FISHERIES TRAINING PROGRAMME
P.O. Box 1390, Skulagata 4
120 Reykjavik, Iceland

Final Project 2007

A DATA SAMPLING STRATEGY FOR COASTAL HERRING OF SRI LANKA

Sisira Haputhantri
Marine Biological Resources Division (MBRD)
National Aquatic Resources Research and Development Agency (NARA)
Colombo 15
Sri Lanka
sisirahaputhantri@yahoo.com

Supervisors
Gudmundur Thordarson
Marine Research Institute (MRI)
gudthor@hafro.is
and
Gunnar Stefansson
University of Iceland
gstefans@gmail.com

ABSTRACT

Herring (*Amblygaster sirm*) is the most valuable species found in the small scale coastal gillnet fishery of Sri Lanka. The species has a short life span of little over a year and reaches around 20 cm in length. Several studies have been conducted on this species in order to aid sustainable management. Recently conducted studies have revealed that the present level of exploitation of this species in Sri Lanka is unsustainable. Conducting stock assessments using length frequency data and length-based stock assessment models is useful for managing fisheries resources. At present there is no sufficient length frequency data available from present data collection programmes. This was found to be the major obstacle faced when attempting to conduct length-based stock assessment on coastal herring of Sri Lanka. Therefore, attention was paid by this study to reviewing the status of herring and also to proposing a better sampling strategy to obtain length frequency data for future stock assessments. The non-parametric bootstrap technique was used to determine the minimum sample size of length measurements and number of samples required to conduct stock assessment on herring. Capelin (*Mallotus villosus*) length-frequency survey data obtained from annual capelin surveys in Icelandic waters was used as the proxy population for this analysis. The issue of possible internal correlations in the length measurements in a sample (i.e. only getting small fish in a sample or only large ones) was also addressed while applying the bootstrap method for capelin data. The results from the bootstrapping indicate that around 60 samples, each with around 40 measurements, are needed per month to follow the rapid growth of *A. sirm* over the year and conduct assessment. The minimum sample size and number of samples are determined by resource expenditure for taking length measurements and the degree of precision of the estimates made from collected length measurements. Therefore, 20 - sample size and 40 – number of samples might perhaps be adequate when cost for data collection is taken into account and this might be the optimal trade-off between cost and parameter precision.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Sri Lanka is an island in the Indian Ocean and is situated in the Bay of Bengal, south of India between 6° – 10°N and 80° – 82°E. The island has a coastline of 1,770 km and a 2,800 km^2 narrow continental shelf. Since the declaration of an Exclusive Economic Zone (EEZ) in 1978, Sri Lanka has sovereign rights over 517,400 km^2 of the ocean (MFAR 2007). The sea area from the shore to the edge of the continental shelf is considered as the coastal area.

The contribution of the fisheries sector to the Gross Domestic Product (GDP) has varied between 1.9 and 2.5% over the four year period 2001-2004 (MFAR 2007). It declined to 1% in 2005 due to the tsunami on 26 December 2004. The sector is, however, important to the economy in terms of employment, foreign exchange earnings and protein supply to the nation (Haputhantri 2004). The sector currently provides direct employment to about 650,000 people (MFAR 2007). In the last two decades, export of marine and aquatic products has developed into one of the fastest growing export sub-sectors, bringing in a substantial amount of foreign exchange earnings. The sector contributes around 70% of the animal protein consumed in the country (MFAR 2007).

The fisheries sector of Sri Lanka was severely affected by the tsunami. Heavy damages were inflicted on fishing crafts operated in shallow coastal waters. A substantial number of coastal fishing crafts were either damaged or destroyed and some disappeared.

Total fish production of Sri Lanka in 2006 was estimated at 251,270 tons (MFAR 2007). Coastal fisheries constituted 48% of the fisheries production, while offshore fisheries contributed with 38% and inland fisheries and aquaculture 14%. Traditionally, coastal fisheries have represented up to 89% of the annual fisheries production (MFAR 2007). This decline in the relative contribution of the coastal sector can be mainly attributed to the rapid development in the offshore fisheries and poor management of coastal fisheries (Haputhantri 2004).

The coastal fisheries of Sri Lanka are multispecies and multigear with high species diversity. Coastal fisheries resources mainly consist of small and large pelagic fish, demersal and coral reef fish, invertebrates, shrimps and crabs (Samaranayake 2003). Most species are subject to increasing fishing pressure by different vessel-gear combinations (Haputhantri 2004). The total number of crafts in the major coastal craft type (i.e. fibber reinforced plastic boats) in 2006 has increased by 54% compared to pre-tsunami levels (MFAR 2007). A number of various donor agencies and NGOs voluntarily got involved in the repair of crafts damaged by the tsunami and provided new fishing crafts. This happened without proper coordination with the agencies responsible. Another reason for this apparent increase is underestimates of operating fishing crafts in previous years but the 2006/2007 census was more thorough than previous censuses.

Small pelagics account for about 40% of the coastal fish production (Haputhantri 2004). There are about 100 species of small pelagics around Sri Lanka, of which not more than 25 contribute significantly to the commercial production (Samaranayake 2003). Small pelagic fish production in 2006 was 56,230 tons (MFAR 2007).

At present, coastal fishery resources of Sri Lanka are not managed on a scientific basis as is the case for most other marine fisheries in tropical developing countries. The lack of communication between scientists, administrators, decision-makers and the fishing communities is one of the major obstacles to the formulation of effective management strategies (Haputhantri 2004). In addition, lack of assessment studies, especially on major coastal fish stocks which are currently subjected to high fishing pressure, was found to be another obstacle to the implementation of a sound management policy on the utilisation of coastal resources. A proper data collection strategy in order to monitor coastal fish landings is required for conducting stock assessments on coastal species. Therefore, it is important to undertake research to address the above-mentioned issues in view of formulating an appropriate management strategy for the long-term sustainable utilisation of coastal fishery resources in Sri Lanka.

The proposed study was undertaken to review the status of the herring stock (*Amblygaster sirm,* the major small pelagic stock) in the coastal waters of Sri Lanka using commercial fish landing data and other available information and also to propose a sampling strategy for the west coast to obtain data for future assessments. Findings of the study could be used to formulate recommendations and strategies for the long term sustainable utilisation of this fishery. *A. sirm* is one of the most valuable species in the coastal fishery of Sri Lanka and it has been revealed that the present level of exploitation of the species is not sustainable (Haputhantri 2004, Haputhantri *et al.* 2008).

2 AN OVERVIEW OF THE HERRING FISHERY IN SRI LANKA OPERATED BY SMALL MESHED GILLNETS

The narrow continental shelf which rarely exceeds 40 km and averages 22 km in width is the most productive fishing area for coastal species. All small pelagic fish species including herring are most abundant in this zone. Most small pelagic fish are depth dependant. Sardine, anchovy and herring like species are mostly concentrated in the area of the continental shelf within the first 50 m in depth (Sivasubramaniam 1999).

2.1 Herring in the coastal waters of Sri Lanka

A. sirm (Clupeidae) is a small pelagic schooling species occurring in coastal waters and lagoons, at depth ranges of 10-75 m (Froese and Pauly 2007). It is abundant in the southern, western and the northwestern coastal waters of Sri Lanka and is frequently found in the catches of gillnets at depths varying from 5-70 m (Haputhantri 2006). This species is dominant in the fish catches of the western coast (Karunasinghe and Wijeyaratne 1991b). *A. sirm* is a commercially important species as a fish taken for food and is also used as bait in the tuna fishery.

2.2 Fishing vessels and gears

Three types of coastal vessels with small mesh gillnets target *A. sirm* in the shallow coastal waters of Sri Lanka. They are fiber reinforced plastic boats (FRP), mechanised traditional crafts (MTC) and non-mechanised traditional crafts (NTC). FRP boats are 5 - 7.4 m in length and fitted with 8-40 hp outboard engines. Motorised traditional crafts are fixed with outboard engines of 8-10 hp. All traditional crafts, even the crafts without engines are 5 to 12 m in length.

Small mesh gillnets widely used for catching small pelagic species are the main fishing gear used for fishing *A. sirm*. *A. sirm* is mostly caught by gillnets with a mesh size range of 22-38 mm (Haputhantri 2004). A single net piece with mesh sizes between 22 –38 mm is normally 3000 meshes long and 300-500 meshes deep. The relative contribution of *A. sirm* to the small mesh gillnet catch is over 50% and therefore, this species can be considered as the target species in the small mesh gillnet fishery (Haputhantri 2004). Around 90% of small pelagics are caught by small mesh gillnets (Karunasinghe *et al.* 2000). The rest, 10%, is mainly caught by beach seines.

2.3 Small mesh gillnet fishing operations

The set time for gillnets is normally 0.5 - 4 hours and this is relatively longer for FRP boats than traditional boats (Haputhantri 2004). In general, only one fishing operation takes place each day and involves two persons per boat. Fishing of small pelagic species using gillnets is conducted throughout the year but most coastal boats do not operate in unfavourable weather conditions. The number of gillnets used per fishing operation varies from 3 to 28 but in general mechanised crafts use more gillnets than non-mechanised crafts (Haputhantri 2004). Variations among fishing areas have been observed. For example, 98% of FRP boats in Negombo fisheries district use more than 10 gillnets per fishing operation (Haputhantri 2004).

2.4 Previous research on herring fisheries in Sri Lanka

Dayaratne and Gjϕsaeter (1986) studied age, growth and spawning season of *A. sirm* collected by gillnets in the coastal waters of Sri Lanka. This species reaches 20 cm after one year and the total life span is 1 - 1.5 years. This is the only age based study on this species found from Sri Lanka.

Sizes at 50% maturity for males and females are 15.9 cm and 15.0 cm respectively (Karunasinghe and Wijeyaratne 1998). Eggs and larval stages have been reported from coastal waters off Sri Lanka. There is some correlation between spawning of *A. sirm* and the two seasonal monsoons: the southwest monsoon which begins in April and continues until August and the northwest monsoon which lasts from November to February (Dayaratne and Gjϕsaeter 1986). The main spawning season of *A. sirm* at the west coast is observed at the beginning of the southwest monsoon, i.e. from April to June and they further suggest that there is another spawning season for *A. sirm* at the beginning of the northeast monsoon, i.e. in November-December at the east coast.

