Stock Assessment of Tropical Fishes

Training Manual

Stock Assessment of Tropical Fishes

Training Manual

For bibliographic purpose this document should be cited as:

ICAR-CMFRI/BOBP-IGO/GoI, 2016. Training Manual on Stock Assessment of Tropical Fishes, November 2016, 156p.

Editing: *E Vivekanandan, Somy Kuriakose, Y S Yadava*
Layout, Design & Graphics: *S Jayaraj, M Krishna Mohan*
Photographs: *S Jayaraj, Y S Yadava*
(All photographs used in this document are from the archives of the BOBP-IGO)

Published by:

Director
ICAR-Central Marine Fisheries Research Institute
Post Box No. 1603
Ernakulam North P.O.
Kochi – 682 018, Kerala, India

Director
Bengal of Bengal Programme
Inter-Governmental Organisation
91, St Mary's Road
Abhiramapuram
Chennai – 600 018, Tamil Nadu, India

Printed at L S Graphic Prints, Chennai, Tamil Nadu, India

Acknowledgements

The following staff members of the ICAR-Central Marine Fisheries Research Institute (CMFRI), Kochi; Fishery Survey of India (FSI), Kochi; and Bay of Bengal Programme Inter-Governmental Organisation (BOBP-IGO), Chennai served as the Faculty, Coordinators and Facilitators for the conduct of the 14-day Training Course on ‘Stock Assessment of Tropical Fishes’ for 12 Officials of the Government of the People’s Republic of Bangladesh, organized at ICAR-CMFRI, Kochi, Kerala, India from 23 November to 06 December 2016:

Faculty/Resource Persons

Dr A Gopalakrishnan, Director, ICAR-CMFRI
Dr T V Sathianandan, Principal Scientist & Head of Division, ICAR-CMFRI
Dr Somy Kuriakose, Principal Scientist, ICAR-CMFRI
Dr E M Abdussamad, Principal Scientist, ICAR-CMFRI
Dr Shoba Joe Kizhakudan, Principal Scientist, ICAR-CMFRI
Dr Grinson George, Senior Scientist, ICAR-CMFRI
Dr D K Gulati, Zonal Director, FSI
Dr E Vivekanandan, National Consultant, BOBP-IGO

Coordinators

Dr Y S Yadava, Director, BOBP-IGO
Dr Boby Ignatius, Principal Scientist, ICAR-CMFRI
Mr Rajdeep Mukherjee, Project Coordinator (OPP Project), BOBP-IGO
Mr S Jayaraj, Publication Officer, BOBP-IGO

Facilitators

Mr T B Ratheesh, Technical Assistant, ICAR-CMFRI
Mr K K Sajikumar, Technical Assistant, ICAR-CMFRI
Mr N Ragesh, Technical Assistant, ICAR-CMFRI

The Training Manual is based on the following two publications:

Vivekanandan, E. 2005. Stock Assessment of Tropical Marine Fishes. Indian Council of Agricultural Research, New Delhi, 115 p.

CMFRI, 2014. Fish Stock Assessment and Management. Training Manual. Central Marine Fisheries Research Institute, Kochi, 150 p.

***

Preface

Fisheries in the tropical ecosystem present unique challenges to the fisheries managers/researchers in estimating the stock abundance and drawing management plans for optimum harvest of the resources. Unlike temperate waters, tropical fisheries are characterized by high diversity and multiplicity of gear, of which many are non-selective and harvest a variety of species in a single haul. Such attributes of tropical fisheries make estimations more challenging than the temperate fisheries.

Fisheries stocks, if unmanaged, are exhaustible and can lead to consequences that can impact food and nutritional security, livelihoods and in many cases also the national economy. Thus the need for good estimates arising out of stock assessment is being increasingly felt by coastal nations to ensure that their stocks are not over-harvested and maintained at sustainable levels. Besides meeting such objectives, sound stock assessment figures have also become a necessity for meeting the obligations under international agreements/arrangements to which countries are signatory.

This Training Manual for ‘Stock Assessment of Tropical Fishes’ has been prepared to facilitate the training of 12 Officials nominated by the Government of the People’s Republic of Bangladesh to undertake 14 days Training Course on ‘Stock Assessment of Tropical Fishes’ at the ICAR- Central Marine Fisheries Research Institute (CMFRI), Kochi, India. The Training Course has been organized by the Bay of Bengal Programme Inter-Governmental Organisation (BOBP-IGO), Chennai in pursuance of (i) recommendations made in the Second Meeting of the Joint Working Group between India and Bangladesh on Cooperation in the Field of Fisheries held on 28 - 29 October 2015 in Goa, India and (ii) towards capacity building of fisheries managers/researchers from Bangladesh in stock assessment as a part of the capacity building activities under the World Bank/GEF funded ‘Ocean Partnership Project for Sustainable Fisheries and Biodiversity Conservation – Models for Innovation and Reforms: Bay of Bengal Project (TF 018233).’

The Training Manual is set under 12 chapters dealing with overview of stock assessment, biostatistics, sampling, growth estimation, mortality estimation, virtual population analysis, prediction models, macro-analytical models, productivity based biomass estimates, and recent approaches in stock assessment. In the concluding part, the Training Manual also provides a glossary and suggested reading. It is hoped that this Training Manual will help in promoting an enhanced knowledge of fisheries science and assessment concepts and increased understanding of fisheries management, leading to more informed decision-making, and giving fisheries managers greater confidence as they participate in fisheries management programmes. It is also felt that the utility of this Manual will go much beyond this Training Course and fisheries managers and practitioners of stock assessment in the BOBP-IGO member-countries and also in other regions will be able to make use of the Manual.

The preparation of the Training Manual and conduct of the 14 days Training Course involved many staff members of CMFRI, BOBP-IGO and the Fishery Survey of India and their contribution is greatly acknowledged. Thanks are also due to the Department of Animal Husbandry, Dairying and Fisheries, Ministry of Agriculture and Farmers Welfare, Government of India for assigning the organization and conduct of this Training Course to CMFRI/BOBP-IGO.

Dr A Gopalakrishnan
Director, ICAR-CMFRI

Dr Y S Yadava
Director, BOBP-IGO

Contents

1.0 Overview of Stock Assessment

2.0 Biostatistics

3.0 Sampling

4.0 Growth Estimation

8.0 Virtual Population Analysis

9.0 Prediction Models

12.0 Recent Approaches in Stock Assessment

Chapter 1

Overview of Stock Assessment

1.1

Concept and Objectives of Stock Assessment*

Among all the exploited natural animal resources, fisheries constitute the largest. The magnitude, dynamics and resilience of fish stocks pose great challenge to their assessment as well as management. Fisheries resources are unique at least on three accounts, (i) many species have wide spatial distribution, (ii) several species show wide temporal variations in abundance, and (iii) since the resources cannot be seen visually, gaining an insight into the structure and function of the resources is a challenge. To exploit these resources, to manage and develop the fisheries, and to conserve the fish stocks, it is essential to have accurate information on the stocks. The success of fisheries depends critically on the state of fish stocks. Fish stocks are controlled by several natural factors such as weather, physical, chemical and biological oceanographic conditions and predator-prey relationships. They are also affected by man's activities, and to an increasing extent, by fishing. The assessment of a fish stock must consider all the relevant factors, especially the direct impact of a fishery on a single species. Those concerned with making policy decisions about fisheries must take into account, the state of the fish stocks and the effect of the proposed decisions on the stocks. The science of stock assessment provides scope for extending advice on such aspects.

For instance, if the fish stock assessment studies indicate decline in stocks, fishing regulatory measures such as closed fishing season(s), no fishing zone(s) or restrictions on the expansion of fishing fleet may be contemplated. The stock assessment work would also calculate the amount of increase in the catch, the time required to increase the catch, and the possibility of sustaining the catch, if any one of the measures mentioned above is implemented. Likewise, if the stock suffers from growth overfishing (exploitation of large quantities of juveniles), mesh size regulations can be suggested. The study can determine that if the juveniles are not caught (by increasing the mesh size), and allowed to grow in the sea, the juveniles would grow to a better size, which may result in, say, 20 percent increase in the total catch.

Objectives of stock assessment

Maximum Sustainable Yield (MSY) and Maximum Economic Yield (MEY)

Fisheries resources, although renewable, are exhaustible. The objective of fish stock assessment is to predict changes in the size of the stock and the size of yields as functions of both fishery dependant (fishing effort, etc) as well as fishery independent factors, so that optimum levels of effort and yield could be determined. Figure 1 (see page 8) illustrates the increase in yield with increase in fishing effort up to a certain level, after which, the renewal of stock (reproduction + growth) does not compensate the loss of biomass due to fishing, and hence, further increase in fishing effort leads to decline in yield.

Stock assessment pursues short-term as well as long-term objectives. Assessments for short-term objectives depend to a large extent on the current state of the stock and suggest what is likely to happen to it in the near future, say, next year or the year after. Those pursuing

---

** E. Vivekanandan, National Consultant, World Bank/GEF Ocean Partnership Project, Bay of Bengal Programme Inter-governmental Organisation, 91, St Mary's Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.*

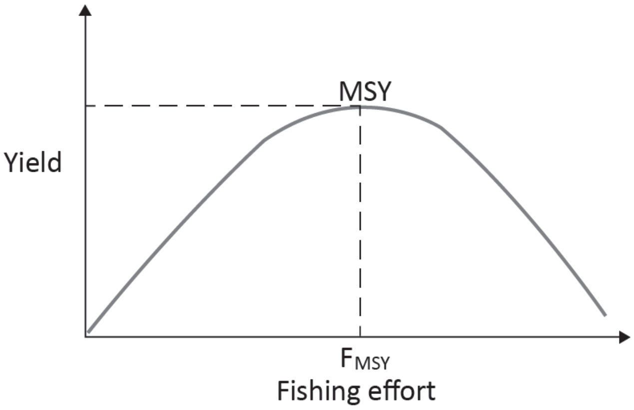


Fig. 1. Concept of Maximum Sustainable Yield (MSY)

long–term objectives (such as estimating the Maximum Sustainable Yield, or in short the MSY), on the other hand, depend little on the present state of the stock, but more on recruitment and growth. While long-term objectives seek to formulate strategies for the long-term management of fisheries, short-term objectives relate to the tactics required for implementation of strategies for which they are concerned, for example, the effort required in the immediate future.

The MSY is a useful tool for describing the fish stocks in relation to exploitation. It explains the fact that more fishing does not necessarily mean more fish and that fishing beyond a certain point, can mean less fish. The fishing effort, which in the long-term gives the highest yield, is indicated as F_{MSY}.(Fig.1)

The *MSY may be defined as the largest average catch, which can continuously be taken from a stock*. The MSY estimate has the important objectives of (i) maximizing the catch, (ii) ensuring that the maximized catch can be sustained, and (iii) interpreting the catch as an approximate measure of the well-being of a fishery. The role of MSY for advocating management measures is as follows: In simple cases, if the abundance of a stock is above the MSY, the stock is considered as underexploited and fishing can be increased; if below, the stock is overexploited and fishing should be restricted; and if the stock abundance is equal to the MSY, the fishing is considered as well maintained.

One criticism of the MSY concept is that the actual yield in a particular year can be subject to considerable variations due to non-fishery causes, such as environmental factors. It is often felt that in the complex modern fisheries situation, MSY is not an adequate tool either to understand the resource or as an index of management success. In recent years, economic and social considerations are receiving increasing attention.

The economic considerations can be seen by converting the curve of Figure 1 into relationship between the cost of fishing and the value of catch (Fig. 2 on page 9). If economic return is considered as the measure of success, fishing at the point of Maximum Economic Yield (MEY) would be the appropriate objective. However, the MEY also ignores several factors such as the environmental parameters, fishermen empowerment, etc. A consensus is now emerging that a single objective of management (MSY or MEY) should not be applied in all situations regardless of changes in the status of the natural resources and in the needs of the society.

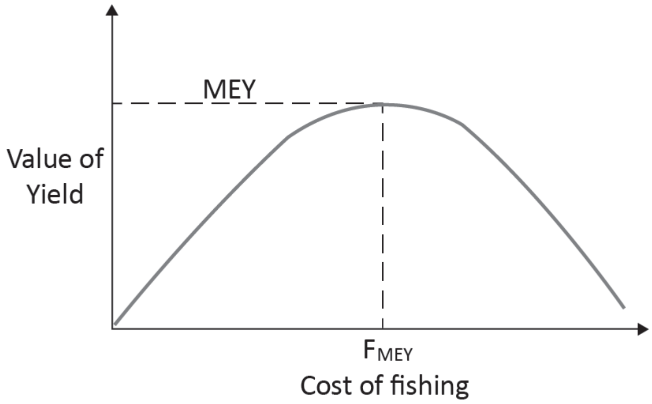


Fig. 2. Concept of Maximum Economic Yield (MEY)

The unit stock

For gaining a proper understanding of the dynamics of the exploited fisheries resources, information on the fundamental units of such resources, referred to as the unit stocks, is essential. The stock is a term applied in a special way in fisheries management. It is a subset of a species characterized by the same growth and mortality parameters, and inhabiting a particular geographical area. The members of a stock share a common gene pool, and hence belong to a particular race within a species. There are distinctions between the fisheries concept of a stock and the biological concepts of a population (Table 1). Capture fisheries research is usually concerned with the stock of fish exploited by a particular fishery, rather than with an individual fish or with the total population of a species. For instance, the Indian mackerel, *Rastrelliger kanagurta* is exploited along the east and west coasts of India; but the biological characteristics like growth, reproduction, mortality, etc of the different stocks of this species differ greatly from one area to another. The stocks, therefore, should be treated and investigated separately for fisheries management purposes.

