ddply(cas_long,group_variables,summarise,
          w_sampled=sum(w),
          p_sampled=sum(p, na.rm=TRUE))
```

Calculate the number of trips per vessle-gear code combination that are behind the above sample sums and the mean number of days fished in the week:

```
group_variables2 <- c("vCode","gCode")
i <- cas.raw$year %in% 2014
trip <- ddply(cas.raw[i,],group_variables2,summarise,
           n=length(district),
           nd_perWeek=mean(nDaysWeek,na.rm=TRUE))
```

Join the species and the trip information for each vessel gear code combination:

```
cas <- join(cas,trip)
```

```
## Joining by: vCode, gCode
```

Join the frame survey with the above data to get the total number of boats per district-vessel-gear combination. The first step is to calculate the number of vessel-gear combinations in the frame survey:

```
frame <- ddply(frame.raw,group_variables2,summarise,
               N=length(vCode))
cas <- join(cas,frame)

## Joining by: vCode, gCode
```

Where vessel gear combination is not available in the frame survey assume that the number in the cas survey is the total (this is not an issue in the "whole lake" approach):

```
i <- is.na(cas$N)
cas$N[i] <- cas$n[i]
```

Then raise the sum of the sampled catch to annual catch per species by vessel-gear combination by:

```
cas$landings <- cas$w_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
cas$price <- cas$p_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
```

Calculate the annual of each species:

```
landings2014_by_lake <- ddply(cas,c("sCode"),summarise,
                              landings=round(sum(landings),3),
                              price=round(sum(price),3))
landings2014 <- landings2014_by_lake
landings2014$Method <- "Lake wise"
```

Now we need to split the data to districts

Number of vessels per district

```
frame <- ddply(frame.raw,c("district"),summarise,N=length(district))
frame$P <- frame$N/sum(frame$N)
x <- expand.grid(district=sort(unique(frame.raw$district)),
               sCode=landings2014_by_lake$sCode)
x <- join(x,frame)

## Joining by: district

x <-join(x[,c("district","sCode","P")],landings2014_by_lake)

## Joining by: sCode

x$landings <- x$landings * x$P
x$price <- x$price * x$P
x$Method <- "lake"
landings2014_by_district_from_lake <- x[,c("district","Method","sCode","landings","price")]
```

B. District ("bottom-top") approach

To calculate the landings by district needs only a minor modification of the codes above.

```
group_variables  <- c("district","vCode","gCode","sCode")
group_variables2 <- c("district","vCode","gCode")
```

The remainder is then the same:

```
cas <- ddply(cas_long,group_variables,summarise,
             w_sampled=sum(w),
             p_sampled=sum(p, na.rm=TRUE))
i <- cas.raw$year %in% 2014
trip <- ddply(cas.raw[i,],group_variables2,summarise,
              n=length(district),
              nd_perWeek=mean(nDaysWeek,na.rm=TRUE))
cas <- join(cas,trip)

## Joining by: district, vCode, gCode

frame <- ddply(frame.raw,group_variables2,summarise,
               N=length(vCode))
cas <- join(cas,frame)

## Joining by: district, vCode, gCode

i <- is.na(cas$N)
cas$N[i] <- cas$n[i]
i <- cas$n > cas$N
cas$N[i] <- cas$n[i]
cas$landings <- cas$w_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
cas$price <- cas$p_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
```

To calculate the landings by species and district on then does:

```
landings2014_by_district <- ddply(cas,c("district","sCode"),summarise,
                                  landings=round(sum(landings),3),
                                  price=round(sum(price),3))
```

And to calculate the total landings based on landings by district one does:

```
landings2014_from_district <- ddply(landings2014_by_district,c("sCode"),summarise,
                                  landings=round(sum(landings),3),
                                  price=round(sum(price),3))
landings2014_from_district$Method <- "by_district"
landings2014 <- rbind(landings2014,landings2014_from_district)
```

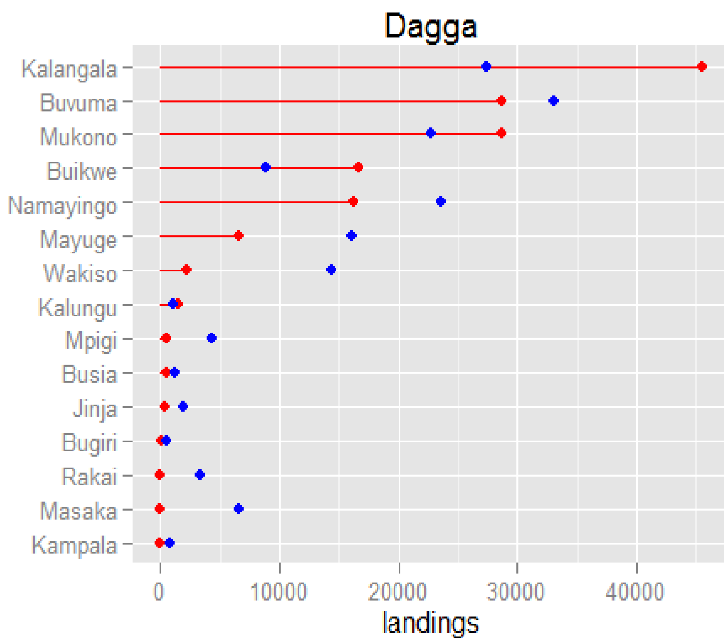
Merge the by distict calculation with that obained from the whole lake method