The population parameters of *A. sirm* in the coastal waters of Sri Lanka have been estimated by various authors and show substantial variability: L_{∞} = 24.6 - 25.8, K =

0.95 – 1.48 and M = 1.3 - 2.2 (Table 1). Haputhantri (2006) evaluated the parameter uncertainties in those estimations and concluded that uncertainties may result in misleading management recommendations if those estimations are used to determine the optimal fishing strategies.

Table 1: Literature values of population parameters for *A. sirm*

L_∞	K	M	Source
24.8	0.95	2.1	Siddeek *et al.* (1985)
24.9 - 25.8	1.10 – 1.48	1.3	Karunasinghe and Wijeyaratne (1991b)
24.6	1.3	2.2	Dayaratne and Sivakumaran (1994)

Gear selection patterns for *A. sirm* in the small-mesh gillnet fishery in the west coast of Sri Lanka was studied by Dayaratne (1988) and Karunasinghe and Wijeyaratne (1991a). For this, a set of selection curves for gillnets with different mesh sizes were first plotted by overlapping normal probability curves with the same standard deviation. Selection factors and optimum selection lengths were then estimated from the method proposed by Holt (1963) for determining gear selectivity. Estimated values for selection factors ranged from 5.49 – 5.71 and 5.11 - 6.03 whereas optimum selection lengths varied from 13.91 – 21.67 cm and 12.9 – 19.7 cm between the two studies respectively. Haputhantri (2004) studied gillnet catch efficiency of *A. sirm* in the coastal waters of Sri Lanka using commercial catch/effort data and other variables related to gillnet operations. A tendency for increased catch rates of *A. sirm* with increased sea depth up to 32 m was observed. A seasonal and inter-annual variation in the catch rates was also observed for *A. sirm*.

The role of herring species in the coastal ecosystem of Sri Lanka was studied by analysing the food web of the ecosystem (Haputhantri 2004 and Haputhantri *et al.* 2008). Thirty-nine functional groups, including herring as one group, were considered for this ecosystem based on the modelling approach. In simulation modelling of fishing effort, a possibility for greatly affecting major small pelagic populations such as herrings and sardines in terms of their biomass and available catch by increased fishing effort of small mesh gillnets was further observed.

A little attention has yet been drawn to length-based stock assessments. Sanders and Dayaratne (1999), however, conducted a stock assessment using a length-based Thompson and Bell model and life history data for herring. Data availability for conducting length-based stock assessments might be one possible reason for the lack of attention towards such assessment methods.

2.5 Fisheries statistics

The National Aquatic Resources Research and Development Agency (NARA) which is the research arm of Ministry of Fisheries and Aquatic Resources (MFAR) of Sri Lanka has been collecting fisheries statistics for many years and fishing operations have been regularly monitored under its annual research programmes for a long period of time. Time series fisheries data on herring used for this study were provided by the small pelagic fishery landings monitoring programme conducted by NARA.

The data collection programme for small pelagics is at present carried out in eight fisheries districts (Puttalam, Chilaw, Negombo, Kalutara, Matara, Tangalle, Kalmunai and Trincomalee) out of the total fifteen marine fisheries districts (Figure 1). Trained enumerators collect data under supervision of the research officers of NARA by monitoring the landings from three types of coastal vessels described above. This data collection programme was started in January 1996. During the first four years (i.e. 1996 – 1999) the statistics collected were restricted to the collection of species level information of fish landings including herrings (i.e. species wise catch of the landings) and gear/vessel employed for respective fishing operations. During that period, the total number of monitored fisheries districts was five (Negombo, Kalutara, Matara, Tangalle and Kalmuna) and collected data were stored in Excel spreadsheets.

In 2000, the NARA small pelagic fisheries monitoring programme was widened in order to collect other information such as details on fishing operations: fishing time, depth of fishing, number of fishing devices used and their sizes (for example, the number of net pieces used in the gillnet fishery and their mesh sizes). Monitoring fishing effort and measuring the lengths of commercially important species also started then. Apart from that, the number of fisheries districts covered by the small pelagic fishery data collection programme was extended up to seven fisheries districts to include Chilaw and Trincomalee. In 2006, Puttalam fisheries district in the northwest coast was included in the monitoring programme.

For data collection, 12 samplers are stationed in the above districts by NARA except in Puttalam and Chilaw. In these two fisheries districts, data are collected by research assistants attached to MBRD. Important fish landing sites of the above-mentioned fisheries districts have been selected for data collection. The whole sampling scheme was designed based on stratified random sampling. Sampling stratification mainly consists of spatial strata (landing sites), technical strata (vessel/gear types) and temporal strata (months). Weekly field visits are made to the landing sites for data collection. On each sampling day, at least 10% of the total number of fishing crafts targeting small pelagic species is sampled randomly at each sampling site.

Total lengths of the commercially important species selected on a random basis are also measured to the closest millimetre while recording the landings of the catches. However, only limited length measurements are taken per day. It is not clear how many length measurements should be taken for each species. Taking a limited number of length measurements might be problematic for length-based stock assessments.

In the field, Clupeids and other dominant species such as a few Carangids, *Sphyraena spp* and many *Stolephorus spp* are identified to the species level. *Gazza spp*, *Leiognathus spp* and *Thryssa spp* etc. are identified to the genus level. Most Carangids, flying fish (Exocoetidae) and other species (Belonidae, Mullidae etc.) are grouped into their respective families. Over seventy-five such species/genus/families groups are considered in the small pelagic fishery monitoring programme of NARA. Since 2000, data have been stored in an Access database.

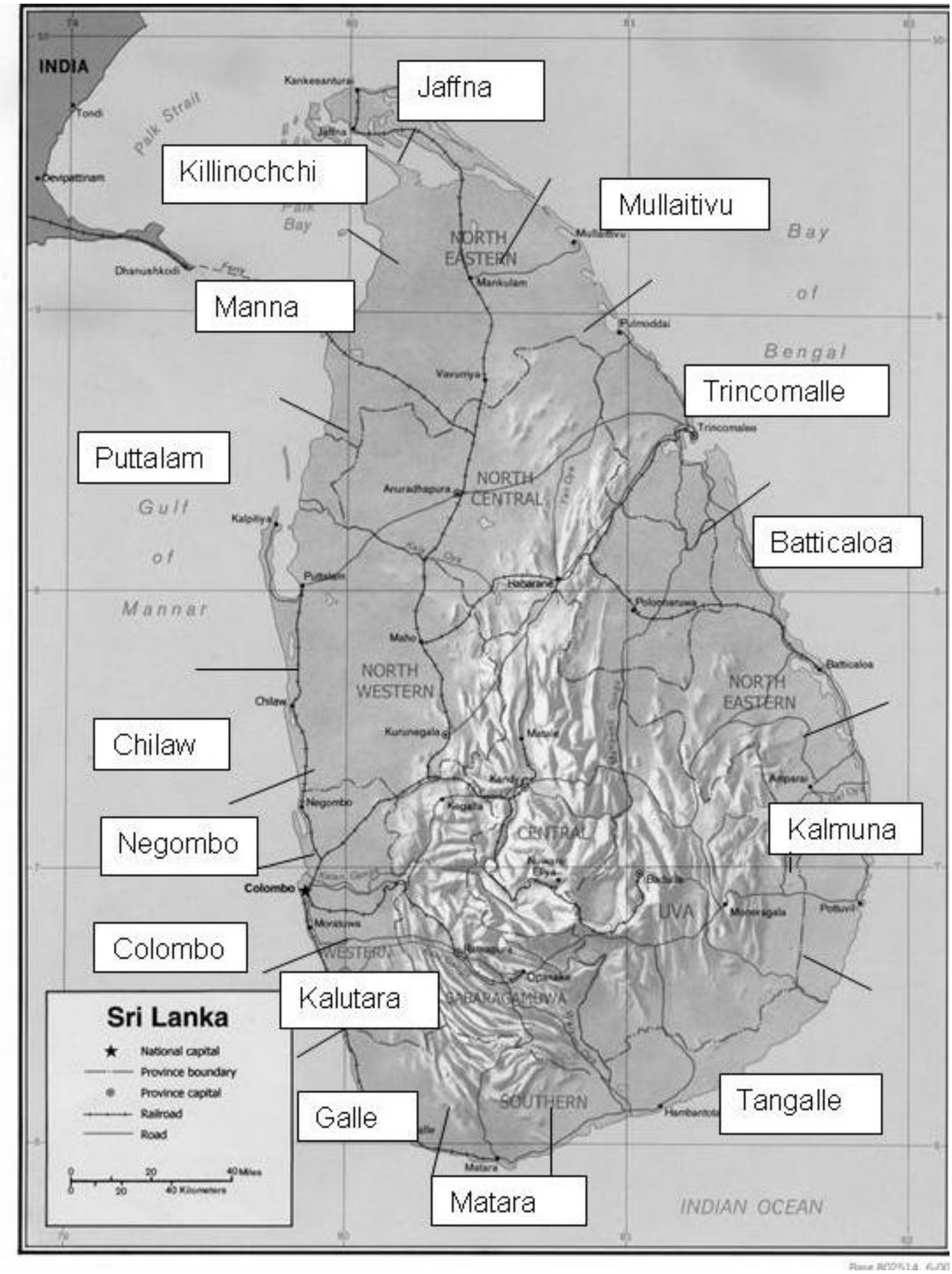


Figure 1: Marine fisheries districts in Sri Lanka (Haputhantri 2004)

2.6 West coast herring fishery

Many studies have been conducted on the west coast herring fishery in Sri Lanka (Dayaratne 1984, Siddeek *et al.* 1985, Karunasinghe 1990, Haputhantri 2004). The west coast herring fishery has been monitored for many years, in particular by NARA. Negombo and Chilaw are the most important fisheries districts on the west coast for small pelagics (Figure 1). These two districts reported landings of 6,130 tons and 8 850 tons of small pelagic fish in 2006 and these are 11% and 16% of the total small pelagic fish landings of the country respectively (MFAR 2007).