Table 1. Differences in the concept between a population and a stock

Population		Stock	
i.	Basically a biological concept	I.	Basically a fisheries concept for management purpose
ii.	Breeding unit of a species	ii.	Basic fish sampling unit
iii.	Each member shares a common spawning ground	iii.	Basic fish sampling unit (stock) has production characteristics (like K,Z,F,M, etc) as any other individual of the stock
iv.	In a species' geographic range, these individuals have rapid gene flow among all members of the group	iv.	If geographic clusters of a species differ in the above characteristics, more than one stock is set up for management purpose
v.	Larvae develop in the same geographic area	v.	Stock may be a portion of a population or include more than one population
vi.	Mixing between populations very rare		---

Ideally, a unit stock is a self-contained and self-perpetuating group, with no mixing from outside. There are well-defined geographical limits of spawning and gene exchange within stocks of the non-migratory or short distance migratory species, unlike the highly migratory species. Therefore, it is much easier to identify the stocks of such non-migratory species than those of the species undertaking long distance feeding and spawning migrations, like the tunas.

Russell was one of the first to express the factors in the year 1931, which affect the size of a fish stock by using the following formula:

$$S_2 = S_1 + (R + G) - (C + M) \qquad \text{(Eq. 1.1)}$$

where S_1 = size of the stock at the beginning of the year; S_2 = size of the stock at the end of the year; R = recruitment; G = growth; C = catch; and M = death due to natural causes. If the stock is to be in equilibrium ($S_2 = S_1$), then $R + G = C + M$ or $C = R + G - M$. This is a very simple way of considering the factors that govern stock size. During the last 85 years, fishery biologists have devised methods to estimate the parameters in the equation.

Complexities in stock assessment of tropical fishes

The dynamics of the tropical fish stocks are more complex than those of the temperate stocks. Nevertheless, the methods of fish stock assessment available today are basically those designed for the temperate stocks. Perhaps the most conspicuous difference in stock assessment between the tropical and temperate fishes is in the nature of the basic input data, rather than in the models, as explained here: (i) As age determination is difficult in tropical fishes, length frequencies have to be converted into age frequencies. There are several techniques now available for the conversion of length groups into age groups; (ii) Unlike in the temperate fishes, prolonged spawning makes it extremely difficult to assign seasonality to spawning patterns in tropical fishes. Hence, identification of different cohorts and tracing the length frequency progression of each cohort of tropical fishes has to be carried out under conditions of high subjectivity. The recruitment patterns are also not properly understood at present; (iii) Tropical fishes are characterized by faster growth and shorter life span unlike the temperate fishes. It is more realistic and appropriate to estimate the population parameters of tropical fishes for shorter time units of age, say, one month, and then raise the values to annual basis; and (iv) Another complexity of the tropics is that they support multispecies fisheries where large number of species are caught in the same ground in some important gears like the bottom trawl in almost every haul. Hence, the interspecies relationship and natural mortality under tropical conditions must be very different from those under temperate conditions. As the stock assessment models are tailored to suit the biological characteristics of temperate fisheries, it becomes very difficult to apply them to tropical fisheries. These models are very sensitive to seasonal patterns of recruitment, catchability and mortality. Therefore, appropriate adjustments or modifications are necessary in the existing models to suit tropical fisheries.

Limitations

All the stock assessment and prediction models contain uncertainties in the estimates of specific parameters. This is particularly true for length-based assessment methods, which are mostly applied to tropical species. The length-based assessments depend critically on the estimation of the highly sensitive and variable growth parameters. In a length-based Virtual Population Analysis (VPA), overestimating the K will mean that the time required to grow through a length interval will be underestimated. This implies that the fishing mortality for that interval will be overestimated. In turn, the exploitation rate, which is an input for later analysis, will be overestimated.

Most of the studies on stock assessment suffer from one or other deficiency relating to the estimation of population parameters. A few typical cases are: (i) Estimation of growth and mortality parameters based on samples of larger pelagics collected from selective gears like the large mesh gillnets, which exploit mostly larger fishes; (ii) Growth and mortality estimates of shoaling smaller pelagics such as the oil sardine and the Indian mackerel sampled from the purseseines. The smaller pelagics tend to form schools of fish of the same size; (iii) Estimating the stocks of migratory fishes like the tunas without considering the characteristics of the cohorts and the stocks in the fishing areas from where the samples were collected. It is possible that the samples represented different cohorts and also different stocks. Systematic aerial surveys are conducted regularly in some countries for assessing the stocks of migratory pelagic fishes; (iv) Often, there is bias in the selection of the length frequency modes; (v) Collection of data from an array of gears without properly standardizing the effort; and (vi) Selecting inappropriate methods, especially for the estimation of the total mortality coefficient. Results obtained from discrepant analyses would lead to distorted conclusions on the status of the stocks. When working with mathematical models, it is essential that the fisheries scientists check whether the basic assumptions of the models are fulfilled.

Discards and their effects on stock assessment

An important factor that could not be ignored in stock estimates based on fish landings is the discard. Discards are fish thrown back into the sea because they are too small, of little market value or unmarketable. The problem of discards arises mainly because of mobile gears like the trawl, which catch everything accessible to it in front of its sweep, and there is no space in the fishold of the vessel to accommodate the entire catch. In general, discards are not recorded. The discard factor has assumed alarming proportions in recent years. The trawlers discard almost the entire tiny fishes caught, especially during the early part of the cruise. Ignoring the discard leads to underestimation of the catch as well as the number recruited to the fishery. The exact quantity and nature of discards could be observed only onboard the fishing vessel. The discards comprise large number of species, ranging from gastropods and echinoderms to the juveniles of economically important fishes, the crustaceans and cephalopods. Considerable effort has been made in many countries to account for the effects of discards on stock assessment by adjusting the data on the reported landings and their age composition. It is apprehended that the exploitation of small fishes may affect the food balance in the ecosystem.

Fisheries assessments are highly sophisticated scientific exercises calling for a variety of skills, a sound knowledge on the biology of the species and a good understanding of the fishing operations and the industry. Fisheries scientists often face the problem of lack of information, or even if information is available, it is either inadequate or cannot be processed in time. This is because the stock assessment studies have to rely on the quantity and quality of the data and knowledge, which depend to a large extent on the cooperation provided by the fishing industry.

Further, the resource system itself varies with time in such a way that the basic scientific conclusions of today may have to be modified, often radically, within a short time in the future. For data analysis one has to wait. By the time the catch and biological data become available and put to analysis, the assessments get outdated by several years. The importance of such delays cannot be underestimated, considering the need for timely assessments to understand the status and resilience of the tropical multispecies fish stocks in withstanding overfishing over a good deal of time because of their characteristic multiple spawning frequency and fast growth. Most of the fast growing and short-lived tropical fishes, penaeid prawns and cephalopods have high potential increase rates, *vis-a-vis* rapid decline rates within a short duration. Furthermore, the interactions between the trophic levels are too great that one cannot expect consistence in the stocks and in the stock estimates over the time-scale. The existing single species stock assessment models are often found inadequate to accommodate the resilience of tropical stocks.

Moreover, any change in the exploitation pattern of commercial fisheries, like the introduction of a new gear or a change in the mesh size may considerably alter the assessment estimates. Change of fishing areas from time to time is another major causative factor for the tentativeness and inaccuracy of the estimates. With the induction of more large vessels and consequent extension of fishing to deeper grounds all over the Exclusive Economic Zone (EEZ), this factor has assumed greater importance in stock assessment.

Stock assessment models provide estimates of the optimum yield, usually taking into consideration only the biological factors. These models consider that the environment is invariable, which is not true. In addition to the environmental factors, the economic factors, such as the escalation of operational cost and fluctuations in the value of the catch, also play an important role in arriving at appropriate management decisions. It is necessary that each of these considerations and their alternatives are investigated and addressed thoroughly.

If fish stock assessments and predictions fail, the ways by which the stocks respond to exploitation provide an opportunity to gain an understanding of the status of the fishery. Changes in catch per unit effort (CPUE), mean length in the catch, length at first maturity and other biological characteristics are the responses of fish stocks to exploitation. For instance, as the intensity of fishing increases, there is a progressive decrease in the abundance of the stock, the mean length of the fish in the catch and the length at first maturity. By continuously monitoring these changes, a clue, though of limited value, could be gained on the effects of fishing pressure on the stock. There are many responses, such as drastic fluctuations in the catch, which do not provide specific answers. The factors mentioned in Table 2, though not exhaustive, reflect the response of stocks to fishing pressure or to environmental changes.

Table 2. Indicators of decline in fishery resources and the causes

Nature of decline	Indicators	Causes
Decrease in catch	Decrease in catch rate	Environment, fishing competition
	Change in species composition	Environment, changes in gear & area of fishing, market preference, etc.
Decrease in recruitment	Sudden increase in mean age/ length	High vulnerability to fishing
	Spawners exploitation	Target fishing
F = M or F>M	Reduction in mean age/length	Environment, fishing pressure exploitation of juveniles
Deviations from normal pattern	Changes in spawning pattern	Environment, biological
	Changes in length at first maturity	Environment, biological
	Changes in fecundity	Environment, biological
	Changes in size composition	Fishing, market preference

Problems of specific fisheries

Small pelagics

Small shoaling pelagic stocks usually do not fit well into the conventional population dynamics models because of their highly variable characteristics, thus making their assessment and management difficult and uncertain. Considering the uncertainties in the small pelagic stocks, fishing for the small pelagics is a high risk activity as compared to fishing for the more reliable and robust demersal stocks, including the penaeid prawns for which the conventional stock assessment models work much better. There are at least 5 major reasons for the uncertainties involved in the assessment of the small pelagics. These are: (i) The highly variable nature of the small shoaling pelagics is now fully recognized. The fluctuations in the catches of oil sardine and the Indian mackerel in India are good examples; (ii) The small pelagics undergo fast growth and live a short span of life. Any decrease in recruitment will be followed quickly by a decrease in the stock size. When estimates of yearclass strength are available from the catch data, the yearclass is often out of the fishery. The quick changes in the stock size necessitate decision-making at very short notice; (iii) Stock sizes and catches depend almost completely on one or two yearclasses; (iv) The catch rate may not be a reliable indicator of the abundance since the catch per haul or catch per hour will be high if one school is sighted. In this case, high catch rate is possible even if the stock has declined to a very low level. Catchability may be more reliable if the effort includes searching time; and (v) A fish shoal is normally made of fish of about the same size and hence, it is difficult to get a representative sample of the length composition of the population. These problems are likely to affect the length-based methods for estimating growth, mortality and other population parameters.

For these highly variable stocks, stochastic modeling could be followed. Stochastic modeling is one in which the inputs and results are given as a range of possible realizations accounting for the natural randomness. This approach requires adequate reliable data so that acceptable confidence levels could be obtained for model predictions. Acoustic survey also appears to be suitable and may provide reliable indices of abundance. The relation between the biomass and the acoustic properties is still uncertain, and species identification generally requires direct biological sampling, usually with nets. Acoustic surveys are useful when used in a long-term monitoring programme, but may be misleading when employed as a one-time estimate of absolute biomass. Different vessels and equipments could give different acoustic measurements and cross calibration is necessary for comparing the survey results.

Tunas

As in the case of the small shoaling pelagics, the available methods are often insufficient when applied to large migratory pelagics, like the tunas. When stocks are not vulnerable to fishing due to horizontal or vertical migrations to areas not covered by fishing vessels during parts of their life span, it is not possible to sample the entire life stages of such stocks. This situation may lead to gaps in the samples for shorter or longer periods and in such cases, it is likely that the data are misinterpreted. Hence, the length-based methods such as the modal progression analysis and catch curve analysis need to consider the migration of the species. This problem could be overcome if the migratory routes are known. Then the stock assessment exercises could be performed based on the samples collected at the designated areas representing the migratory routes during the specific periods. There are several ways of acquiring knowledge on the possible migratory route. Landing records of various types of fishing vessels from the contiguous areas could provide some clues. The classical approach is to undertake a tagging programme. Sophisticated acoustic and radio and satellite tags have been developed which allow continuous observations on the movements of single fish. Latest developments include tags that are released at a preset moment; they pop up to the water surface, and transmit the position of the fish to the satellite.

In recent years, advanced methods like MULTI FAN-CL assessment and stock synthesis are being used for tunas in the Indian Ocean, using tag-recovery catch rate and environmental data.

Conclusion

The most common use of the results of stock assessment is to provide advice to the fisheries administrators about the development and management of fisheries. In spite of the limitations and uncertainties in stock estimates, as mentioned in the preceding sections, it is highly desirable that advice is given on the basis of stock estimates. It is also desirable that all uncertainties in the estimates, especially the uncertainties regarding recruitment, which could have a great effect on the fisheries, are informed along with the advice. This would in no way diminish the value of the advice. A reasonable professional estimate of future trends appears acceptable than lack of any information.

***

1.2

Fish Stock Assessment and Management*

It is a known fact that when resources are owned by a large group of people, no one takes responsibility for maintaining the resource. In such a situation the tendency of each person is to use the resource to the maximum extent without bothering to conserve the resource or invest in the resource because those who do not contribute would also get the benefit. This is generally known as “the tragedy of the commons”. To prevent “the tragedy of the commons”, the most common property resources are held in trust and managed for the people by the government. Fish living in public waters are a common property resource. The government has the responsibility of managing this resource for the benefit of all citizens, even those who do not fish.

Fish is a renewable resource which has to be managed using some basic principles for the benefit of the citizens, generation after generation. Renewable resources like fish are living things that replenish themselves naturally and can be harvested, within limits, on a continuing basis without being eliminated. The scientific principles behind this renewability are well known and provide the basis for its management. As a kind of biological insurance against natural calamities, all living organisms produce more offspring than that is necessary to survive to become adult. Most individual fish produce tens of thousands to millions of eggs, most of which do not survive to become juveniles, and even fewer live to become adults.