```
landings2014_by_district$Method <- "district"
landings2014_by_district <- rbind(landings2014_by_district[,names(landings2014_by_district_from_lake)],
                                    landings2014_by_district_from_lake)

x <- landings2014_by_district
j <- x$Method %in% "district"
```

Looking at the three major fisheries of dagaa, Nile perch and tilapias

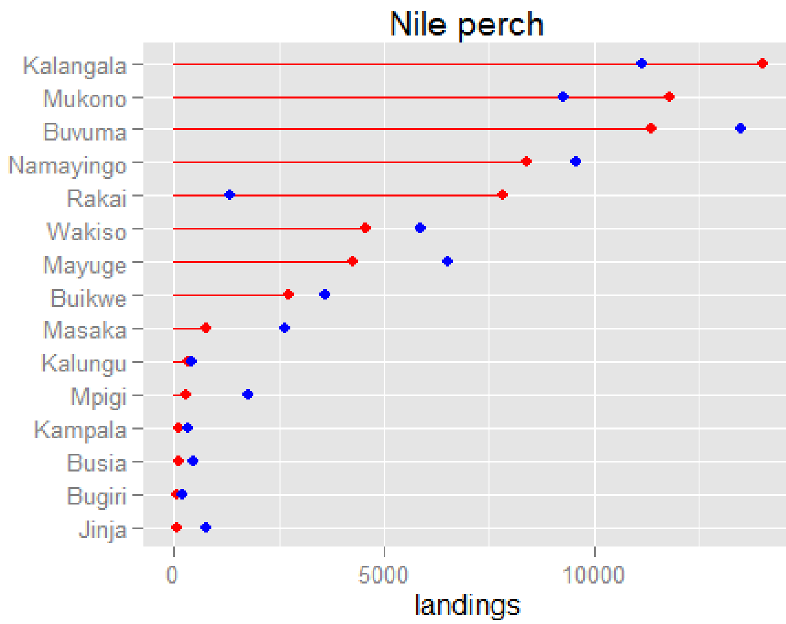
Dagaa

```
i <- x$sCode %in% "DA"
ggplot() +
  geom_point(data=x[i & j,],aes(reorder(district,landings),landings),col="red") +
  geom_linerange(data=x[i & j,],aes(reorder(district,landings),ymin=0,ymax=landings),col="red") +
  geom_point(data=x[i & !j,],aes(reorder(district,landings),landings),col="blue") +
  coord_flip() +
  labs(x="",title="Dagga")
```



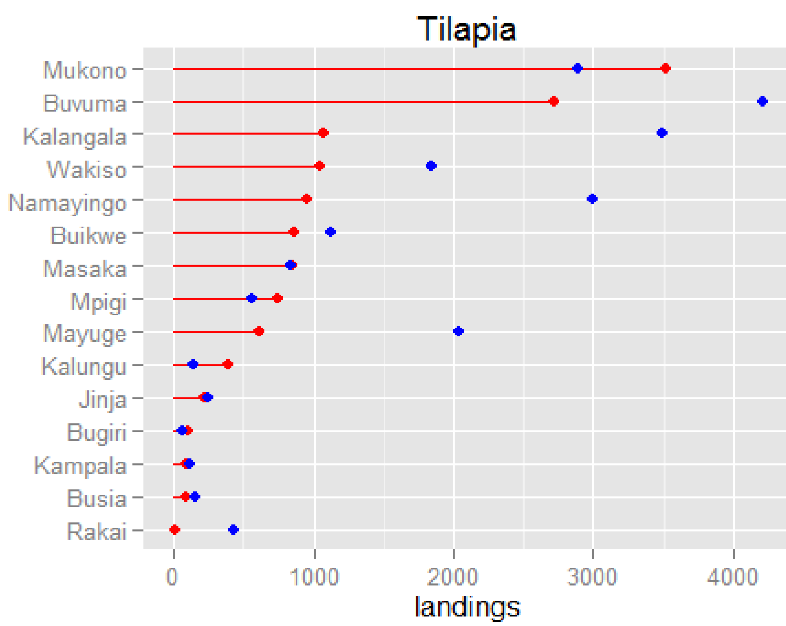
Nile perch