The following section is based on the information from 1996 to 2006 in Negombo fisheries district provided by the NARA small pelagic fishery landings monitoring programme. According to the census of fishing vessels in 2006/2007, there were 3904 coastal fishing crafts operating in this district: FRP – 1845, MTC – 1 and NTC – 2058 (MFAR 2007). The NTC crafts do not target herrings because these small boats mostly operate in shallow coastal areas and lagoons and are not seaworthy enough to access the fishing grounds were herrings are abundant. Consequently, herring fishing operations in this district are mainly performed by FRP boats.

Annual catch rates of herrings have considerably fluctuated (factor of 4 approximately) for the 1996 - 2006 period (Figure 2). The highest catch rates were reported in 1999 whereas the lowest catch rates were in 2000. The 1998 El Niño is a possible reason for declining catch rates in 2000 (Haputhantri 2004, Haputhantri *et al.* 2008).

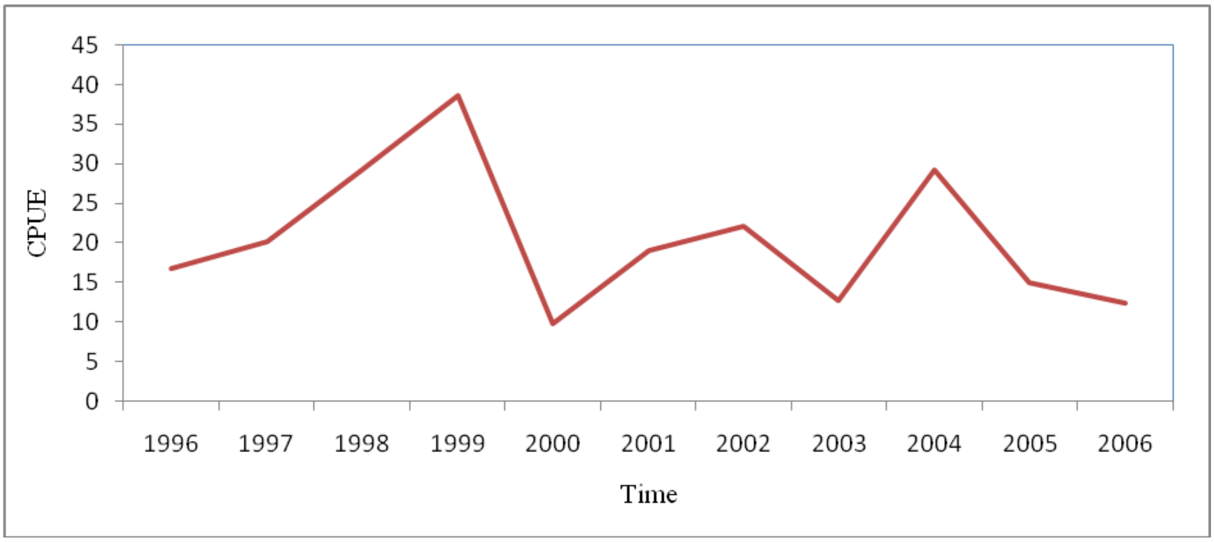


Figure 2: Annual variation of *A. sirm* CPUE in Negombo (catch in kg per FRP gillnet boat) for the 1996 -2006 period

A strong seasonal variation in the catch rates of herrings is also observed (Figure 3). The highest catch rates were mostly reported June – October whereas the lowest catch rates were reported in November – February. Catch and effort data of all years were lumped for drawing Figure 4 and it clearly shows month wise variation in the CPUE within a year. Accordingly, June–October is the peak season for herrings. The average CPUE of herrings per FRP gillnet boat day was estimated at 22.4 kg per FRP gillnet boat day.

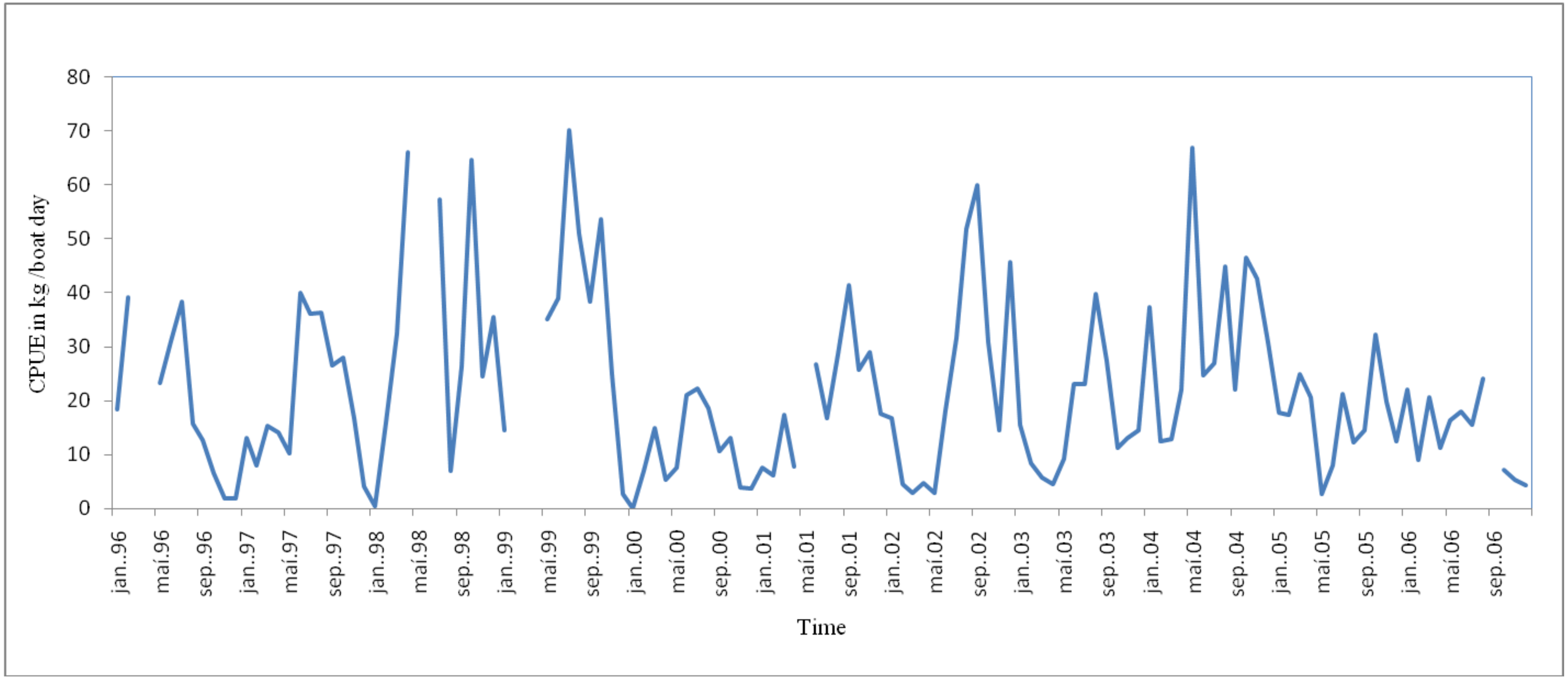


Figure 3: Monthly CPUE of *A. sirm* in Negombo based on small pelagic fishery landings data of NARA (January – December 2006)

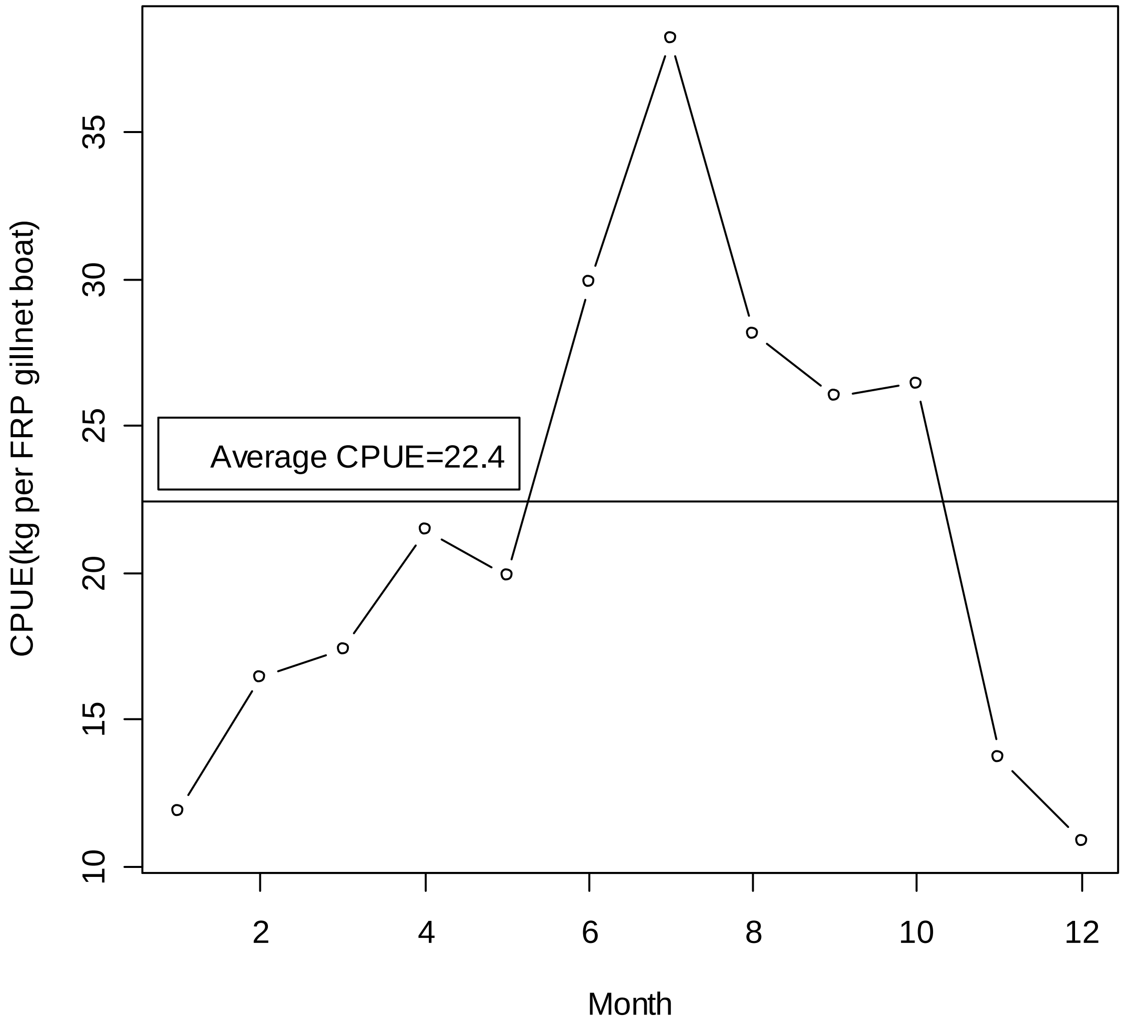


Figure 4: Month wise variation in *A. sirm* CPUE in kg per FRP gillnet boat days in Negombo fisheries district (based on 1996 – 2006 herring catch and effort data)

An inconsistency was noted in the available length data of west coast herring landings provided from the present data collection programme (Table 2). The collected length measurements were found to be few and therefore contain too much random variation for analysing them separately in relation to different years for further assessments. After rounding up to 0.5 cm of the length measurements, they were lumped together for the whole period for drawing the length distributions quarterly (Figure 5).