In a habitat of unfished population, the population biomass increases over time due to reproduction and growth of individual fish and it will approach a maximum beyond which the biomass cannot grow, which is also known as the carrying capacity of the habitat. Compared to a fished population, the unfished population will have lots of older, larger fish that dominate the habitat, limiting the survival of the young fish produced each year to become old fish. When fishing begins, many large older fish are removed, which brings down the biomass below the carrying capacity and increases the chances of survival for smaller, younger fish. This extra production of the younger fish together with the effects of harvesting fish can result in sustainable production. The unfished population is relatively stable population with moderate production and the fished population is dynamic with a higher turnover of individual fish as the older fish are replaced by younger, faster growing fish. Some of this new production must be allowed to survive and reproduce to maintain the population. The remaining is the surplus production, which is available for harvest.

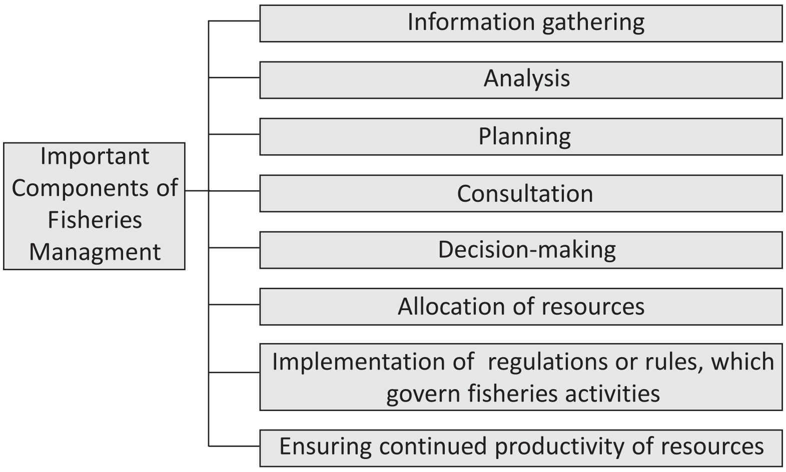
Fisheries Management

The basic goal of fisheries management is to find out the quantity of fish that can be harvested (optimum yield) in a sustainable manner, which may be subjected to changes based on political, economic, and social considerations. Highly conservative management can result in loss of fish production due to under-harvesting, while too liberal or no management can result in population depletion due to over-harvesting. Unlike mineral resources, if the fishery resources

---

* *T.V Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

are well managed, their duration is unlimited. The fundamental basis for conservation and management of fisheries resources is its biological characteristics. Fisheries management draws on fisheries science in order to find ways to protect fishery resources so that sustainable exploitation is possible. The important components of fisheries management are depicted in the following infographic.



The reality is that fisheries management is not about managing fish but about managing the associated people and industry through regulations. The success of fisheries management is associated with human factors like reaction of fishermen and the implications of management regulations for the stakeholders. The Food and Agriculture Organization (FAO) of the United Nations advises that the ‘precautionary approach’ should be applied when "ecosystem resilience and human impact are difficult to forecast and hard to distinguish from natural changes”. It suggests that when an action risks harm, it should not be proceeded until it can be scientifically proven to be safe.

Fisheries Management History

Even 700 years ago there were restrictions on fishing imposed by Maori tribes of New Zealand, against taking more than what could be eaten. Also, as an offering to their sea god ‘Tangaroa’, they released back to the sea the first fish caught. Attempts started in the 18th century to regulate the north Norwegian fishery, which resulted in forming a law in 1816 on the Lofoten fishery, known as territorial use rights. The first resource protection based fisheries management by a government was for the North European fisheries, after the first over fishing conference held in London in 1936. Ray Beverton and Sidney Holt, the two British fisheries researchers published a seminal work on North Sea commercial fisheries dynamics In 1957 and this became the theoretical platform for North European fisheries management schemes in 1960. In south-west coast of India, the oil sardine fishery had totally collapsed during the early 1940s and the oil sardine fishery was officially closed for 5 years (1943-47) by the erstwhile British Government. Many countries have set up government departments for controlling aspects of fisheries within their Exclusive Economic Zone (EEZ).

Traditional management practices aim to reduce the number of old, slow-growing fish, leaving more room and resources for younger, faster-growing fish. The assumption was that younger spawners would produce plenty of viable larvae. However, research revealed that large, elderly females are far more important than younger fish in maintaining productive fisheries. The larvae produced by these older maternal fish grow faster, survive better, and are much more likely to survive than the offsprings of younger fish. Failure to account for the role of older fish may help explain recent collapses of several fisheries. One way to prevent such collapses is to establish marine reserves, where fishing is not allowed and fish populations age naturally. Some of the fisheries management mechanisms are listed in the above infographic.

Fisheries Management Mechanisms

- Prohibiting devices such as bows and arrows, spears, or firearms.
- Setting minimum mesh sizes.
- Limiting the average potential catch of each vessel in the fleet.
- Limiting the fishing season.
- Implementing Marine Protected Areas.
- Restricting the number of simultaneous fishing vessels.
- Limiting a vessel's average operational intensity per unit time at sea.
- Limiting average time at sea.

In most countries fisheries management rules are based on the 1995 FAO Code of Conduct for Responsible Fisheries (CCRF), agreed by the FAO member-countries in 1995. International agreements are required to regulate fisheries in international waters. The desire for such agreement led to three conferences on the 'Law of the Sea' and to the treaty known as the United Nations Convention on the Law of the Sea (UNCLOS) and derived the concept of an EEZ, extending up to 200 nautical miles (370 km) from a nation's coast and allocated certain sovereign rights and responsibilities for resource management to individual countries.

Fish Stock Assessment

Living resources are limited but renewable. The term stock in fisheries refers to genetically distinct populations within a species that are unique biological identities. Fish stock assessment can be described as the search for the exploitation level that in the long run will give maximum yield from the fishery. The aim of fish stock assessment is for a fishing strategy that gives the highest steady yield year after year. The primary objective of fish stock assessment is to provide advice on the optimum exploitation of aquatic living resources such as fish and shrimp. Social well-being of the fishermen and the economic success of the fishing industry are based on the status of fish stocks. Different regulatory measures imposed on the fishery are closed seasons, marine protected areas, protected species, ban on destructive fishing gear and methods, mesh size regulations, use of excluder devices, etc.

Catch Quotas

Systems that use individual transferable quotas (ITQ), also called individual fishing quota limit the total catch and allocate shares of that quota among the fishers who work in that fishery. Fishers can buy/sell/trade shares as they choose. A study in 2008 provided strong evidence that ITQs can help to prevent fishery collapse and even restore fisheries that appear to be in decline.

Ecosystem-based Fisheries

Tony Pitcher and Daniel Pauly proposed that rebuilding ecosystems, and not sustainability *per se*, should be the goal of fishery management. According to them, sustainability is a deceptive goal because human harvesting of fish leads to a progressive simplification of ecosystems in favour of smaller, high turnover, lower trophic level fish species that are adapted to withstand disturbance and habitat degradation.

Ecopath with Ecosim

Ecopath with Ecosim is an ecosystem modeling software suite, which is an initiative by the National Oceanic and Atmospheric Administration (NOAA), led by Jeffrey Polovina and later developed at the University of British Columbia. In 2007, it was named as one of the ten biggest scientific breakthroughs in NOAA's 2000 year history. This software revolutionized scientists' ability to understand complex marine ecosystems world wide and is used widely in fisheries management as a tool for modeling and visualizing complex relationships that exist in real world marine ecosystems.

***

Chapter 2

Biostatistics

2.1

Measures of Central Tendency[*]

Description of a variable usually begins with the specification of its single most representative value, often called the ‘measure of central tendency’. The best way to reduce a set of data and still retain part of the information is to summarize the set with a single value. A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. Measures of central tendency are sometimes called ‘measures of central location’ or ‘summary statistics’. Measures of central tendency are measures of the location of the middle or the center of a distribution. The definition of “middle” or “center” is purposely left somewhat vague so that the term “central tendency” can refer to a wide variety of measures. There are several measures for this statistic.

Measures of Central Tendency

- Arithmetic mean
- Median
- Mode
- Weighted mean
- Geometric mean
- Harmonic mean

Arithmetic mean

The arithmetic mean of a set of values is the quantity commonly called the mean or the average. For a data set, the mean is the sum of the values divided by the number of values. The mean of a set of numbers x_1, x_2, ...,x_n is typically denoted by $\underline{x}$, pronounced “*x* bar”.

$$\text{Arithmetic mean} = \bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} \qquad \text{Or} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Arithmetic mean from a grouped data

i) Discrete frequency distribution

Data arising from organising *n* observed values into a smaller number of disjoint groups of values, and counting the frequency of each group is often presented as a frequency table. In this case the values of the variable are multiplied by their respective frequencies and this total is then divided by the total number of frequencies.

Arithmetic mean, $$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

---

[*] *Somy Kuriakose, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

where $x_1, x_2, \ldots x_n$ are values of the variable x and $f_1, f_2, \ldots f_n$ are their corresponding frequencies.

ii) Continuous frequency distribution

We take mid values of each class as representative of that class, multiply this mid value by their corresponding frequencies, total these products and divide by the total number of items. If $x_1, x_2 \ldots x_n$ represent the mid values of classes and $f_1, f_2, \ldots f_n$ the frequencies, then

Arithmetic mean= $$\frac{f_1 x_1 + f_2 x_2 + \ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

where N= $\sum_{i=1}^{i=n} f_i$

The mean is valid only for interval data or ratio data. Since it uses the values of all the data points in the population or sample, the mean is influenced by outliers that may be at the extremes of the data set. The mean uses all the observations and each observation affects the mean. Even though the mean is sensitive to extreme values (*i.e.* extremely large or small data can cause the mean to be pulled toward the extreme data), it is still the most widely used measure of location. This is due to the fact that the mean has valuable mathematical properties that makes it convenient for use with inferential statistics analysis. For example, the sum of the deviations of the numbers in a set of data from the mean is zero, and the sum of the squared deviations of the numbers in a set of data from the mean is the minimum value. The following are the merits and demerits of arithmetic mean.

Merits and demerits of arithmetic mean

Merits

- It is rigidly defined.
- It is easy to calculate and simple to follow.
- It is based on all the observations.
- It is determined for almost every kind of data.
- It is finite and not indefinite.
- It is readily put to algebraic treatment.
- It is least affected by fluctuations of sampling.
- It is easy to calculate.

Demerits

- The arithmetic mean is highly affected by extreme values.
- It is not an approprite average for highly skewed distributions.
- It cannot be computed accurately if any item is missing.
- The mean sometimes does not coincide with any of the observed value.

Median

Median is the value in the middle of the dataset, when the data points are arranged from smallest to largest. If there is an odd number of data points, then just arrange them in ascending or descending order and take the middle value. If there are even number of data points, you will need to take the average of the two middle values. Hence median is determined by sorting the dataset from lowest to highest values and taking the data point in the middle of the sequence. There is an equal number of points above and below the median.

For example, in the dataset {1,2,3,4,5} the median is 3; there are two data points greater than this value and two data points less than this value. In this case, the median is equal to the mean. But consider the data set {1,2,3,4,10}. In this dataset, the median still is three, but the mean is equal to 4. If there are even number of data points in the set, then there is no single point at the middle and the median is calculated by taking the mean of the two middle points.

Calculation of median in a grouped data

i) Discrete series

In this case also, data should be arranged in ascending or descending order of magnitude and find out the cumulative frequencies. Now find out the value of (n+1/2)th item. It can be found by first locating the cumulative frequency, which is equal to (n+1/2) and then determine the value corresponding to it. This will be the value of median.

ii) Continuous series

For computing the value of the median in a continuous series, first determine the particular class in which the value of the median lies. Use N/2 as the rank of Median where N= total frequency. Hence it is N/2, which will divide the area of the curve into two parts. The following formula is used for determining the exact value of the median.

$$\text{Median}= l + \frac{(\frac{N}{2} - m)*c}{f}$$

where N= $\sum f_i$ = Total frequency, l- the lower limit of the median class, m- cumulative frequency up to the median class, f- frequency of the median class and c- class width.

The median can be determined for ordinal data as well as interval and ratio data. Unlike the mean, the median is not influenced by outliers at the extremes of the dataset. Generally, the median provides a better measure of location than the mean when there are some extremely large or small observations (*i.e.* when the data are skewed to the right or to the left). For this reason, the median often is used when there are a few extreme values that could greatly influence the mean and distort what might be considered typical. It may be noted that if the median is less than the mean, the dataset is skewed to the right. If the median is greater than the mean, the dataset is skewed to the left. Median does not have important mathematical properties for use in future calculations.

Merits and demerits of median

Merits

- Median is rigidly defined.
- It is simple to understand and easy to calculate.
- Median is not affected by extreme observations.
- Median can be computed even for open-end classes.
- Median can sometimes be located by inspection.
- Median value is real value and is a better representative value of the series compared to arithmetic mean average, the value of which may not exist in the series at all.
- Median can be obtained graphically.
- Median is only the average to be used while dealing with qualitative characteristics such as intelligence, beauty, etc.

Demerits

- Arrangement of data according to magnitude is necessary.
- Median is not based on all observations.
- For an ungrouped data, if the number of observation is even, median cannot be determined exactly.
- Median is not suitable for further mathematical treatment.
- For a small size sample, median is affected by fluctuation of sampling.

Mode

Mode is the most common value or most frequently occurring value in the dataset. For finding the mode, just look at the data, count how many of each value you have, and select the data point that shows up most frequently. If no value occurs more than once, then there is no mode. If two values occur as frequently as each other and more frequently than any other, then there are two modes. In the same way, there could also be more than two modes.