```
i <- x$sCode %in% "LN"
ggplot() +
  geom_point(data=x[i & j,],aes(reorder(district,landings),landings),col="red") +
  geom_linerange(data=x[i & j,],aes(reorder(district,landings),ymin=0,ymax=landings),col="red") +
  geom_point(data=x[i & !j,],aes(reorder(district,landings),landings),col="blue") +
  coord_flip() +
  labs(x="",title="Nile perch")
```



Tilapias

```
i <- x$sCode %in% "TL"
ggplot() +
  geom_point(data=x[i & j,],aes(reorder(district,landings),landings),col="red") +
  geom_linerange(data=x[i & j,],aes(reorder(district,landings),ymin=0,ymax=landings),col="red") +
  geom_point(data=x[i & !j,],aes(reorder(district,landings),landings),col="blue") +
  coord_flip() +
  labs(x="",title="Tilapia")
```



Variability in CAS estimates

2005 CAS surveys

In the year 2005 survey the CAS survey was done four times over the year while in the last two CAS surveys (2011 and 2014) the estimates are base on only one CAS survey done in the month of May. The 2005 data allows one to investigate the potential problems associated with limiting the CAS survey to only one month of the year.

```
i <- cas.raw$year %in% 2005 & cas.raw$month %in% c(7,8,9,11)
cn <- c("id","month","district","landing","vCode","gCode","wLN","wTL","wDA","wHA","wBD","wPA","wCG","wOT")
cas_wide <- cas.raw[i,cn]
cas_long <- melt(cas_wide,c("id","month","district","landing","vCode","gCode"),
            variable.name = "sCode",
            value.name = "w")

i <- cas.raw$year %in% 2005 & cas.raw$month %in% c(7,8,9,11)
cn <- c("id","month","district","landing","vCode","gCode","pLN","pTL","pDA","pHA","pBD", "pPA","pCG","pOT")
tmp <- cas.raw[i,cn]

tmp <- melt(tmp,c("id","month","district","landing","vCode","gCode"),
           variable.name = "sCode",
           value.name="p")

cas_long$p <- tmp$p
# get rid of "w" in front of the species name
cas_long$sCode <- str_sub(cas_long$sCode,2,3)

i <- is.na(cas_long$w)
cas_long$w[i] <- 0
#i <- is.na(cas2$p)
#cas_long$p[i] <- 0

group_variables  <- c("district","month","vCode","gCode","sCode")
group_variables2 <- c("district","month","vCode","gCode")

cas <- ddply(cas_long,group_variables,summarise,
           w_sampled=sum(w),
           p_sampled=sum(p, na.rm=TRUE))
i <- cas.raw$year %in% 2005 & cas.raw$month %in% c(7,8,9,11)
trip <- ddply(cas.raw[i,],group_variables2,summarise,
            n=length(district),
            nd_perWeek=mean(nDaysWeek,na.rm=TRUE))
cas <- join(cas,trip)

## Joining by: district, month, vCode, gCode

frame <- ddply(frame.raw,c("district","vCode","gCode"),summarise,
             N=length(vCode))
cas <- join(cas,frame)

## Joining by: district, vCode, gCode

i <- is.na(cas$N)
cas$N[i] <- cas$n[i]
i <- cas$n > cas$N
```

```
cas$N[i] <- cas$n[i]
cas$landings <- cas$w_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
cas$price <- cas$p_sampled *
  cas$nd_perWeek/7 * 31 *
  cas$N/cas$n * 12 / 1e3
```

To calculate the landings by species, month and district on then does:

```
landings2005_by_district <- ddply(cas,c("district","month","sCode"),summarise,
                    landings=round(sum(landings),3),
                    price=round(sum(price),3))
```

And to calculate the total landings based on landings by district and month one does:

```
landings2005_from_district <- ddply(landings2005_by_district,c("month","sCode"),summarise,
                    landings=round(sum(landings,na.rm=T),3),
                    price=round(sum(price,na.rm=T),3))
ggplot(landings2005_from_district,aes(sCode,weight=landings,fill=factor(month))) + geom_bar(position="dodge"
) + labs(x="Species code", y="Catch landings (t)")
```