Table 2: Reported length measurements of herrings from the NARA small pelagic fishery monitoring programme

Year	Month	Number of samples	Range of the sample size
2002	January	2	14-16
2002	February	4	14-20
2002	March	2	14-15
2002	April	8	11-21
2002	May	4	16-23
2002	June	9	13-21
2002	July	17	17-25
2002	August	13	18-23
2002	September	13	17-19
2002	October	13	11-19
2002	November	5	14-19
2002	December	3	16-17
2003	January	7	15-22
2003	February	1	15
2003	March	2	16-18
2003	April	3	14-17
2003	May	2	16
2003	June	6	13-15
2003	July	6	13-17
2003	August	10	9-17
2003	September	8	15-23
2003	October	6	9-16
2003	November	6	11-15
2003	December	3	14-20
2004	January	4	17-20
2004	February	6	14-19
2004	March	10	12-17
2004	April	4	14-23
2004	May	9	11-19
2004	June	15	11-16
2004	July	5	14-17
2004	August	6	12-16
2004	September	8	13-22
2004	October	10	7-16
2004	November	7	7-15
2004	December	6	11-21
2005	January	1	16
2005	February	5	15-19
2005	March	6	14-23
2005	April	5	13-18
2005	May	1	17
2005	June	0	0
2005	July	11	15-22
2005	August	4	19-23
2005	September	3	14-16
2005	October	16	8-29
2005	November	6	12-19
2005	December	10	11-24
2006	January	14	14-23
2006	February	15	9-23
2006	March	14	7-24
2006	April	6	11-20
2006	May	13	7-24
2006	June	10	11-22
2006	July	9	12-25
2006	August	10	12-25

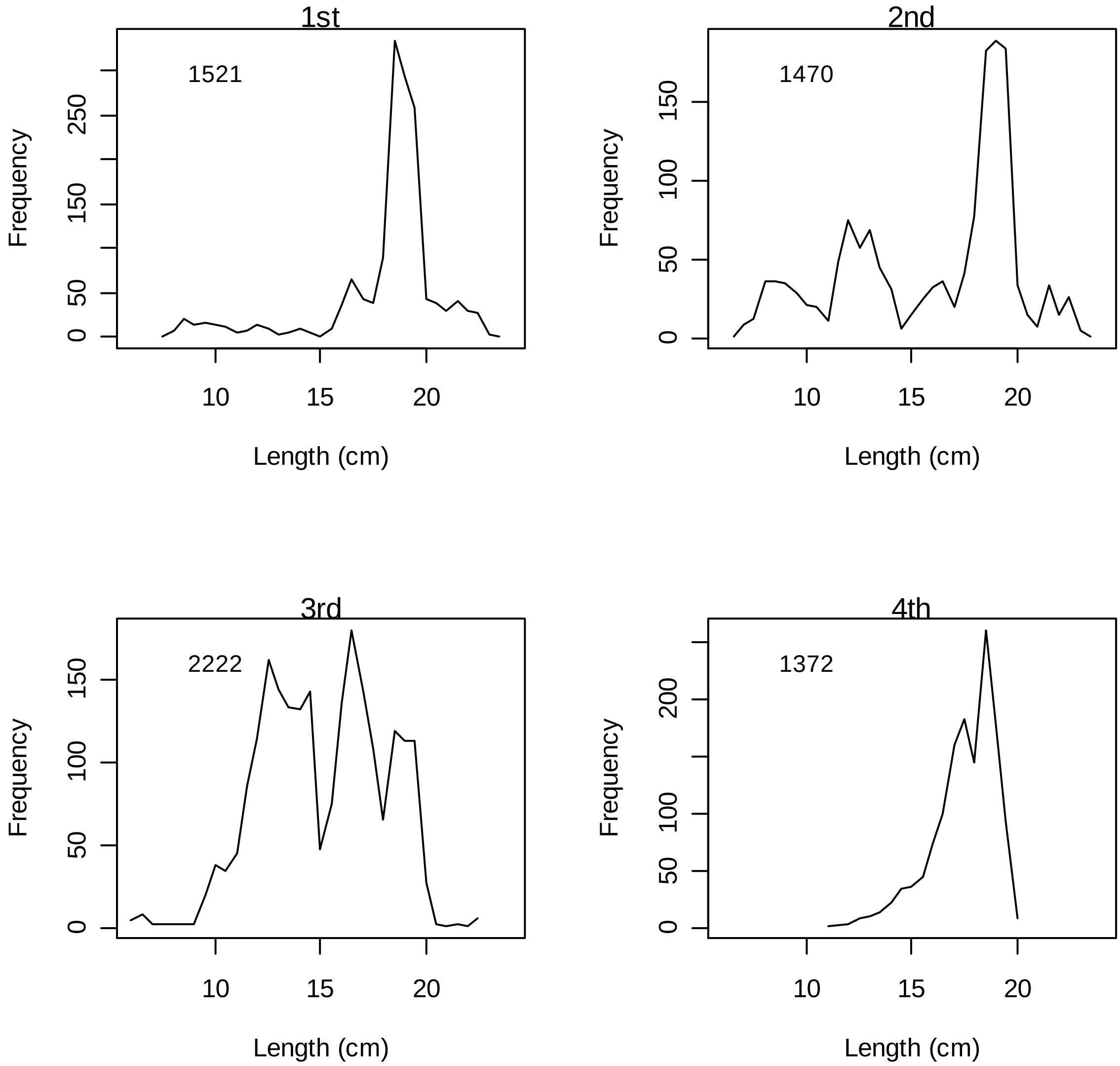


Figure 5: Quarterly variation in the length distribution of *A. sirm* measured from gillnet catches of the vessels operated in Negombo fisheries district in the 2002- 2006 period. Numbers in upper left are the numbers of *A. sirm* measured

From Figure 5 one can distinguish a single "cohort" in the 1st and 4th quarters whereas there appear to be two "cohorts" in the 2nd and 3rd quarters. There are some difficulties in distinguishing the two separate "cohorts" because of insufficient collected length measurements.

2.7 Conclusion

Collected length measurements of *A. sirm* in Negombo fisheries district by the NARA small pelagic fishery landings monitoring programme are not sufficient for applying length-based stock assessment methods to the data directly. Therefore, an alternative approach might be required for stock assessment. For example, a simulation approach could be attempted. However, it is important to quantify the actual data requirements so that a length-based stock assessment on *A. sirm* could be conducted in the future. The aim of the present study is to propose a better data collection strategy for future stock assessments on herrings.

3 MATERIAL AND METHODS

3.1 Review of previous work conducted to determine sample design and sample size in fisheries related studies

Many fisheries organisations conduct their own fisheries data collection programmes for monitoring of fisheries resources. Management decisions are widely taken based on quantitative estimates made using data obtained from those data collection programmes. Data sampling strategies adopted by respective fisheries organisations can greatly impact the accuracy of estimates. Therefore, a risk is always associated with management decisions which are to no small degree related to the data collection programmes. Sampling strategies are further constrained by the cost of data collection. Because of the above reasons, many fisheries organisations are either progressively searching for better data collection strategies or making improvements to currently practicing ones to be better able to give advice to managers for sustainable utilisation of fishery resources (Goodyear 1995, Brouwer and Griffiths 2005).

Few studies have yet been conducted to address the effect of sample design and sample size in estimating or quantifying the parameters related to fisheries (Goodyear 1995, Brouwer and Griffiths 2005, Yamaguchi and Matsuishi 2007). The data may be collected to estimate mean length of fish at a given age or species distribution/abundance or to estimate growth and mortality parameters. Because reference points calculated for fisheries management normally depend on the estimates mentioned above, they should always be estimated with the highest possible precision. Suppose if a fewer number of observations were used for an estimation of a parameter than to the actual number of observations required then this could lead to an imprecise estimation and hence erroneous management recommendations. On the other hand, if more samples were taken to estimate the same parameter then an additional cost has to be borne. Therefore, practising an appropriate data collection strategy is advisable (Kritzer *et al.* 2001, Brouwer and Griffiths 2005).

The methods mostly employed to determine the (minimum) sample size and a sample strategy are based on statistical bootstrapping (Efron and Tibshirani 1993) and this method uses re-sampling procedures to determine them. Bootstrap procedures are widely used for genetic studies in fisheries (Wenburg *et al.* 1998, Rüber *et al.* 2004, Martínez *et al.* 2006) but only few applications are found in fish stock assessment (Goodyear 1995, Kritzer *et al.* 2001, Yamaguchi and Matsuishi 2007).

Influence of sample design on estimates of growth and mortality parameters of fish have been studied previously by Goodyear (1995) and Brouwer and Griffiths (2005). Random sampling and size-based stratified sampling are widely used sampling techniques in fisheries science. The influence of sampling protocol on the estimation of mean lengths at age was evaluated by Goodyear (1995) using a computer simulation of a population of red grouper, *Epinephelus morio*. A positive bias in the mean lengths of older age classes and a negative bias in the younger age classes were observed. Samples drawn from size-selective gears or fisheries results in biased estimates of mean length at age. Accordingly, length-stratified samples and samples drawn from length-selective sampling gears all provide biased estimates. This means if a growth model was fitted using samples obtained, the parameter estimates would be imprecise. Therefore, avoiding length-stratified sampling was recommended by Goodyear (1995) for growth estimates and for developing models of mean length at age.