For example, in the data set {1,2,3,4,4}, the mode is equal to 4. A dataset can have more than a single mode, in which case it is multimodal. In the dataset {1,1,2,3,3} there are two modes: 1 and 3.

The mode can be very useful for dealing with categorical data. The mode also can be used with ordinal, interval, and ratio data. However, in interval and ratio scales, the data may be spread thinly with no data points having the same value. In such cases, the mode may not exist or may not be very meaningful. Following are the various merits and demerits of mode:

Merits and demerits of mode

Merits

- Mode is very simple measure of central tendency. Because of its simplicity, it is a very popular measure of the central tendency.
- Compared to mean, mode is less affected by marginal values in the series since it is determined only by the value with highest frequencies.
- Mode can be located graphically, with the help of histogram.
- Mode is that value which occurs most frequently in the series. Accordingly, mode is the best representative value of the series.
- The calculation of mode does not require knowledge of all the items and frequencies of a distribution.

Demerits

- Mode is an uncertain and vague measure of the central tendency.
- Unlike mean, mode is not capable of further algebraic treatment.
- If frequencies of all items are identical, it is difficult to identify the modal value.
- Calculation of mode involves cumbersome procedure of grouping the data.
- It ignores extreme marginal frequencies and is not a representative value of all the items in a series.

Weighted mean

When two or more means are combined to develop an aggregate mean, the influence of each mean must be weighted by the number of cases in its sub-group.

$$\overline{X}_w = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2 + n_3 \overline{X}_3}{n_1 + n_2 + n_3}$$

Example: $\overline{X}_1 = 12, n = 10$ $\overline{X}_2 = 14, n = 15$ $\overline{X}_3 = 18, n = 40$

Weighted mean =

$$\frac{10(12)+15(14)+40(18)}{10+15+40} = 16.2$$

Geometric mean (GM)

The geometric mean is an average that is useful for sets of positive numbers that are interpreted according to their product and not their sum (as is the case with the arithmetic mean) *e.g.* rates of growth.

$$\bar{x} = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$$

For example, the geometric mean of six values: 34, 27, 45, 55, 22, 34 is:

$$(34.27.45.55.22.34)^{1/6} = 1{,}699,\ 493,\ 400^{1/6} \approx 34.545.$$

Harmonic mean (HM)

The harmonic mean is an average, which is useful for sets of numbers which are defined in relation to some unit, for example speed (distance per unit of time).

$$\bar{x} = n.\ \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1}$$

For example, the harmonic mean of the six values: 34, 27, 45, 55, 22, and 34 is:

$$\frac{6}{\frac{1}{34}\quad \frac{1}{27}\quad \frac{1}{45}\quad \frac{1}{55}\quad \frac{1}{22}\quad \frac{1}{34}} = \frac{60588}{1835} \approx 33.0179836$$

Relationship between AM, GM, and HM

AM, GM, and HM satisfy these inequalities:

$$AM > GM > HM$$

Equality holds only when all the elements of the given sample are equal.

The mean (often called the average) is the most common measure of central tendency, but there are others, such as the median and the mode. The mean, median and mode are all valid measures of central tendency, but under different conditions some measures of central tendency become more appropriate to use than others.

***

Measures of Dispersion*

Measures of variation describe the spread or scatteredness of a set of data. They are also known as measures of dispersion or measures of spread. Measures of variation determine the range of distribution relative to the measures of central tendency. Measures of average, such as the mean and median, represent the typical value for a dataset. Within the dataset, the actual values usually differ from one another and from the average value itself. The extent to which mean and median are good representatives of the values in the original dataset depends upon the variability or dispersion in the original data. Where the measures of central tendency are specific data points, measures of variation are lengths between various points within the distribution. It provide us with a summary of how much the points in our data set vary, *e.g.* how spread out they are or how volatile they are. Measures of variation together with measures of central tendency are important for identifying key features of a sample to better understand the population from which the sample comes from. Datasets are said to have high dispersion when they contain values considerably higher and lower than the mean value.

The most common measures of variation are:

1. Range
2. Quartile déviation or semi-interquartile range
3. Mean deviation
4. Variance
5. Standard deviation
6. The Coefficient of Variation

Range

The range is the distance between the lowest data point and the highest data point. In other words, it is the difference between the highest value and the lowest value.

Range = Highest value – lowest value

The range is the simplest measure of variation. Since the range only uses the largest and smallest values, it is greatly affected by extreme values, that is, it is not resistant to change.

Example: Find the range of the following dataset:

8 4 13 35 63

Lowest value = 4 , Highest value =63

Range = Highest value - Lowest value

= 63 – 4 = 59

The range is simple to compute and is useful when you wish to evaluate the whole of a dataset. It is useful for showing the spread within a dataset and for comparing the spread between similar datasets.

---

* *Somy Kuriakose, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Since the range is based solely on the two most extreme values within the dataset, if one of them is either exceptionally high or low (sometimes referred to as outlier), it will result in a range that is not typical of the variability within the dataset. The range does not really indicate how the scores are concentrated along the distribution. The range only involves the smallest and largest numbers, and is affected by extreme data values or outliers. In order to reduce the problems caused by outliers in a dataset, the inter-quartile range is often calculated instead of the range.

Quartile deviation or the semi inter-quartile range

The inter-quartile range is a measure that indicates the extent to which the central 50 percent of values within the dataset are dispersed. If the sample is ranked in ascending order of magnitude two values of *x* may be found, the first of which is exceeded by 75 percent of the sample, the second by 25 percent; their difference is the interquartile range. It is based upon and related to the median. In the same way that the median divides a dataset into two halves, it can be further divided into quarters by identifying the upper and lower quartiles. The lower quartile Q1 is found one quarter of the way along a dataset when the values have been arranged in order of magnitude; the upper quartile Q3 is found three quarters along the dataset. Therefore, the upper quartile lies half way between the median and the highest value in the dataset whilst the lower quartile lies halfway between the median and the lowest value in the dataset. Between Q1 and Q3, there are half the total number of items. Q3 - Q1 affords a convenient and often a good indicator of the absolute variability. Usually one half of the Q3 - Q1 is used and given the name semi-interquartile range or quartile deviation.

Quartile deviation $= \frac{Q3 - Q1}{2}$

The relative measure of quartile deviation is known as the coefficient of Q.D.

Coefficient of Q.D. $= \frac{\frac{Q3 - Q1}{2}}{\frac{Q3 + Q1}{2}} = \frac{Q3 - Q1}{Q3 + Q1}$

The larger the semi-interquartile range, the larger the spread of the central half of the data. Thus the semi-interquartile range provides a measure of spread and indicates how closely the data are clustered around the median.

Mean deviation

Mean deviation is the average of the absolute values of the deviation scores; that is, mean deviation is the average distance between the mean and the data points. It is calculated as:

$$\sum \frac{|X - \mathrm{X_i}|}{n}$$

Closely related to the measure of mean deviation is the measure of *variance*.

Variance

The variance is the most commonly accepted measure of variation. It represents the average of the squared deviations about the mean. Variance also indicates a relationship between the mean of distribution and the data points; it is determined by averaging the sum of the squared deviations. Squaring the differences instead of taking the absolute values allows for greater flexibility in calculating further algebraic manipulations of the data. It is the average of the squared deviations between the individual scores and the mean. The larger the variance the more variability will be there among the scores. When comparing two samples with the same unit of measurement (age), the variances are comparable even though the sample sizes may be different. Generally, however, smaller samples have greater variability among the scores than larger samples.

The average deviation from the mean is: $Ave.\ Dev = \frac{\sum (x - \mu)}{N}$

The problem is that this summation is always zero. Therefore, the average deviation will always be zero. That is why the average deviation is never used. To keep it from being zero, the deviation from the mean is squared and called the "squared deviation from the mean". This "average squared deviation from the mean" is called the variance. It is in squared units. The formula for variance depends on whether you are working with a population or sample:

The formula for the variance in a population is: $\sigma^2 = \frac{\sum (x - \mu)^2}{N}$ where ¼ is the mean and N is the number of scores.

When the variance is computed in a sample, the statistic $s^2 = \frac{\sum (X - M)^2}{N - 1}$

where M is the mean of the sample and gives an unbiased estimate of σ^2.

For example, for the numbers 1, 2, and 3, the mean is 2 and the variance is:

$$\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = 0.667$$

Standard Deviation (SD)

Standard deviation is the most familiar, important and widely used measure of variation. It is a significant measure for making comparison of variability between two or more sets of data in terms of their distance from the mean.

The standard deviation is the square root of the variance. It is denoted by σ and is:

computed as $\sigma^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

The standard deviation has proven to be an extremely useful measure of spread in part because it is mathematically tractable. Many formulae in inferential statistics use the standard deviation. It possesses the majority of the properties which are desirable in a measure of dispersion and is based on all observations. Because of these merits, the SD should always be used as the measure of dispersion unless there is some definite reason for selecting any other measure of dispersion.

Coefficient of variation

The coefficient of variation is the ratio of the sample standard deviation to the sample mean. It is calculated as:

$$\text{Coefficient of variation (C.V.)} = \frac{\sigma}{\bar{x}} * 100$$

It expresses the standard deviation as a percentage of the mean, so it can be used to compare the variability of two or more distributions even when the observations are expressed in different units of measurement. The coefficient of variation is a dimensionless number. Therefore, when comparing between datasets with different units or widely different means, one should use the coefficient of variation for comparison instead of the standard deviation. A standard application of the coefficient of variation is to characterize the variability of geographic variables over space or time. Coefficient of variation is particularly applied to characterize the interannual variability of climate variables or biophysical variables. When coefficient of variable is lesser in the data, it is said to be more consistent or has less variability. On the other hand, the series having higher coefficient of variable has higher degree of variability or lesser consistency.

When the mean value is close to zero, the coefficient of variation will approach infinity and is hence sensitive to small changes in the mean. Unlike the standard deviation, it cannot be used to construct confidence intervals for the mean.

***

2.3

Correlation[*]

Introduction

Correlation is a statistical technique used to assess a possible linear association between two continuous variables. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. If there is any relation between two variables *i.e.* when one variable changes, the other also changes in the same or in the opposite direction, we say that the two variables are correlated. If it is proved that in a large number of instances two variables, always tend to fluctuate in the same or in the opposite direction, then it is established that a relationship exists between the variables. This is called a “Correlation.” It means the study of existence, magnitude and direction of the relation between two or more variables.

Types of correlation

1. Positive and negative correlation

A) If two variables change in the same direction (*i.e.* if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation.

B) If two variables change in the opposite direction (*i.e.* if one increases, the other decreases and *vice versa*), then the correlation is called a negative correlation.

2. Linear and non-linear correlation

If the ratio of change between two variables is uniform, then the correlation is said to be linear. If the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable, then the correlation is said to be non-linear or curvilinear. The nature of the graph gives us the idea of linear type of correlation between two variables. If the graph is in a straight line, the correlation is called a **linear correlation** and if the graph is not in a straight line, the correlation is **non-linear** or **curvi-linear.**

Degrees of correlation

Through the coefficient of correlation, we can measure the degree or extent of the correlation between two variables. On the basis of the coefficient of correlation, we can also determine whether the correlation is positive or negative and also its degree or extent.

1. **Perfect correlation:** If two variables change in the same direction and in the same proportion, the correlation between the two is **perfect positive**. According to Karl Pearson, the coefficient of correlation in this case is +1. On the other hand if the variables change in the opposite direction and in the same proportion, the correlation is **perfect negative**. Its coefficient of correlation is -1. In practice we rarely come across such types of correlations.

---

[*] *Somy Kuriakose, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

2. **Absence of correlation:** If two series of two variables exhibit no relations between them or change in one variable does not lead to a change in the other variable, then we can firmly say that there is **no correlation** or **absurd correlation** between the two variables. In such a case the coefficient of correlation is 0.
3. **Limited degrees of correlation:** If two variables are not perfectly correlated or if there is a perfect absence of correlation, then we term the correlation as Limited correlation. It may be positive, negative or zero but lies within the limits of 1.

Correlations can be placed under three categories: high degree, moderate degree or low degree. The following table reveals the effect (or degree) of coefficient or correlation.