Another study was conducted by Brouwer and Griffiths (2005) to determine minimum sample size and data collection strategy on estimates of growth and mortality in carpenter (*Argyrozona argyrozona*) but different results were obtained. They tested both random and stratified sampling methods for carpenter and finally suggested length-based stratification because this sampling method was found to be unbiased and cost effective. It was also found that timing and frequency of sampling events could make biased estimates of growth and mortality. They further concluded that a minimum of 300 random samples or at least 10 fishes per 2 cm size class (i.e. total of 193 fishes) were necessary for reliable estimates of growth and mortality. A stratified sample of 10 fishes per 2 cm size class that is taken during the spawning season was the best sampling strategy for estimates of growth and mortality parameters since this optimised the trade-off between cost and parameter precision.

Length-frequency data are widely used to derive growth estimates, especially in tropical fishes. A study was conducted by Gomez-Buckley *et al.* (1999) for sample size determination to estimate length-frequency distributions for Pacific albacore tuna (*Thunnus alalunga*). A large sample of length-frequency data from longline fisheries collected by three countries during the period 1962 – 1995 was used for this. The main question addressed by this study was regarding the appropriate number of length frequency samples to collect to obtain representative distributions for albacore tuna. Accordingly, it was investigated (1) how many fish lengths to record from each vessel and (2) how many vessels would need to be sampled. For this, cumulative distribution functions were computed for the data stratified by country, year and quarter, and a bootstrapping model based on statistical re-sampling was used to investigate maximum differences between original distributions and random samples of sizes ranging from 25 to 200. Ninety-nine percent of the time, the maximum difference exceeded 0.05 when $n = 25$ whereas this was reduced to 58% of the time when $n=200$. The effect of internal correlations in length measurements of a sample was, however, not considered for this analysis.

Sampling methods were tested to quantify fish discards at fleet level by Northern Ireland trawlers (Allen *et al.* 2001). Two sampling methods, equal probability and probability proportional to x (ppx), were investigated here to determine the optimal sampling strategy. The study revealed that in some instances the improvement in precision when using the ppx sampling method was marginal over the simple equal probability method.

Effects of sample size and population structure on the parameter estimates were extensively discussed by Kritzer *et al.* (2001) using real data sets of four reef fishes. They examined the precision of size, age, growth and mortality parameters at sample sizes ranging from 25 to 1000 using bootstrapped population samples. Parameters such as mean length, mean age and modal age were consistently estimated with high precision at small sample sizes compared to other parameters and therefore they suggested estimations of those for monitoring less commercially important populations under limited data availability. Moreover, longer lived species require fewer samples per age-class. For example, 7 – 10 fish per age-class for those species living more than 25 or 30 years. They further suggested a very useful method to figure out the sample size necessary to be taken from the population in order to obtain the results with high precision in case enough data are not available to go for such estimations. An extant data set of a species similar to focal species: e.g. exhibiting similar demographic traits can be used here to approximate the population in question. Accordingly, additional sampling is required whenever the newly researched population is found to be different from proxy population to a large extent.

3.2 Bootstrapping capelin length frequency data of Icelandic surveys

The *A. sirm* length measurements collected in the NARA small pelagic fishery monitoring programme were found to be inadequate for conducting a stock assessment. Therefore, an attempt was made to estimate the number of length measurements actually required for attaining this goal. The non-parametric bootstrap technique was used to determine the adequate sample size of length measurements and number of samples. The bootstrap technique is a statistical method that allows random re-sampling of data from an original data set. The bootstrapping routine was implemented in R (R Development Core Team 2005).

Capelin (*Mallotus villosus*) length-frequency survey data obtained from annual capelin surveys in Icelandic waters conducted by the Marine Research Institute (MRI) of Iceland were used to estimate optimal sample size and number of samples for *A sirm*. Capelin is a small pelagic species widespread in the oceans of the northern hemisphere. Maximum size of the capelin is 20.0 cm TL (male/unsexed) and 25.2 cm TL (female) (Froese and Pauly 2007). It spawns at the age of 3-4 years and the males and most of the females die after spawning (Icelandic Ministry of Fisheries 2008).

It should be noted that the length range of capelin data was quite similar to the measured length range of *A. sirm* in the coastal waters of Sri Lanka. Also the capelin data shows the same characteristics as the *A sirm* data in terms of internal correlation, i.e. some samples only contain small fish whereas other contain only large individuals. Fish landings are usually not uniformly mixed; therefore individual samples taken from a boat are not random samples. There were well over 81,000 capelin length measurements having the length range 5.5 – 27.5 for this analysis. The data matrix of the capelin length-frequency data contained three data columns: row number, id and length. The data were basically sorted by the row number. Lengths were measured up to the nearest 1 mm. Accordingly, length measurements were first rounded up to the closest 0.5 cm before bootstrapping.

Random samples were then taken (without replacements) using the “sample function” in R.. The length distribution of the generated new data set by bootstrapping was compared with the original length distribution by estimating the variance between observed and the actual length distributions. The observed variance for a given

combination was estimated from averaging the variances obtained for each bootstrap sample. The number of bootstrap iterations was set at 100.
The following values were assigned for sample size:

A $\equiv \left\{ \frac{x}{x} \in N and\ 1 \le x \le 19 \right\} \cup \{20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$

Following values were assigned for number of samples,

B $\equiv \left\{ \frac{x}{x} \in N and\ 1 \le x \le 20 \right\} \cup \{30, 40, 50, 75, 100\}$

Since there is a possibility of having some correlations among the length measurements, the above procedure was again slightly modified for better results. Accordingly, random samples were first taken (without replacements) from randomly selected ids. Length measurements were then randomly taken from each selected sample. The procedure was repeated 100 times for each sets of values of A and B.

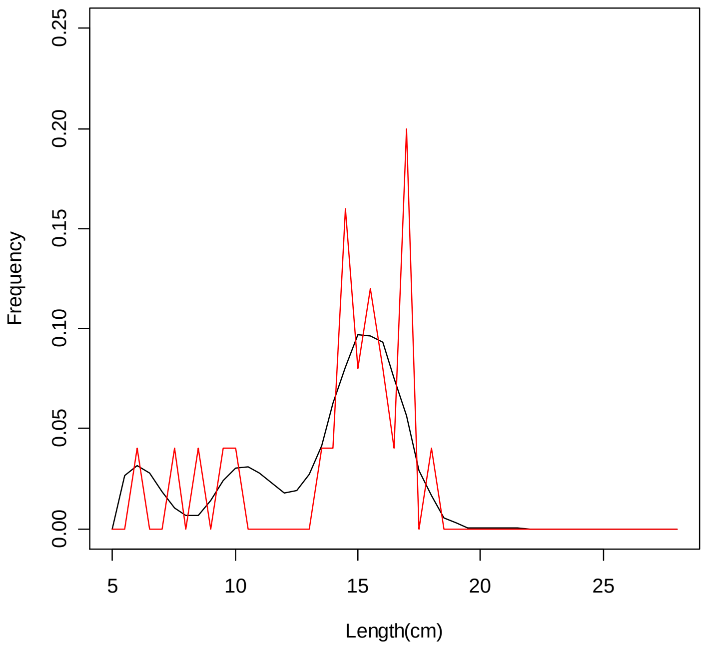
Contour plots of variance against sample size and number of samples were also drawn for visualising of above bootstrap results. These plots were drawn for generated data from different bootstrap scenarios. Accordingly, a suitable combination of sample size and number of samples was determined using the variance plot obtained after taking into consideration internal correlations in the length measurements. This combination was finally suggested as a minimum length-frequency data requirement for conducting length based stock assessments on herring in the coastal waters of Sri Lanka.

4 RESULTS

4.1 Random sampling, ignoring internal sampling correlation

As sample size and number of samples increase the variance between the original length distributions and bootstrapped length distributions decreases (Figure 6).

(a)



(b)

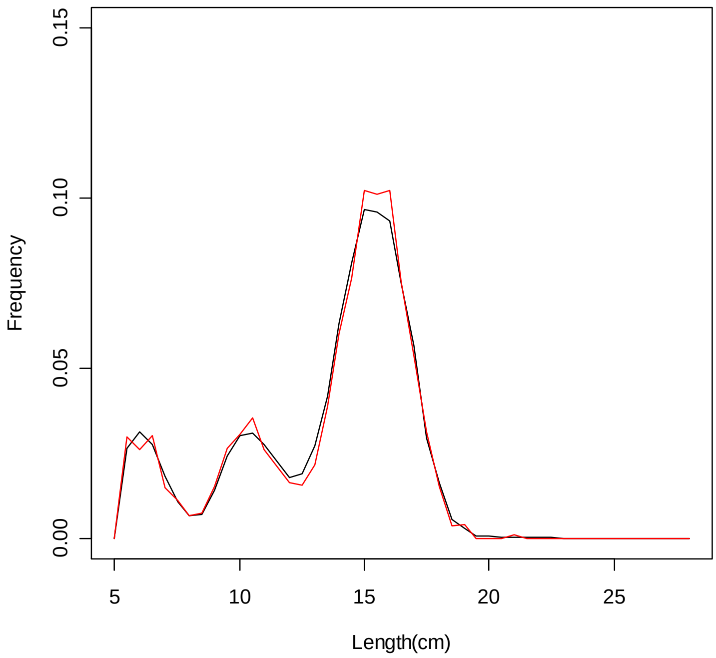


Figure 6: Length distribution of Icelandic capelin data (black lines) and distribution of the length data generated by bootstrapping (in red) when sample size and number of samples are set at five (a) and set at fifty (b) (ignoring internal sampling correlation)

Comparatively high stability in the contour plot of the variance was noticed for either increased sample size or number of samples (Figure 7). The variance can substantially be reduced either by increasing the sample size or number of samples.

As shown in Figure 7, variance contours are parallel to two principal axes of the graph towards beyond some definite values. Respective values for both sample size and number of samples are apparently 20. Accordingly, 20 - sample size and 20 – number of length samples might be substantial for a length-based stock assessment subject to the assumption that there are no internal correlations between length measurements in a sample. This should be attained on a monthly basis throughout the year.

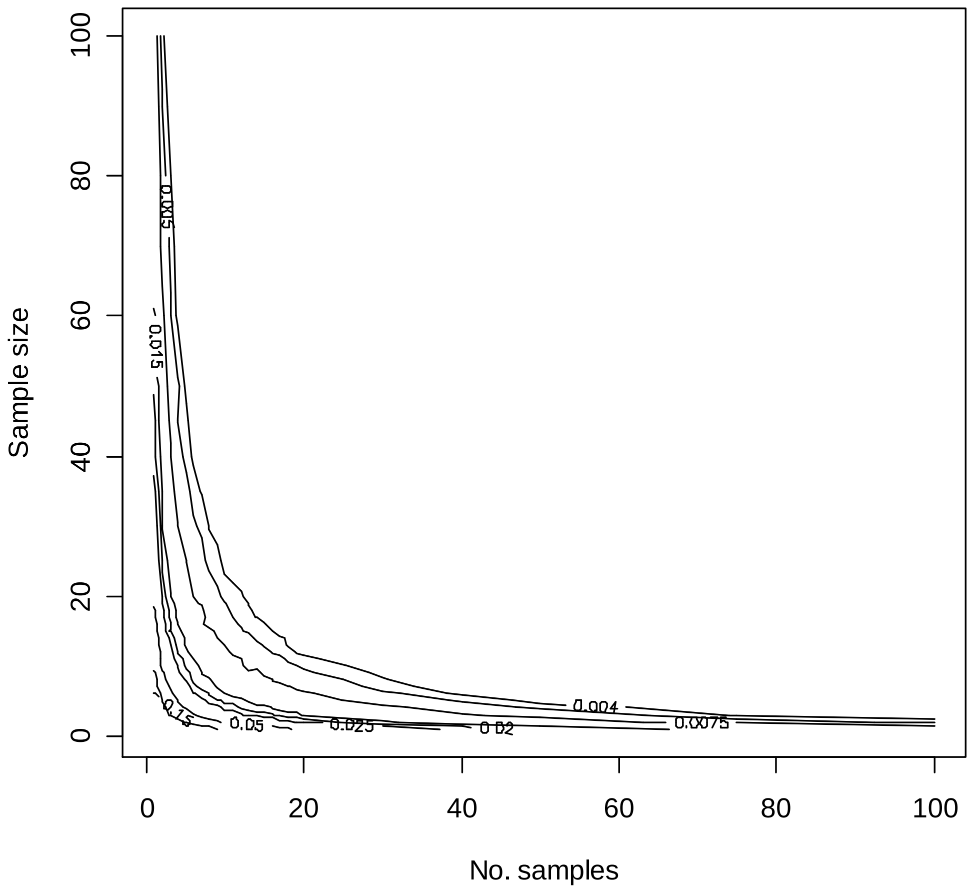


Figure 7: The contour plot of variance against sample size and number of samples obtained by bootstrapping the capelin length-frequency survey data in Icelandic waters from dividing the whole length data set into samples (sample size and number of samples were set over the values on sets A and B respectively)

4.2 Random sampling, taking internal sampling correlation into account

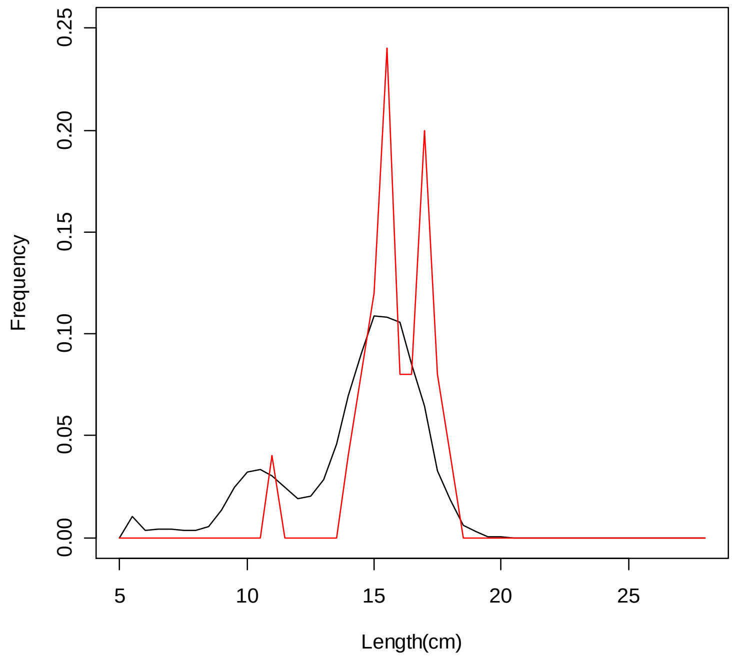
As sample size and number of samples increase, the variance between the original length distributions and bootstrapped length distributions decreases considerably in this case too (Figure 8).

The contour plot of variance against sample size and number of samples has apparently changed after taking into consideration possible correlations in length measurements in a sample (Figure 9). The observed pattern in variance contours are however more similar to those obtained previously in Figure 7. 40 - sample size and 60 – number of samples is the data requirement now for a precise length-based stock assessment. Accordingly, more length measurements are required now owing to correlations in the length measurements. However, 20 - sample size and 40 – number of samples might perhaps be substantial to recommend when cost for data collection is concerned and this would optimise the trade-off between cost and parameter precision.

5 DISCUSSION

Stock assessments are conducted to aid management recommendations of fishery resources. Management reference points such as $F_{0.1}$ and F_{max} are normally determined from the data obtained via fishery catch monitoring programmes or annual resource surveys. Obtaining good quality data from either method is equally important for good management of resources and therefore, attention to data sampling strategies is of great importance. Priority is always given by the organisations engaged in fisheries data collection to improve the data collection mechanism but a little emphasis is paid to the determination of actual data requirements for conducting stock assessments. Examples of this can be found both in the NARA data collection programme as well as in the MRI data collection programme. Ignoring this important factor might lead to poor management of fisheries resources via imprecise estimates. Quantifying data requirements might be important in at least two ways. Data requirements can be predetermined with respect to stated sampling cost before conducting a stock assessment so enabling the institutions engaged in data collection to plan the data collection programmes in order to obtain precise estimates. Also, this creates an avenue to optimise the trade-off between cost and parameter precision. However, minimum data requirements might vary from one estimate to another of two parameters even within the same fish species. As mentioned previously, parameters such as mean length and mean age are estimated with high precision at small sample sizes when compared with other parameters such as growth and mortality parameters. Therefore, additional care must be taken on the above factors in sampling design in the process of conducting stock assessment (Kritzer *et al.* 2001, Brouwer and Griffiths 2005).

(a)



(b)

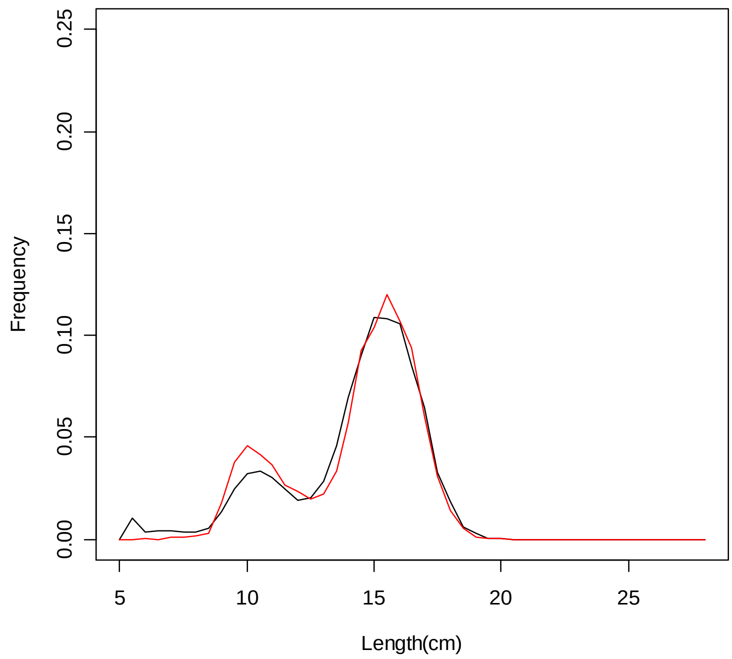


Figure 8: Length distribution of Icelandic capelin data (black lines) and distribution of the length data generated by bootstrapping (in red) after taking internal sampling correlation into account when sample size and number of samples are set at five (a) and set at fifty (b).