Degrees	Positive	Negative
Absence of correlation	Zero	0
Perfect correlation	+ 1	-1
High degree	+ 0.75 to + 1	- 0.75 to -1
Moderate degree	+ 0.25 to + 0.75	- 0.25 to - 0.75
Low degree	0 to 0.25	0 to - 0.25

Methods of determining correlation

The following are the most commonly used methods of determining correlation:

(1) Scatter Plot

(2) Kar Pearson's coefficient of correlation

Scatter Plot (Scatter diagram or dot diagram)

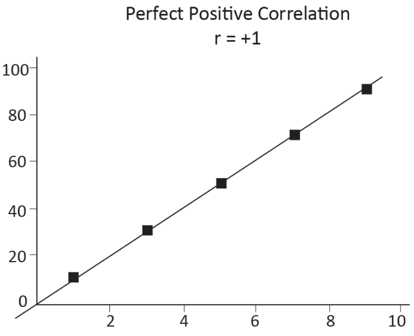
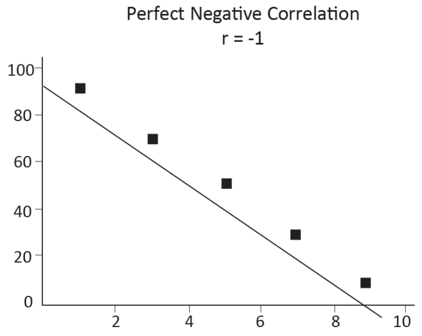
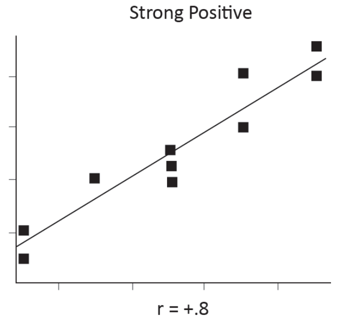
The scatter diagram may be described as the diagram which helps us to visualize the relationship between two phenomena. This is the simplest method for finding out whether there is any relationship between two variables. In this method, the values of the two variables are plotted on a graph paper. One is taken along the x-axis and the other along the y-axis. By plotting the data, we get points on the graph which are generally scattered and hence the name 'Scatter Plot'. The manner in which these points are scattered suggests the degree and the direction of correlation. The greater the scatter of the points on the chart, the lesser is the relationship between the two variables. The more closely the points come to a straight line, the higher the degree of relationship. The degree of correlation is denoted by ' r ' and its direction is given by the signs positive and negative. Scatter diagrams will generally show one of six possible correlations between the variables:

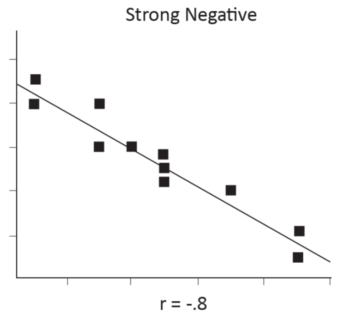
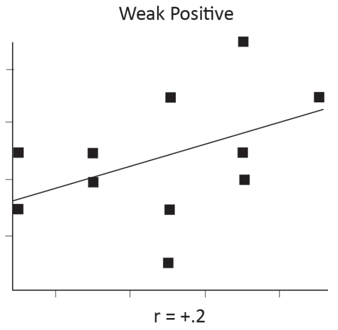
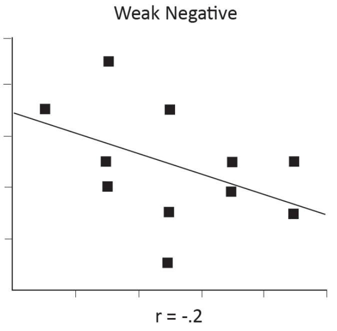
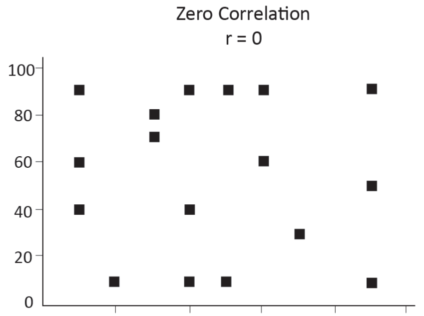
- *Strong positive correlation :*The value of Y clearly increases as the value of X increases.
- *Strong negative correlation:* The value of Y clearly decreases as the value of X increases.

- *Weak positive correlation :*The value of Y increases slightly as the value of X increases.
- *Weak negative correlation:* The value of Y decreases slightly as the value of X increases.
- *Complex correlation:* The value of Y seems to be related to the value of X, but the relationship is not easily determined.
- *No correlation:* There is no demonstrated connection between the two variables.

Though this method is simple and provides a rough idea about the existence and the degree of correlation, it is not reliable. As it is not a mathematical method, it cannot measure the degree of correlation.

Illustrations

1	If all points lie on a rising straight line the correlation is perfectly positive and r = +1	
2.	If all points lie on a falling straight line the correlation is perfectly negative and r = -1	
3.	If the points lie in narrow strip, rising upwards, the correlation is high degree of positive	

4	If the points lie in a narrow strip, falling downwards, the correlation is high degree of negative	
5	If the points are spread widely over a broad strip, rising upwards, the correlation is low degree positive	
6	If the points are spread widely over a broad strip, falling downward, the correlation is low degree negative	
7	If the points are spread (scattered) without any specific pattern, the correlation is absent. *i.e.* r = 0.	

Karl Pearson's coefficient of correlation

The most widely-used type of correlation coefficient is *Pearson r*, also called *linear* or *product- moment* correlation. It gives the numerical expression for the measure of correlation. The value of ' r ' gives the magnitude of correlation and the sign denotes its direction. It is defined as:

$$r = \frac{\sum XY}{n\sigma_x\sigma_y}$$

where

$X = (X_i - \overline{X}),\ Y = (Y_i - \overline{Y}),$

$\sigma_x = s.d.of\ X,\ \sigma_y = s.d.of\ Y$

and n is the number of pairs of observations.

Properties of correlation coefficient

1. The value of correlation does not depend on the specific measurement units used; for example, the correlation between height and weight will be identical regardless of whether *inches* and *pounds*, or *centimeters* and *kilograms* are used as measurement units.
2. The value of correlation coefficient lies between -1 and +1; -1 means perfect negative linear correlation and +1 means perfect positive linear correlation.
3. The correlation coefficient ‘r’ only measures the strength of a linear relationship. There are other kinds of relationships besides linear.
4. If the two variables are independent, then the value of the correlation coefficient is zero. If the value of the correlation coefficient is zero, it does not mean that there is no correlation, but there may be non-linear correlation.
5. The value of ‘r’ does not change if the independent (x) and dependent (y) variables are interchanged.
6. The correlation coefficient ‘r’ does not change if the scale on either variable is changed. You may multiply, divide, add, or subtract a value to/from all the x-values or y-values without changing the value of ‘r’.
7. The correlation coefficient ‘r’ has a Student’s ‘t’ distribution.

Assumptions to use the Pearson product-moment correlation

1. The measures are approximately normally distributed.
2. The variance of the two measures is similar (homoscedasticity).
3. The relationship is linear.
4. The sample represents the population.
5. The variables are measured on an interval or ratio scale.

***

2.4

Regression Analysis[*]

Introduction

Correlation gives us the idea of the measure of magnitude and direction between correlated variables. Now it is natural to think of a method that helps us in estimating the value of one variable when the other is known. The fact that the variables X and Y are correlated does not necessarily mean that x causes y or *vice versa*. Regression analysis is a statistical tool for the investigation of relationship between variables. It is a powerful technique used for predicting the unknown value of a variable from the known value of another variable. When there is only one independent variable then the relationship is expressed by a straight line. This procedure is called simple linear regression. More precisely, if X and Y are two related variables, then linear regression analysis helps us to predict the value of Y for a given value of X or *vice verse*. Multiple regression is an extension of bivariate regression in which several independent variables are combined to predict the dependent variable. In multilple regression analysis, the value of Y is predicted for given values of X_1, X_2, …, X_k.

Dependent and independent variables

By simple linear regression, we mean models with just one independent and one dependent variable. The variable whose value is to be predicted is known as the dependent variable and the one whose known value is used for prediction is known as the independent variable. Similarly, for multiple regression the variable whose value is to be predicted is known as the dependent variable and the ones whose known values are used for prediction are known independent (exploratory) variables.

The regression model

The line of regression of Y on X is given by Y = a + bX, where a and b are unknown constants known as intercept and slope of the equation. This is used to predict the unknown value of variable Y when value of variable X is known. The simple regression model is:

$$\boldsymbol{Y = a + bX}$$

The coefficient of X in the line of regression of Y on X is called the regression coefficient of Y on X. It represents change in the value of dependent variable (Y) corresponding to unit change in the value of independent variable (X).

In general, the multiple regression equation of Y on X1, X2, …, Xk is given by:

$$\boldsymbol{Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + b_k X_k}$$

Here b_0 is the intercept and b_1, b_2, b_3, …, b_k are analogous to the slope in linear regression equation and are also called regression coefficients. They can be interpreted as the change in the value of dependent variable (Y) corresponding to unit change in the value of independent variable X_i.

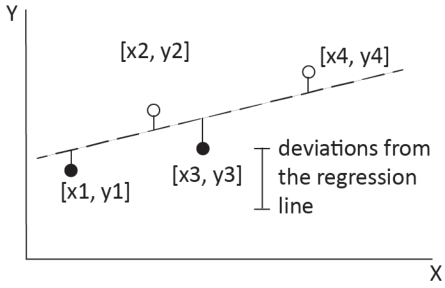
---

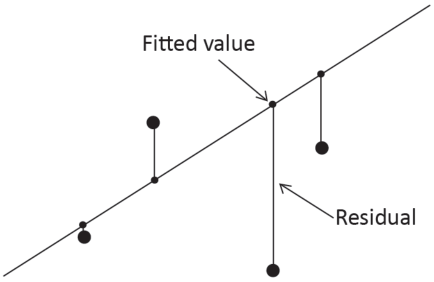
[*] *Somy Kuriakose, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Fitting of regression line

In scatter plot, we have seen that if the variables are highly correlated then the points (dots) lie in a narrow strip. If the strip is nearly straight, we can draw a straight line, such that all points are close to it from both sides. Such a line can be taken as an ideal representation of variation. This line is called the line of best fit if it minimizes the distances of all data points from it and is also called as the line of regression. Now prediction becomes easy because all one need to do is to extend the line and read the value. Thus to obtain a line of regression, it is necessary to have a line of best fit.

The problem of choosing the best straight line then comes down to finding the best values of ‘a’ and ‘b’. By ‘best’ we mean the values of ‘a’ and ‘b’ that produce a line closest to all ‘n’ observations. This means that we find the line that minimizes the distances of each observation to the line. Choose the ‘a’ and ‘b’ values that give the line such that the sum of squared deviations from the line is minimized. This method of estimation of parameters is called the least square method. The best line is called the regression line, and the equation describing it is called the regression equation. The deviations from the line are also called residuals.





R^2 - coefficient of determination

Once a line of regression has been constructed, one can check how good it is (in terms of predictive ability) by examining the coefficient of determination (R^2), which is defined as the proportion of variance of the dependent variable that can be explained by the independent variables. The coefficient of determination is a measure of how well the regression equation $y = a + bx$ performs as a predictor of ‘y’. R^2 always lies between 0 and 1. Higher values of this are generally taken to indicate a better model. The closer R^2 is to 1, the better is the model and its prediction.

***

2.5

Analysis of Length-Weight Relationship*

Length-weight relationship (LWR) is of great importance in fishery assessments. Length and weight measurements can give information on the stock composition, life span, mortality, growth and production. Length-weight (L-W) relationships are very useful for fisheries and ecological research because they are used:

- to convert growth-in-length equations to growth-in-weight, for stock assessment models;
- for the estimation of the biomass of a species from length frequency distributions from both onboard surveys and underwater ('eyeballing') observations;
- as an estimate of the condition of fish; and
- for between-region comparisons of life histories of a certain species.

The relationship between weight (W) and length (L) in fishes has the form:

$$W = aL^b$$

In this equation, the parameters ***a*** and ***b,*** usually termed as length-weight parameters are to be estimated with the available length-weight data. Each species of fish will have a specific length-weight relationship or specific length- weight parameters. It may also differ between sexes and between stocks or those belonging to different geographical regions. The parameter ***a*** is a scaling coefficient for the weight at length of the fish species. The parameter ***b*** is a shape parameter for the body form of the fish species.

The length-weight relationships were originally used for estimating the weight corresponding to a given length and to provide information on the condition of fish.

In theory, one might expect that the exponent ***b*** would have a value of roughly ***b*** = 3 because the volume of a 3-dimensional object is roughly proportional to the cube of length for a regularly shaped solid. Length is one-dimensional, whereas weight which depends on volume is three-dimensional. Hence, there is a thinking that weight of a fish is proportional to cube of the length of the fish. That is, there exists cubic relationship between weight and length of a fish.

In practice, fish that have thin elongated bodies will tend to have values of ***b*** that are less than 3 while fish that have thicker bodies will tend to have values of ***b*** that are greater than 3. Thus this also help to determine whether somatic growth is isometric ($b=3$) or allometric. Values of ***b*** smaller, equal and larger than 3 indicate isometry, negative allometry and positive allometry respectively. When $b>3$, large specimens increase in height or width faster than in length, either as the result of a change in body shape with size, or because the large specimens in the sample are in better condition than the small ones. Conversely, when $b<3$, either the large specimens have changed the body shape, *i.e.*, become more elongated, or the small specimens were in better nutritional condition at the time of sampling.

** Somy Kuriakose, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Thus the growth of fish length and weight is not proportionate or the relationship between length and weight is not linear. This means that when the length is increased the increase in weight is not proportionate to it. It is rather non-linear type of relationship. However, since the estimation procedure for length-weight relationship is through linear regression. It has to be transformed into linear type by applying logarithmic transformation. If we take logarithm (*natural logarithm with base e*) the above model will become linear as:

$$1\mathrm{n}(W) = 1\mathrm{n}(a) + b\ 1\mathrm{n}(L) \quad \text{or } Y = A + b\,X$$

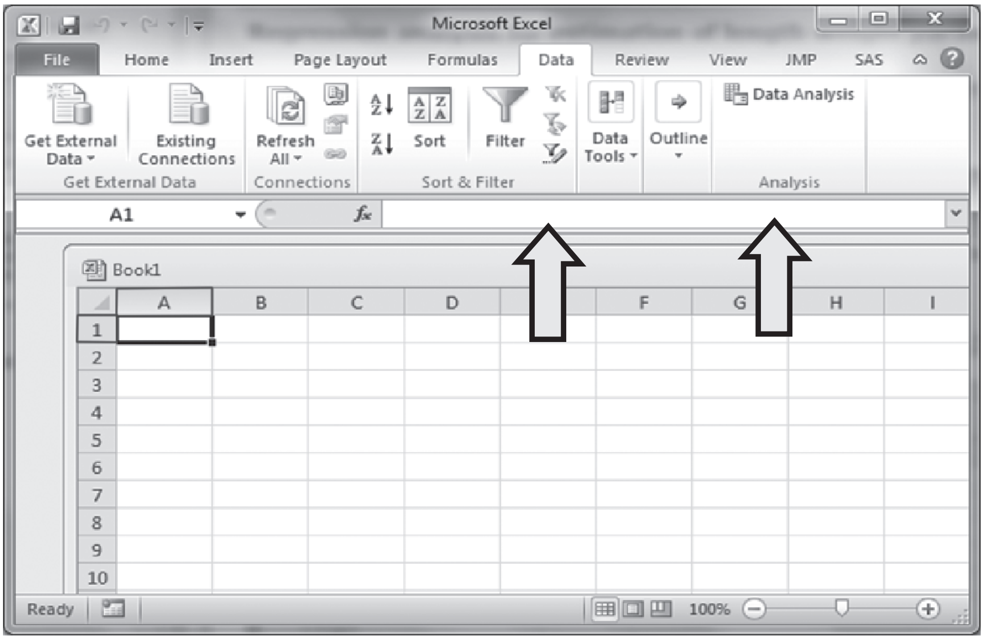
where ln(a) is the intercept and (b) the slope or regression coefficient. The above relationship is now linear and we can use the ordinary linear regression method for estimating the parameters of the relationship.