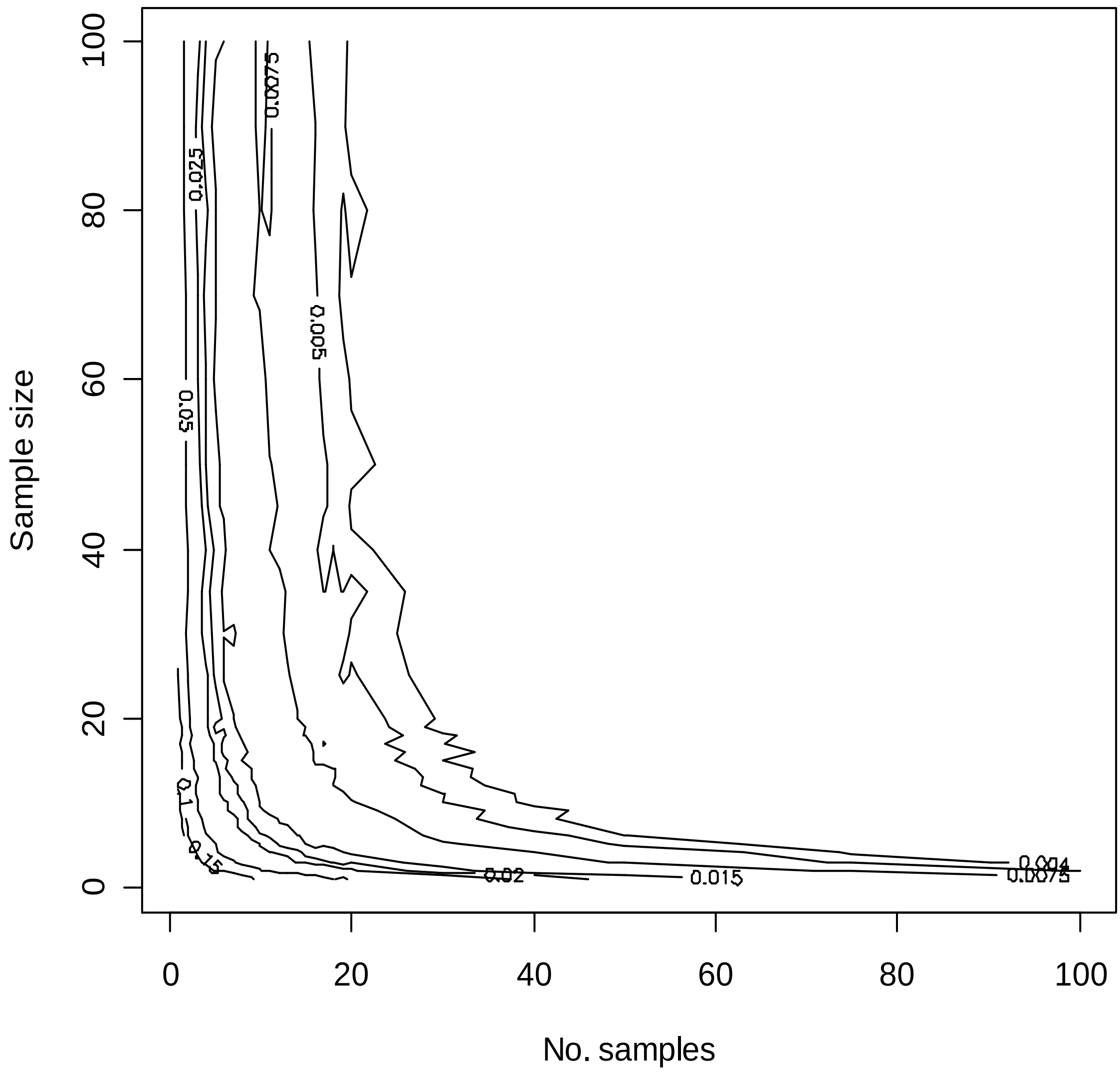


Figure 9: The contour plot of variance against sample size and number of samples obtained from bootstrapping the capelin length-frequency survey data in Icelandic waters (sample size and number of samples were set over the values on sets A and B respectively) after taking into the consideration of possible correlations in the measured lengths.

In this study, Icelandic capelin was used as a proxy population for proposing a sampling strategy for coastal herring of Sri Lanka. There are many differences between the two species in relation to the environment, distribution and biology. Herring is a short lived species (about 1.5 years) whereas capelin lives around 6 years (Fröese and Pauly 2007). The former is a coastal species whereas the latter is oceanic found in the depth range of 0 -725 m. However, there are close similarities in terms of data structure between the two species especially in relation to the length distribution. The maximum lengths of the two species are more or less the same. Therefore, using capelin data to formulate a data sampling strategy for herring is sensible. The data collection programmes of commercial fish landings in Icelandic waters are conducted in a well organised manner and the data obtained though such programmes always have high precision. Therefore, utilisation of such data for this assessment is reasonable. However, these data are not suitable for other purposes and can not be used for stock assessment work on herring.

In this study, the possible internal correlations in the length measurements in a sample was especially taken into account in the process of proposing a data collection sampling strategy for coastal herring in Sri Lanka. The length measurements taken from a fishing boat in the field cannot be considered as completely random. There might be some possibility of biased samples of length measurements and this may happen either when taking samples from catches dominated by small fishes or large fishes. Such biased samples were observed both in herring and capelin data. This study might perhaps be the first attempt at addressing the influence of the above factor when proposing a data sampling strategy for conducting fish stock assessment. This approach makes it possible to propose a sound data collection strategy for coastal herring of Sri Lanka so that length-based stock assessments could be conducted in the future.

Any sampling strategy proposed is bound by the fund allocations of respective fisheries organisations. This is highly applicable for research institutions like NARA in the developing world. Therefore, attention has to be paid equally to both the precision of the results and the cost of data collection. Accordingly, sampling strategy might have to be reconsidered when the two factors are equally important. 40- sample size and 60- number of samples may be the ideal for formulating of a sampling strategy for coastal herring. However, 20 - sample size and 40 – number of samples is probably sufficient when considering budget allocations for data collection.

Collecting length-frequency data throughout the whole year at the same frequency is important in future stock assessments. A considerable fluctuation in measured length samples of herring by the small pelagic fishery landings monitoring programme of NARA was observed for different months (Table 2). This was mostly characterised by a higher number of length measurements during the fishing season for herring whereas comparatively few length measurements were taken during the lean fishing season. This inconsistency in measured length samples is likely to be a considerable constraint when attempting to conduct stock assessment on herring. The minimum data requirement proposed here subject to budget constraints would be an attainable target. 20 - sample size and 40 – number of samples is achievable during the lean fishing season too.

The same data sampling strategy can be adopted for other fisheries districts too. Perhaps, this minimum data requirement could be covered from a few fisheries districts jointly. For example, the sampling programme could be carried out for the whole west coast so covering two important fisheries districts (Negombo and Chilaw) together. However, before going for such a strategy, separate data collection programmes should first be adopted for two districts for a minimum period of one year fulfilling minimum length data requirements. By comparing the monthly length distributions of two districts after the period, it can easily be determined whether the two districts can be merged or not. Therefore, this perhaps makes avenue for going for a better option of measuring more length samples (40- sample size and 60- number of samples) in order to attempt more precise results.

The data obtained for herring from the NARA small pelagic fishery landings monitoring programme can also be used for determining an appropriate data sampling strategy for coastal herring by generating a large number of hypothetical samples by bootstrapping (Kritzer *et al.* 2001). However, use of an empirical data set apparently seems to be sounder than the use of a hypothetical generated data set by bootstrapping to conduct this analysis. However, testing of this hypothesis could also be suggested as part of future research in determining a better sampling strategy.

6 CONCLUSION

Conducting stock assessments using length-based stock assessment models is important for proposing management recommendations for sustainable utilsation of fisheries resources. Currently, the data sampling strategy of the NARA small pelagic fishery landings monitoring programme needs to be substantially revised for conducting such stock assessments on the herring fishery of Sri Lanka in the future. An inconsistency was noted in the available length data of west coast herring landings provided from the present data collection programme. This study proposed a better length data collection sampling strategy which is cost effective and may give precise data for conducting stock assessments on herring in the future. The bootstrapping scenarios employed here for quantifying the minimum sample size and number of samples to determine the minimum length data requirement for conducting length-based stock assessments on coastal herring of Sri Lanka in future seem to be useful.

ACKNOWLEDGEMENTS

I would like to thank the United Nations University for granting me a fellowship to participate for this Fisheries Training Programme in Iceland. My sincere gratitude especially to Dr. Tumi Tómasson, Programme Director - UNU/FTP, Mr. Þór Ásgeirsson, Deputy Director of the programme, Mr. Guðni Magnús Eiríksson, the Programme Manager and Mrs. Sigríður Kr. Ingvarsdóttir, the Programme Officer for their invaluable support and encouragement provided to me to complete this programme successfully. I would like to extend my thanks to the Director and staff of the Marine Research Institute (MRI), Iceland for their warm welcome and assistance. My sincere thanks to my two supervisors, Professor Gunnar Stefánsson and Dr. Guðmundur Þórðarson for their excellent guidance. My thanks are to be extended to the Chairman and Director General of the National Aquatic Resources Research and Development Agency (NARA), Sri Lanka for nominating me for this training programme. I would like to thank my colleagues and staff members of the Marine Biological Resources Division of NARA for their valuable assistance. Finally, my wife and two kids are among those who deserve my sincere thanks.

REFERENCES

Allen, M., Kilpatrick, D., Armstrong, M., Briggs, R., Perez, N. and Course, G. 2001. Evaluation of sampling methods to quantify discarded fish using data collected during discard project EC 95/094 by Northern Ireland, England and Spain. *Fisheries Research.* 49, 241–254.

Brouwer, S.B. and Griffiths, M.H. 2005. Influence of sampling design on estimates of growth and mortality in *Argyrozona argyrozona* (Pisces: Sparidae). *Fisheries Research* 74, 44-54.

Dayaratne, P. 1984. Fishery and biology of some small pelagic fish species (Clupeioides) from the West Coast of Sri Lanka. Unpublished Ph.D. dissertation, University of Bergen, Norway.

Dayaratne, P. 1988. Gillnet selectivity for *Amblygaster* (= *Sardinella*) *sirm*. *Asian Fisheries Science* 2: 71-82.

Dayaratne, P. and Gjϕsaeter, J. 1986. Age and growth of four *Sardinella* species from Sri Lanka. *Fisheries Research* 4:1-33.

Dayaratne, P. and Sivakumaran, K.P. 1994. Bio-socio-economics of the fishery for small pelagics along the south-west coast of Sri Lanka. BOBP/WP/96:38 p.

Efron, B. and Tibshirani, R.J. 1993. *An introduction to the Bootstrap*. Chapman Hall: New York.

Fröese, R and Pauly, D. 2007. Fishbase. *World Wide Web electronic publication*. [November 20, 2007] <http://www.fishbase.org>.

Gomez-Buckley, M., Conquest, L., Zitzer, S. and Miller, B. 1999. Use of statistical bootstrapping for sample size determination to estimate length-frequency distributions for pacific albacore tuna (Thunnus alalunga). USA: National Marine Fisheries Services [January 15, 2008]

Goodyear, C.P. 1995. Mean size at age: An evaluation of sampling strategies with simulated Red Grouper data. *Transactions of the American Fisheries Society* 124, 746-755.

Haputhantri, S.S.K. 2004. Assessment of a coastal fishery of Sri Lanka using single-species and multi-species approaches. Unpublished Ph.D. dissertation, Institut National Polytechnique de Toulouse, France.

Haputhantri, S.S.K. 2006. A simple method for determining optimal fishing strategies using Beverton and Holt yield per recruit model in fish stocks with parameter uncertainties. *Journal of the National Science Council of Sri Lanka* 34: 125-129.

Haputhantri, S.S.K., Villanueva, M.C.S. and Moreau, J. 2008. Trophic interactions in the coastal ecosystem of Sri Lanka: An ECOPATH preliminary approach. *Estuarine Coastal and Shelf Science* 76: 304 – 318.