Data for fitting the length-weight relationship is collected randomly from the commercial catches or surveys and should represent fishes of all sizes, smallest to the biggest, and there should be enough samples for the analysis and estimation through regression. If our aim is to examine the difference in length-weight relationship between different sexes, then the data should be collected separately for males and females.

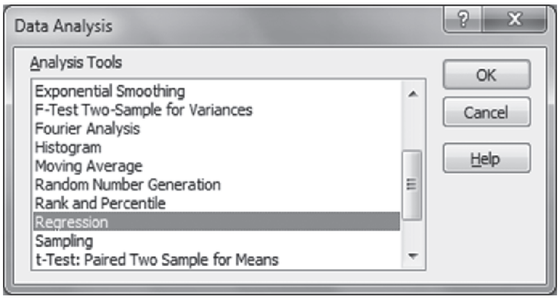
Regression analysis for estimation of length-weight parameters

Those who are familiar with MicroSoft Excel can do the analysis directly using the regression analysis tool.

Select Data from the Main Menu, and Select Data Analysis.

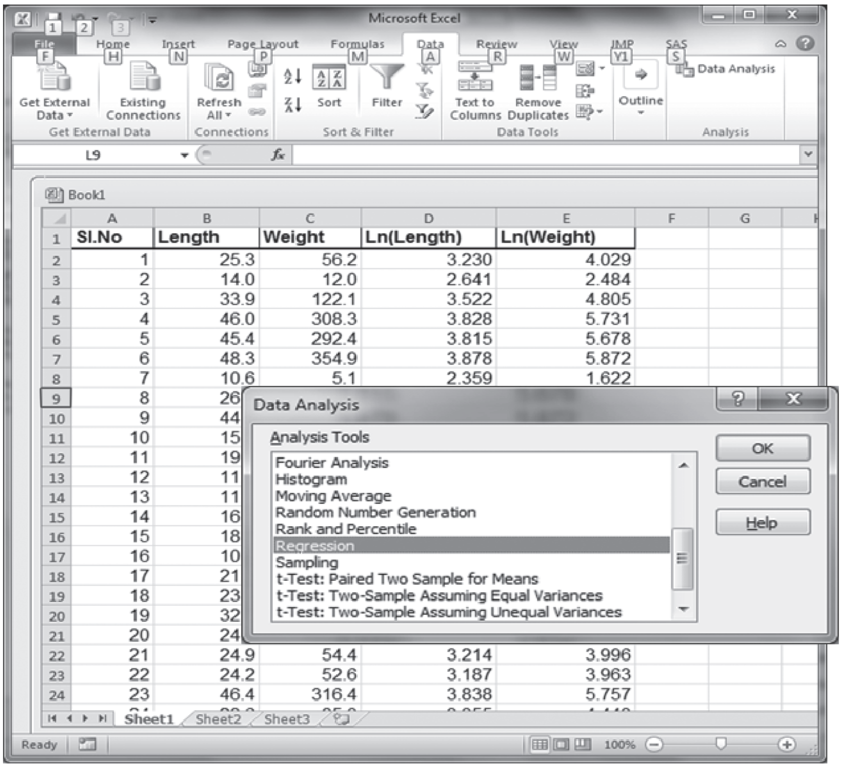


Select ‘Regression’ from the ‘Data Analysis’ dialog box and click OK.



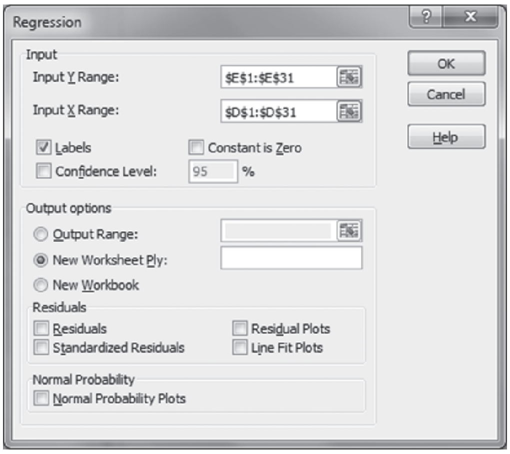
The following example demonstrates the use of this tool for estimation of length-weight parameters.

Enter the data on length and weight of samples in two columns as shown below. Generate two columns as the logarithmic values of the length and weight by using the natural logarithm function ‘ln’. The transformed data will be used for estimation of parameters. To run the regression routine Select Data from the Main Menu, and Select Data Analysis. From the dropdown menu select regression.



	A	B	C	D	E
1	Sl.No	Length	Weight	Ln(Length)	Ln(Weight)
2	1	25.3	56.2	3.230	4.029
3	2	14.0	12.0	2.641	2.484
4	3	33.9	122.1	3.522	4.805
5	4	46.0	308.3	3.828	5.731
6	5	45.4	292.4	3.815	5.678
7	6	48.3	354.9	3.878	5.872
8	7	10.6	5.1	2.359	1.622
9	8	26			
10	9	44			
11	10	15			
12	11	19			
13	12	11			
14	13	11			
15	14	16			
16	15	18			
17	16	10			
18	17	21			
19	18	23			
20	19	32			
21	20	24			
22	21	24.9	54.4	3.214	3.996
23	22	24.2	52.6	3.187	3.963
24	23	46.4	316.4	3.838	5.757

You will be presented with the following dialog: Specify the Log transformed weight data and label for “Input Y Range:” (D1:D31) and the log transformed length data and label for “Input X Range:” (E1:E31). Check the “Labels” box (since you included data labels in your input ranges), provide a new worksheet name under “Output options” and click OK.



Input Y Range: $E$1:$E$31

Input X Range: $D$1:$D$31

☑ Labels

☐ Constant is Zero

☐ Confidence Level: 95 %

○ Output Range:

● New Worksheet Ply:

○ New Workbook

☐ Residuals

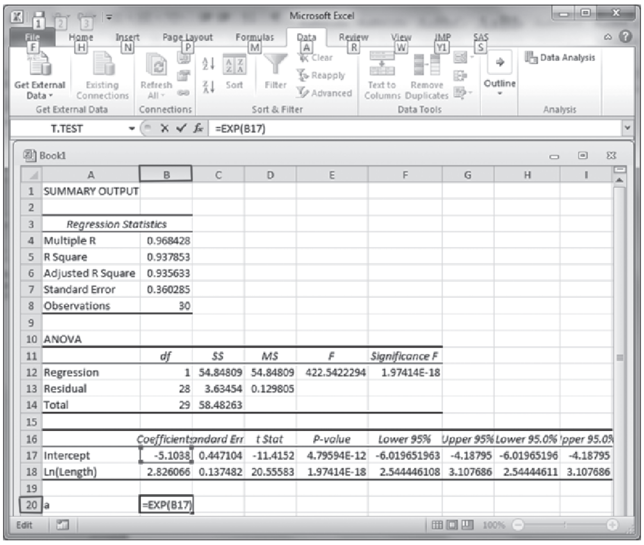
☐ Residual Plots

☐ Standardized Residuals

☐ Line Fit Plots

☐ Normal Probability Plots

The output will be obtained in a new sheet as given below:



T.TEST =EXP(B17)

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.968428							
5	R Square	0.937853							
6	Adjusted R Square	0.935633							
7	Standard Error	0.360285							
8	Observations	30							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	54.84809	54.84809	422.5422294	1.97414E-18			
13	Residual	28	3.63454	0.129805					
14	Total	29	58.48263						
15									
16		*Coefficient*	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*pper 95.0%*
17	Intercept	-5.1038	0.447104	-11.4152	4.79594E-12	-6.019651963	-4.18795	-6.01965196	-4.18795
18	Ln(Length)	2.826066	0.137482	20.55583	1.97414E-18	2.544446108	3.107686	2.54444611	3.107686
19									
20	a	=EXP(B17)							

The output will give regression statistics, ANOVA and the estimates of coefficients. The estimate of parameter 'a' is calculated from the value given against intercept and the estimate of parameter 'b' is that given against Ln(length) coefficient (here it is the value against 'ln(Len)' which is 2.826). The estimate of 'a' is calculated as the exponent of the intercept value which can be obtained by using the 'exp' function. For example, here the intercept value is in cell B17 and to obtain the estimate of 'a' in a blank cell use the function '=exp(B17)' to get the value of a as 0.00607.

The goodness of fit of the regression model is indicated by the 'R square' value in the output. It should be high for the relationship fitted to be good. In the example it is 0.96, indicating a good fit. The maximum value of 'R square' is 1.0 and the minimum is zero.

Using the estimated values of the parameters and the original data we can calculate the expected values of weight for the lengths in the sample data. This is done by substituting the estimated values in the relationship $W=aL^b$ and calculating the weights corresponding to each length in the sample.

Statistical test for b=3 (Isometric Relationship)

In a statistical test of hypothesis, this is testing for the null hypothesis H_0: $b=3$ against the alternative hypothesis H_1: $b \neq 3$. The test criterion for this statistical test is a Student's 't' statistic with (n-2) degrees of freedom where 'n' is the total number of observations.

Since this test criterion is for a linear regression, for the length-weight relationship situation we should use the log transformed values for the X and Y variables. Therefore, X values are the log transformed values of length and Y values are the log transformed values of the weight.

The test statistics for this is:

$$t_{n-2} = \frac{(b-3)\sqrt{(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2 - b^2\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

This value has to be compared with the table value of 't' for n-2 d.f for making inferences about the null hypothesis.

If the table value of Student's 't' is higher than the calculated value, we accept the null hypothesis that b=3. In that case we infer that the length-weight relationship is said to be isometric or there is cubic relationship between length and weight.

The length-weight relationship in fishes can be affected by a number of factors including season, habitat, gonad maturity, sex, diet and stomach fullness, health and preservation techniques, and differences in the length ranges of the specimens caught. The exact relationship between length and weight differs among species of fish according to their inherited body shape, and within a species according to the condition (robustness) of individual fish. Condition sometimes reflects food availability and growth within the weeks prior to sampling. But, condition is variable and dynamic. Individual fish within the same sample vary considerably, and the average condition of each population varies seasonally and yearly. Sex and gonad development are other important variables in some species.

***

Chapter 3

Sampling

3.1

Introduction to Sampling Theory*

Basic of Sampling

A “sample” is a miniature representation of a larger group or aggregate. In other words, the sample provides a specimen picture of a larger whole. This larger whole is termed as the “population” or “universe”. The procedure of collecting data for each and every unit belonging to the population, which is the aggregate of all units of a given type under consideration, is termed as census. The effort, money and time required for carrying out census will be enormously large. The census is necessary only if information is required for each and every unit in the population under study. Most of the time, it is impossible, difficult or expensive to observe all the elements of a population in order to arrive at a valid conclusion. Moreover, the sizes of populations are often so large that the study of all the units would not only be expensive but also cumbersome and time consuming, for example, the fishery resources in the Indian seas.

For fisheries research, it is impossible to collect information about all the fishery resources in the sea. So, a researcher will have to select a representative few, *i.e.* a sample from the population of the survey. This process is known as sampling. In fisheries research, the sampling surveys are an essential ingredient of fish stock assessment. In sample survey, data is collected from a part or a sample of the population under study in a suitable manner and inference is drawn about the population on the basis of observations made on the selected units. If the samples are selected with adequate criterion, it is possible to measure the precision of the conclusions or inferences about that population.

The objective of a sample survey is to estimate the unknown population parameters like total, ratio or median based on a random sample drawn by some specified rule from the population. A sample is considered as a subset of population. The principal advantages of sampling compared to census are reduction in cost, greater speed, wider scope and higher accuracy and more importantly the estimate error. Most fisheries research requires some planning and sampling designs before the research is undertaken. The objective of conducting sample surveys in fisheries is to collect data from the stocks and their exploitation level, to study the characteristics of the resources, the effects of exploitation on the abundance of these resources and to determine appropriate fishing levels to obtain the best possible catches at present and during future years. Fisheries research is most often concerned with the estimation of population mean and totals, *e.g.* the total catch by different types of gears at a particular landing centre.

Sampling error and non-sampling error

Sampling error arises from the fact that samples differ from their populations in that they are usually small subsets of the total population. Non-sampling errors can be defined as errors arising during the course of survey activities rather than resulting from the sampling procedures. Non-sampling errors can occur because of problems in coverage, response, non-response, data processing, estimation and analysis.

---

* *K.G. Mini, Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Methods of sampling

Sampling methods can be broadly classified into two categories:

Probability sampling

Non-Probability sampling

Probability sampling: Probability sampling is based on random selection of units from a population. In other words, the sampling process is not based on the discretion of the researcher but is carried out in such a way that the probability of every unit in the population is pre-defined.

Some of the characteristics of a probability sample are:

- each unit in the sample has some probability of entering the sample,
- weights appropriate to the probabilities are used in the analysis of the sample, and
- the process of sampling is automatic in one or more steps of the selection of units in the sample.

Non-probability sampling: Non-probability sampling is based on the judgement of the researcher. The guiding factors in non-probability sampling include the availability of the units, the personal experience and convenience of the researcher in carrying out a survey. Since these samples are not prepared through random sampling techniques, they are known as non-probability samples. Depending on the technique used, non-probability sampling includes purposive, incidental, snowball and quota sampling.

Sampling scheme

In survey sampling, a fixed finite population is under consideration, where the population elements are labeled so that each element can be identified. Probability sampling helps to select a random sample from a fixed population and assigns a probability of selection to each element. The probability of selection need not be equal for all the elements. A specific sampling scheme is used in drawing the sample. The collection of techniques for the selection of the sample is termed as sampling scheme.

A **sample scheme** can be described by two factors.

- **Sampling method**. Sampling method refers to the rules and procedures by which some elements of the population are included in the sample. The basic sampling techniques, which are commonly employed are simple random sampling, stratified sampling, cluster sampling, systematic sampling and sampling with unequal probabilities of selection of units particularly with probability proportional to size.
- **Estimator**. The estimation process for calculating sample statistics is called the estimator. Different sampling methods may use different estimators. For example, the formula for computing a mean score with a simple random sample is different from the formula for computing a mean score with a stratified sample. Similarly, the formula for the standard error may vary from one sampling method to the other.

The principal steps in a sample survey are given below:

i. Statement of the objectives of the survey: The objectives of the survey have to be defined clearly prior to the actual survey.

ii. Definition of the population to be sampled: The population will be used to denote the aggregate from which the sample is chosen. Whenever possible, the population to be sampled should obviously coincide with the population about which information is wanted.

iii. Determination of the data to be collected: It is important to verify that all the data are relevant to the purpose of the survey, and that no essential data are omitted.

iv. Methods of measurement: When the kinds of data that are needed have been decided, there may be a choice as to the methods of measurement to be employed.

v. Choice of sampling unit: As a preliminary to the selection of a sample, the population must be sub-divided in some way into parts, which will be called sampling units, or units. The sampling units must together comprise the whole of the population, and they must be non-overlapping, in the sense that every element in the population belongs to one and only one unit. The construction of a complete list of sampling units, sometimes called a frame, maybe one of the major practical problems. Sometimes the frame is impossible to construct, as with the population of fish in the sea.

vi. Selection of the sample: There are varieties of procedures by which the sample may be selected. The selection also involves a decision about the size of the sample, which in turn requires a provisional estimate of the cost of the survey, to ensure that the sample will fall within the allowable budget.

vii. Organization of the field work: In extensive surveys, many problems of administration are involved. The personnel must receive training in the purpose of the survey and in the methods of measurement to be employed and must be adequately supervised in their work. A procedure for early checking of the quality of the data may be invaluable. Plans must be made for handling non-response, that is, the failure of the enumerator to obtain information from certain of the units in the sample.

viii. Summary and analysis of the data: The first step is to edit the completed questionnaires, in the hope of amending recording errors, or at least of deleting data that are obviously erroneous. Decisions about tabulating procedure are needed in the case where answers to certain questions were omitted by some respondents or had to be deleted in the editing process. Thereafter, the tabulations which lead to the estimates are performed. Different methods of estimation may be available for the same data.

ix. Information gained for future surveys: The more information we have initially about a population, the easier it is to devise a sample which will give accurate estimates. Any completed sample is potentially a guide to improved future sampling, through the data which it supplies about the means, standard deviations, and nature of the variability of the principal measurements, and about the costs involved in getting the data.

Simple random sampling

Simple random sampling is a method of selecting a sample from a finite population in such a way that every unit of the population is given an equal chance of being selected. In practice, you can draw a simple random sample unit by unit through the following steps:

- Define the population.
- Make a list of all the units in the population and number them from 1 to N.
- Decide the size of the sample, or the number of units to be included in the sample.
- Use either the 'lottery method' or 'random number tables' to pick-up the units to be included in the sample.

For example, you may use the lottery method to draw a random sample by using a set of N tickets, with numbers 1 to N if there are N units in the population. After shuffling the tickets thoroughly, the sample of a required size, say 'n', is selected by picking the required 'n' number of tickets. The units which have the serial numbers occurring on these tickets will be considered selected. The assumption underlying this method is that the tickets are shuffled so that the population can be regarded as arranged randomly. When the size of the population is large, this procedure of numbering units on tickets and selecting one after reshuffling becomes cumbersome. Human bias and prejudice may creep in this method. Therefore, this method is generally discouraged.

The best method of drawing a simple random sample is to use a table of random numbers. After assigning consecutive numbers to the units of population, the researcher starts at any point on the table of random numbers and reads the consecutive numbers in any direction horizontally, vertically or diagonally. If the read out number corresponds with the one written on a unit card, then that unit is chosen for the sample.

Simple random sampling can be regarded as the basic form of probability sampling applicable to situations where there is no previous information available on the population structure. There are two methods of simple random sampling, namely simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRSWOR). Sampling with replacement means that each unit selected in the sample is returned to the population before the next is drawn.

In SRSWR, one unit of element randomly selected from the population is the first sampled unit. Then the sampled unit is replaced in the population. The second sample is drawn with equal probability. The procedure is repeated until the requisite sample units *'n'* are drawn. The probability of selection of an element remains unchanged after each draw. The same units could be selected more than once. Let N denote the population size and 'n' is sample size. As the population size remains the same after each draw, not only the probability of each unit being selected in the sample is $1/N$ at each draw, it remains same even when included in the sample more than once. In case of SRSWR, the number of all possible samples is N^n. The probability of drawing any of these N^n is $1/N^n$. For example, if there are three landing centres in a district, denoted by A, B, C (the population size N= 3). Then $3^2 = 9$ possible samples of size 2 can be drawn through SRSWR. They are (A, A) (A B) (A C) (B A) (B B) (B C) (C A) (C B) (C C). Each sample can be selected with equal probability of 1/9.

In SRSWOR, unlike SRSWR, once an element is selected as a sample unit, it will not be replaced in the population pool. The selected sample units are distinct. SRSWOR is a method of selecting 'n' units out of N such that every one of the ${}_{N}C_{n}$ distinct samples has an equal chance of being drawn. The simple random samples are drawn unit by unit. The units in the population are numbered from 1 to N. A series of random numbers between 1 and N is then drawn, either by means of a table of random numbers or by means of a computer program that produces such a table. At any draw, the process must give an equal chance of selection to any number in the population not already drawn. The units that bear these numbers constitute the sample. With reference to the same example above in SRSWR, the possible samples with SRSWOR without giving any importance to the ordering of the units are (A B) (A C) (B C). The total number of samples is ${}_{3}C_{2} = 3$. Any of these samples have equal probability of selection and the probability of selection of each of the samples is 1/ 3 . So in SRSWOR, the probability of selection of each unit at any draw is 1/N and the probability of inclusion of any unit in the sample is n/N.

Estimating the population mean

Let Y be the character of interest and $Y_1, Y_2, ..., Y_i, ..., Y_N$ be the values of the character on N units of the population. Let, $y_1, y_2, ..., y_i, ..., y_n$ be the sample of size 'n' selected by SRSWOR.

The estimator for the population mean is given by:

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

An unbiased estimator of variance of the population mean is given by:

$$\hat{V}(\hat{\bar{Y}}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} \text{, Where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Bound on the error of estimation (B)

$$B_y = t\,SE(\hat{\bar{Y}}) = t\sqrt{\hat{V}(\hat{\bar{Y}})}$$

Where *'t'* is the Student's 't' value for *n-1* degrees of freedom at the $1-\alpha/2$ level of significance. For $\alpha=0.05$, t = 1.96., Confidence Interval is given by C.I = $(\hat{\bar{Y}}) \pm B_y$

Estimating the population total

The estimator for the population total is given by:

$$\hat{Y} = N\hat{\bar{Y}} = \frac{N}{n}\sum_{i=1}^{n} y_i$$

An unbiased estimator of the variance of the population total is given by:

$$V(\hat{Y}) = N^2 V\left(\hat{\bar{Y}}\right) = N^2\left(1 - \frac{n}{N}\right)\frac{s^2}{n}$$

Where S^2 is an unbiased estimator of the population mean square S^2

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

Advantages of simple random sampling

One of the best things about simple random sampling is the ease of assembling the sample. It is also considered as a fair way of selecting a sample from a given population since every member is given equal opportunities of being selected. Another key feature of simple random sampling is that the sample will be a representative of the population. If the sample is not representative of the population, the random variation called sampling error will be large. An unbiased random selection and a representative sample are important in drawing conclusions from the results of a study.

Disadvantages of simple random sampling

One of the most obvious limitations of simple random sampling method is its need of a complete list of all the members of the population. However, from a practical point of view, a list of all the units of a population is not possible to obtain for large populations. Even if it is possible, it may involve a very high cost, which a researcher or an organisation may not be able to afford. Therefore, simple random sampling is difficult to realize. Also, in case of a highly heterogeneous population, a simple random sample may not necessarily represent the characteristics of the total population, even though all selected units participate in the investigation. In such cases, it is wiser to use other sampling techniques.

Stratified random sampling

When the population is heterogeneous, the whole population can be divided into sub-populations, called strata, to increase the precision of the estimates. In stratified sampling, the population of N units is first divided into disjoint groups of N_1, N_2 ,.., N_h ,.., N_L units, respectively. These sub-groups, called strata, together comprise the whole population, so that $N_1+N_2+ ....+ N_h +..+N_L= N$. The strata should not overlap and each stratum should be sampled following some sampling design. The strata are sampled separately and the estimates from each stratum combined into one estimate for the whole population. If a simple random sample selection scheme is used in each stratum then the corresponding sample is called a stratified random sample.

Reasons for stratification

- To obtain estimates of known precision for certain sub-divisions of the population by treating each sub-division as a stratum. Since sampling is done independently in each stratum, separate stratum estimates and their precision can be obtained by treating each stratum as a "population" in its own right. For example, in fishery surveys, estimates may be required by state or province, district, month, landing centre, craft, species, etc.

- For administrative convenience, for example stratification can provide survey organization to control the distribution of fieldwork among its regional offices.
- Sometimes different parts of the population may call for different sampling procedures.
- Stratification may often produce a gain in precision of the estimates of characteristics of the whole population. The amount in the gain depends on the type of stratification. If the population is heterogeneous and if it can be divided using prior information about the population into sub-populations (strata), each of which is internally homogeneous. If each stratum is homogeneous, that is characteristic under consideration vary little from one unit to another, a precise estimate (an estimate with smaller variance) of any stratum parameter can be obtained from a small sample in that stratum. These estimates can then be combined to obtain a precise estimate for the whole population.

Notations

The suffix h (h=1,2,...,L) denotes the stratum and i the unit within the stratum.

N_h = Total number of population units in stratum h.

n_h= Total number of sample units in stratum h.

w_h= N_h/N = The hth stratum weight.

Y_{hi} = Value of the characteristic for the ith unit in stratum h.

$Y_h = \sum_{i=1}^{N_h} Y_{hi}$ = Population total of Y – values for units belonging to stratum

$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi} = \frac{N_h}{N_h}$ = Population mean of Y-values for units belonging to stratum h.

$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$= Population variance of Y-values for units belonging to stratum h.

$\bar{Y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h} Y_{hi}}{\sum_{h=1}^{L} N_h} \sum_{h=1}^{L} W_h \bar{Y}_h$= Population mean of Y-values.

$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} = \frac{y_h}{n_h}$ = Sample mean of Y-values for units belonging to stratum h.

$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{hi} - \bar{y}_h\right)^2$ = Sample variance of Y-values for units belonging to stratum h.

$$\bar{y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h} Y_{hi}}{\sum_{h=1}^{L} n_h}$$ = Sample mean of Y-values.

Estimation of population mean and variance

An unbiased estimator of $\bar{y}$ is given by:

$$\hat{\bar{Y}} = \sum_{h=1}^{L} W_h \bar{y}_h$$

Also, since sampling is done independently within each stratum.

$$V\left(\hat{\bar{Y}}\right) = \sum_{h=1}^{L} W_h^2 \frac{N_h - n_h}{N_h - 1} \sigma_h^2$$

Note that $V\left(\hat{\bar{Y}}\right)$ cannot be computed since it involves Y-values for all the units in the population. However, based on Y-values for the sampled units we can estimate $V\left(\hat{\bar{Y}}\right)$ by using the following formula:

$$\hat{V}\left(\hat{\bar{Y}}\right) = \sum_{h=1}^{L} W_h^2 \frac{N_h - n_h}{N_h n_h} s_h^2$$ which estimates $V\left(\hat{\bar{Y}}\right)$ unbiasedly.

Allocation in stratified random sampling

In planning a study requiring stratification of the population, an important consideration is how to allocate a total sample size ‘n’ among the L identified strata. There are three types of allocations, as shown below:

Allocation of a sample to strata

1. Equal: If the strata are presumed to be of roughly equal size, and there is no additional information regarding the variability or distribution of the response in the strata, equal allocation to the strata is probably the best choice. $n_h = n/L$
2. Proportional: If the strata differ in size, allocation of sample sizes to strata might be

performed proportional to these stratum sizes. $n_h = \left(\frac{N_h}{N}\right) n$

3. Optimum (Neyman): The allocation which minimizes the variance of the estimator

of the mean (and total) is given by $n_h = \dfrac{n N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h}$

Whenever the strata are heterogeneous among themselves and the variance of each stratum is small, the sampling variance of the mean or total value estimators obtained by stratified sampling will always be smaller than the simple random sampling. The relative sizes of strata must be known to obtain the full benefits of the stratification technique. Each stratum should be internally homogeneous. If information about heterogeneity is not available then consider all strata equally variable. A short stratified pilot survey can sometimes provide useful information about internal dispersion within strata. A small-sized sample could be taken from a stratum if the variability among their units is small. A larger sample from a stratum should be taken if the stratum is larger, the stratum is more heterogeneous and the cost of sampling the stratum is low.