Holt, S.J. 1963. A method for determining gear selectivity and its application. *ICNAF Special Publication* 5: 106–115.

Icelandic Ministry of Fisheries 2008. Important statement on responsible fisheries in Iceland. Iceland: Icelandic Ministry of Fisheries [February 06, 2008] <http://www.fisheries.is/stocks/capelin.htm>

Karunasinghe, W.P.N. 1990. Some aspects of biology and fishery of trenched *Amblygaster sirm* (Pisces: Clupeidae) from the coastal waters around Negombo, Sri Lanka. Unpublished M. Phil. dissertation, University of Kelaniya, Sri Lanka.

Karunasinghe, W.P.N. and Wijeyaratne, M.J.S. 1991a. Selectivity estimates for *Amblygaster sirm* (Clupeidae) in the small-meshed gillnet fishery on the west coast of Sri Lanka. *Fisheries Research* 10: 199-205.

Karunasinghe, W.P.N. and Wijeyaratne, M.J.S. 1991b. Population dynamics of trenched sardine *Amblygaster sirm* (Clupeidae) in the western coastal waters of Sri Lanka. *Asian Fisheries Science* 4: 329-334.

Karunasinghe, W.P.N. and Wijeyaratne, M.J.S. 1998. Reproductive biology of the trenched sardine *Amblygaster sirm* (Walbaum) from the western coastal waters of Sri Lanka. *Ceylon Journal of Science (Bioscience)* 25:5-51.

Karunasinghe, W.P.N., Fernando, P.A.T. and Samaraweera, E.K.V. 2000. Status of the gillnet fishery for the small pelagic fish species in Western, South-Western and Southern coastal waters of Sri Lanka during 1995-1997. *Journal of the National Aquatic Resources Agency, Sri Lanka* 36: 66-79.

Kritzer, J.P., Davies, C.R. and Mapstone, B.D. 2001. Characterizing fish populations: effects of sample size and population structure on precision of demographic parameter estimates. *Canadian Jounal of Fisheries and Aquatic Sciences* 58, 1557–1568.

Martínez, P., González, E. G., Castilho, R. and Zardoya, R. 2006. Genetic diversity and historical demography of Atlantic bigeye tuna (*Thunnus obesus*). *Molecular Phylogenetics and Evolution* 39, 404-416.

MFAR 2007. Census of fishing boats 2006/2007. Sri Lanka: Ministry of Fisheries and Aquatic Resources [November 18, 2007] < http://www.fisheries.gov.lk/policy/Census.htm>

R Development Core Team 2005. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rüber, L., Britz, R., Kullander, S. O. and Zardoya, R. 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. *Molecular Phylogenetics and Evolution* 32, 1010-1022.

Samaranayake, R.A.D.B. 2003. "Review of national fisheries situation in Sri Lanka". In Silvestre, G., Garces, L., Stobutzki, I., Ahmed, M., Valmonte-Santos, R.A., Luna, C., Lachica- Aliño, L., Munro, P., Christensen, V. and Pauly, D. eds. *Assessment, Management and Future Directions of Coastal Fisheries in Asian Countries,* pp. 987 - 1012. WorldFish Center Conference Proceedings 67, 1 120 pp.

Sanders, M.J. and Dayaratne, P. 1999. Yield assessment for the small pelagic fishery occurring along the northwest, west and south coasts of Sri Lanka. *Asian Fisheries Science* 12: 25-40.

Siddeek, M.S.M., Joseph, L., Jayasuriya, P.M.A. and Karunasinghe, W.P.N. 1985. A preliminary analysis of length frequency data of *Amblygaster sirm* from Negombo, Sri Lanka, using ELEFAN programs. *Journal of the National Aquatic Resoures Ageny of Sri Lanka* 32: 35-46.

Sivasubramaniam, K. 1999. *One hundred years of fisheries management in Sri Lanka: lessons for the future*. The Department of Fisheries and Aquatic Resources, New Secretariat Building, Maligawatte, Colombo- 10, Sri Lanka. 156 pp.

Wenburg, J.K., Bentzen, P. and Foote, C.J. 1998. Microsatellite analysis of genetic population structure in an endangered salmonid: the coastal cutthroat trout (*Oncorhynchus clarki clarki*). *Molecular Ecology* 7**,** 733–749.

Yamaguchi, H. and Matsuishi, T. 2007. Effects of sampling errors on abundance estimates from virtual population analysis for Walleye Pollock in northern waters of sea of Japan. *Fisheries Science* 73, 1061-10.

APPENDICES (BOOTSTRAPPING)

Appendix i

First few rows of capelin data file at the MRI in Iceland which consists of over 81 000 such data entries

row.names	**id**	**le**
1	221741	12.5
2	221741	12.5
3	221741	10.5
1220	138952	16
1221	138952	16
1222	138952	16

Appendix ii

R code used for bootstrapping of capelin length data from dividing the whole length data set into samples

```
#Reading the capelin data set and rounding up the length measurements to the nearest
0.5 #cm
dat<-read.table("capelinData.dat", header=T)
dat$le<-round(dat$le*2)/2

#plot capelin length distribution
ldst<-table(dat$le)
l.names<-seq(5,28,0.5)
 full<-rep(0,length(l.names))
 names(full)<-l.names
 lens<-names(ldst)
 full[lens]<-ldst
 plot(l.names,full, xlab="Length(cm)",ylab="Frequency",type="l")

set.seed(1)
#select values for sample size and number of samples
s.size<-c(1:19, seq(20,50,5),60,70,80,90,100)
s.no<-c(1:20,30,40,50,75,100)

data.matrix<-NULL

for(k in s.no){
        for(i in s.size){
                for(j in 1:100){
        plot(l.names,full/sum(full), ylim=c(0,0.14), xlab="Length(cm)",
        ylab="Frequency",type="l",xaxt="n", yaxt="n")
        #estimate maximum number of samples
                s.max<-floor(length(dat$le)/k)
        #take a random sample
                s.vec<-sample(rep(1:s.max,k))
```

```
        init.dat<-cbind(s.no=s.vec, dat[1:length(s.vec),])
    #take random length measurements from a selected sample
        init.sample<-sample(s.vec,i)
        s.dat<-init.dat[!is.na(match(init.dat$s.no,init.sample)),]
        s.ldst<-table(s.dat$le)
        s.full<-rep(0,length(l.names))
        names(s.full)<-l.names
        s.lens<-names(s.ldst)
        s.full[s.lens]<-s.ldst
    #plot the generated length distribution
    lines(l.names,s.full/sum(s.full),type="l", col="red")
    #compute the variance between original and generated length
    #distributions
        var.init<-sum(((s.full/sum(s.full))-(full/sum(full)))^2)
        data.init<-c(j,i,k,var.init)
        data.matrix<-rbind(data.matrix,data.init)

    }
    }
}
#arrange variances as a matrix for each combination of sample size, sample #number
and bootstrap run
dimnames(data.matrix)<-list(1:nrow(data.matrix), c("run","s.si","s.no","variance"))
data.matrix<-data.frame(data.matrix)
#compute the mean variance for a given sample size and sample number
variance.matrix<-tapply(data.matrix$variance, list(data.matrix$s.no, data.matrix$s.si),
mean)

#contour plot
x<-s.no
y<-s.size
z<-variance.matrix[1:(length(s.no)*length(s.size))]
func<-function(x,y)z
plot(range(x),range(y), type="n", xlab="No. samples", ylab="Sample size")
contour(x,y,outer(x,y,func),add=T,
levels=c(0.15,0.1,0.05,0.025,0.02,0.015,0,01,0.0075,0.005,0.004))
```

Appendix iii

R code used for bootstrapping of capelin length data after taking into account of possible correlations of measured length data

```
# Reading the capelin data from the data file and rounding up the length measurements to
# the nearest 0.5 cm
le3<-read.table("capelinData2.dat", header=T)

capelin.tot<-table(le3$le)
le.vec<-seq(5.5,24,0.5)
le.dist<-rep(0,length(le.vec))
names(le.dist)<-paste(le.vec)
id<-names(capelin.tot)
le.dist[id]<-capelin.tot
capelin.tot<-le.dist
capelin.tot2<-capelin.tot/sum(capelin.tot)


#select values for sample size and number of samples

s.size<-c(1:19, seq(20,50,5),60,70,80,90,100)
s.no<-c(1:20,30,40,50,75,100)
n.run<-100
#sort data set to the ascending order with respective to id
samp.av<-sort(unique(le3$id))
samp.mat<-NULL
set.seed(1)

for(b in 1:n.run){
for(k in s.size){
        for(a in s.no){
        #random samples are taken from randomly selected ids
                my.samp<-sample(samp.av,a, replace=F)
                le.samp<-NULL
                for(j in my.samp){

            #length measurements are taken from randomly selected ids
                        s.vec<-sample(le3$le[le3$id==j],k, replace=F)
                        le.samp<-c(le.samp,s.vec)
                }
        #find the length distribution and variance of bootstrapped data
                tab.boot<-table(le.samp)/length(le.samp)
                le.dist<-rep(0,length(le.vec))
                names(le.dist)<-paste(le.vec)
                id<-names(tab.boot)
                le.dist[id]<-tab.boot
                tab.boot<-le.dist
                cap.var<-sum((tab.boot-capelin.tot2)^2)
```

```
        #variance estimations
                init<-c(a, k, b, cap.var)
                samp.mat<-rbind(samp.mat,init)
        }
}}
dimnames(samp.mat)<-list(1:nrow(samp.mat),c("s.no","s.size","run","variance"))
samp.mat<-data.frame(samp.mat)
#estimate mean variance for combinations of sample size and sample number
variance.matrix<-tapply(samp.mat$variance, list(samp.mat$s.no, samp.mat$s.si),
mean)

#contour plot
x<-s.no
y<-s.size
z<-variance.matrix[1:(length(s.no)*length(s.size))]
func<-function(x,y)z
plot(range(x),range(y), type="n", xlab="No. samples", ylab="Sample size")
contour(x,y,outer(x,y,func),add=T,
levels=c(0.15,0.1,0.05,0.025,0.02,0.015,0,01,0.0075,0.005,0.004))
```