***

3.2

Systematic, Cluster and Multistage Sampling[*]

Systematic Sampling

The sampling methods, which are more popular and usually precede a discussion on systematic sampling, basically involve the methodology of selection of successive units with the help of random numbers. But there are situations where only the first unit is selected at random and the rest all the other units get selected automatically thereafter. This method is called systematic sampling.

As the entire sample is selected based on one unit, the mode of selection is based on a pattern. Such pattern involves regular spacing of units either spatially or temporally whose first occurrence is amenable to selection randomly. Suppose a population contains N units serially numbered from 1 to N and suppose further that N is expressible as a product of two integers ‘k’ and ‘n’ so that $N=kn$, draw a random number less than ‘k’, say ‘i’, and select the unit with corresponding serial number and every kth unit in the population thereafter. Obviously the sample will contain the n units $i, i+k, i+2k, \ldots, i+(n-1)k$ and is known as a systematic sample.

Some examples:

(a) The selection of a maize (corn) field , every k^{th} mile apart, for the observation on pest incidence.

(b) Observing every k^{th} strip of forest to estimate timber.

(c) Selection of every k^{th} interval of time (say two hourly) for observing the number of fishing craft landing on the coast.

(d) Selection of every k^{th} village from a list of villages.

Systematic sampling obtains the basic property of a probability sample just because the selection of the first unit has many practical advantages. This method is simple to adopt and economical. The field-level workers can easily succeed in performing this type of sampling as it has very little technical maneuvering. Further, this ensures better control over the field work. By its very design, the systematic sampling ensures even spread of the sample through the length and breadth of the population.

On the disadvantage side, the relative position in the population of the different units included in the sample is fixed. Hence, if the population exhibits some cyclicity then this method is liable to pitfalls. For example, if the task is to estimate daily vehicular traffic in a locality and if every 7th day is a holiday then a systematic sample selected with a typical holiday would be highly skewed. Another significant drawback of this method is that as only one cluster or group of preset units is selected, estimation of variability of the estimated mean, technically called as sampling variance, is not obtainable.

---

[*] *J. Jayasankar, Principal Scientist and K.G. Mini, Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Similarities with other methods

Systematic sampling resembles Stratified Random Sampling in that one sampling unit is selected from each stratum of ‘k’ consecutive units. This similarity is strictly casual as the procedure followed while adopting stratified sampling is quite different. Systematic sampling resembles cluster sampling, a systematic sample being equivalent to a sample of one cluster selected out of ‘k’ clusters of ‘n’ units each.

Cluster Number	1	2	…	i …	…	k
1	1	2	.	i	.	k
			.		.	
2	1 + k	2 + k	.	i + k	.	2k
.	.	.	.	.	.	.
j	1+ (j-1)k	2+(j-1)k	.	i+(j-1)k	.	jk
.	.	.	.	.	.	.
n	1+(n-1)k	2+(n-1)k	.	i+(n-1)k	.	nk

The following schematic diagram shows the possible representation of a population considered for systematic sampling in clusters (Sukhatme, 1954).

Since the first number less than or equal to ‘k’ is to be chosen at random, every one of the ‘k’ columns gets an equal chance of being chosen as the systematic sample.

Estimates of parameter

Under a systematic sampling design, the estimate of the commonest parameter, mean, is given by:

Sample mean = sum of all the values of selected units comprising the sample/ the number of units selected or using notations,

Sample mean $\bar{y}i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$

And this is used as the estimate of the population mean given by:

$$\bar{y}_{-} = \frac{1}{nk} \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}$$

The variance of the estimate is given by:

$$\frac{\text{Sum of squared deviations of the means of each cluster from the population mean}}{\text{number of clusters (k)}}$$

or in notations

$$V(\bar{y}_i) = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_i - \bar{y}_{-})^2$$

It is this parameter, the sampling variance of mean, which cannot be estimated as only one $\bar{y}_{i.}$ Is known at the end of the sampling operation and that itself is the estimate of population mean.

Cluster sampling

A sampling design pre-supposes the division of the population into a finite number of distinct and identifiable units called as sampling units. Thus a population of fields under paddy in a given region might be regarded as composed of fields or groups of fields, on the same holdings, villages or other suitable settlements. A human population might similarly be regarded as composed of individual persons, families, or groups of persons residing in houses in villages. The smallest units into which the population can be divided are called the elements of the population and groups of elements are called as clusters. When the sampling unit is a cluster, the procedure of sampling is called cluster sampling. When the entire area containing the population under study is sub-divided into smaller areas and each element in the population is associated with one and only one such small area, the procedure is alternatively called area sampling.

For many types of population, a list of elements is not available and the use of an element as the sampling unit is therefore not feasible. The method of cluster or area sampling is available in such cases. Thus in a city, a list of all the houses is readily available, but that of persons is rarely available. Again lists of fields are not available but those of villages are. Cluster sampling is therefore widely practiced in sample surveys.

The unique situations on the field could leave the clusters to be selected as equal or unequal. The procedures for estimation will slightly vary depending upon the cluster sizes and their equality. Here let us focus on the case of equal sized clusters.

We shall first consider the case of equal clusters. For example, the population is composed of N clusters of M elements each, and that a sample of 'n' clusters is drawn from it by the method of simple random sampling.

For notational simplicity, let:

y_{ij} denotes the value of the character for the jth element (j=1,2,…, M) in the ith cluster (i=1,2,..,N)

$\bar{y}_i = \frac{1}{M}\sum_{j=1}^{M} y_{ij}$ be the mean for the elements of the ith cluster

$\bar{y}_{N.} = \frac{1}{N}\sum_{i=1}^{N} y_{i.}$ be the mean of cluster means in the population

$\bar{y}_{..} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}$ be the mean per element in the population and

$\bar{y}_{n.} = \frac{1}{n}\sum_{i=1}^{n} y_{i.}$ be the mean of cluster means in a simple random sample of n clusters and this is the unbiased estimator for the population mean.

The variance of this mean is given by:

$$V(\bar{y}_{n.}) = \frac{N-n}{Nn}\left(\frac{1}{N-1}\sum_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{N.})^2\right)$$

And its estimate being:

$$\hat{V}(\bar{y}_{n.}) = \frac{N-n}{Nn}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_{i.} - \bar{y}_{n.})^2\right)$$

Thus systematic and cluster sampling procedures form a very important pair of sampling designs, which are quite practical and economically viable in estimating parameters of interest.

Multistage sampling

In the cluster sampling, all the elements of the selected clusters are enumerated. Though cluster sampling is convenient and economical, the method restricts the spread of the sample over the population and it results in reduction in the efficiency of the estimator. There are many ways to modify cluster sampling for more complex sampling situations. One common modification is to take a sample of secondary units from within sampled clusters instead of inspecting every secondary unit within each sampled cluster. The process of drawing samples from the selected clusters is called sub-sampling.

The sampling method which consists of selecting the clusters and then randomly choosing a specified number of units from each selected cluster is known as two-stage sampling. The clusters which form the units of sampling at the first stage are called first stage units (fsu's) or primary stage units (psu's) and the elements within the clusters are called second stage units (ssu's). For example, to obtain a sample of fish caught in a commercial fishery, it may be necessary to first take a sample of boats and then a sample of fish from each selected boat. The simplest type of multistage sampling is two-stage sampling. The estimation mean and variance is given below.

Let N be the number of primary units in the population and 'n' be the number of primary units selected in the first stage. If each primary unit contains M sub-units, of which 'm' are chosen, the,

$\bar{y}_{ij}$ = value obtained for the j^{th} element in the i^{th} primary unit

$\bar{y}_i = \sum_{j=1}^{m}\frac{y_{ij}}{m}$ = sample mean per element in the ith primary unit

$\bar{\bar{y}} = \sum_{i=1}^{n}\frac{\bar{y}_i}{n}$ = over-all sample mean per element

$S_1{}^2 = \sum_{i=1}^{N}\frac{(\bar{Y}_i-\bar{Y})^2}{N-1}$ = variance among primary unit means

$S_2{}^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij}-\bar{Y}_i)^2}{N(M-1)}$ = variance among elements within primary units

The simplest estimates of the population total and mean per sub-unit are, respectively,

$$\hat{Y} = \frac{NM}{n}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n),$$

$$\hat{\bar{\bar{Y}}} = \frac{1}{n}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n)$$

Where $\bar{y}_i$ is the sample mean per sub-unit in the i^{th} primary unit.

If the *‘n’* units and the *‘m’* sub-units from each chosen unit are selected by simple random sampling, $\bar{\bar{y}}$ is an unbiased estimate of $\bar{\bar{Y}}$ with the variance:

$$V(\bar{\bar{y}}) = \left(\frac{N-n}{N}\right)\frac{{S_1}^2}{n} + \left(\frac{M-m}{M}\right)\frac{{S_2}^2}{mn}$$

An unbiased estimate of $V(\bar{\bar{y}})$ is:

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n}{s_1}^2 + \frac{f1(1-f_2)}{mn}{s_2}^2$$

where , ${s_1}^2 = \frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{\bar{y}})^2}{n-1}$ ${s_2}^2 = \sum_i^n \sum_j^m \frac{\left(y_{ij} - \bar{y}_i\right)^2}{n(m-1)}$

The two stage sampling method can be generalized to three or more stages and is termed as multistage sampling. In one-stage cluster sampling, the estimate varies because different samples of primary units yield different estimates. In two-stage cluster sampling, the estimate varies due two sources: different samples of primary units and then different samples of secondary units within primary units. In general, if there are ‘k’ stages of sub-sampling, there will be ‘k’ sources of variability. Thus, variances and variance estimators for multistage sampling with k-stages will contain the sum of ‘k’ components of variability.

Multistage sampling designs are used for a variety of practical reasons and this procedure is being commonly used in large-scale surveys. This sampling procedure is more flexible as it permits the use of different selection procedures at different stages. The multistage sampling may be the only choice in a number of practical situations where a satisfactory sampling frame of ultimate stage units is not readily available and the cost of obtaining such a frame is large and time consuming.

Reference

Sukhatme, P. V. 1954. Sampling Theory of Surveys with Applications: The Indian Society of Agricultural Statistics: New Delhi. 491p.

***

3.3

Sampling Design followed by CMFRI for Estimates of India's Exploited Marine Fishery Resources*

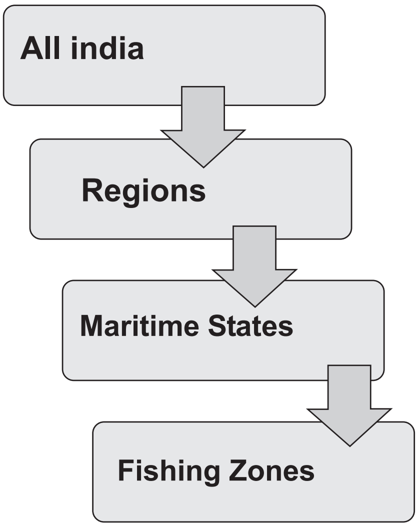


Figure 1. Stratification over space

India has a coastline of about 8129 km. Marine fish landings take place almost all along in the coastline during the day and sometimes during night. According to marine fisheries census, 2010, there are 3288 fishing villages scattered along the coast from where fishermen go for fishing and return to a landing centre, which may be distinct from the fishing village. There are 1511 landing centres scattered along the coastline of the mainland. Under these conditions, collection of statistics by complete enumeration would involve very large number of enumerators and huge sums of money, apart from the time involved in collection of data. In this situation, a feasible solution for obtaining marine fish landings is the adoption of a suitable sampling technique for the collection of fish landing data. The sampling design adopted by the Central Marine Fisheries Research Institute (CMFRI) to estimate resource-wise/region-wise marine fish landings is based on stratified multi-stage random sampling design, in which, the stratification is over space and time. Over space, each maritime state is divided into suitable, non-overlapping fishing zones on the basis of fishing intensity and geographical considerations (Fig. 1).

Selection of fish landing centre days

The number of landing centres may vary from zone to zone. These zones have been further stratified into sub-strata, on the basis of intensity of fishing. Each zone is regarded as a stratum in space. The stratification over time is by calendar month. A zone and a calendar month constitute a space-time stratum. If in a zone, there are 25 landing centres and there are 30 fishing days in the month; we get 25 x 30 = 750 landing-centre days, which constitute the primary stage units (PSU). The fishing boats that land on a landing centre-day forms the second stage units (SSU). The introduction of space-time stratification in the sampling methodology becomes necessary as the fish population is supposed to vary with respect to both space and time. Stratification is intended to reduce variance in the sample estimates. The fish landings are found to vary considerably among the landing centres in a multi-centre zone, especially in different seasons and hence a zone is further stratified as major, minor and very minor centres. The centres in which either mechanised boats or 100 or more

---

* *K.G. Mini, Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

non-mechanised/ motorised boats are operating, are considered as major centres. Similarly, other strata are defined based on the number and type of fishing boats operating.

Further, a month is divided into 3 groups, each of 10 days. From the first five days of a month, a day is selected at random, and the next 5 consecutive days are automatically selected. From this, three clusters of two consecutive days are formed. For example, for a given zone, in a given month, from the five days if the date (day) selected at random is 4, then the clusters formed from the first 10 day group are (4, 5), (6, 7) and (8, 9). In the remaining ten-day groups, the clusters are systematically selected with an interval of 10 days. For example, in the above case, the cluster of days for observation in the remaining groups are (14, 15), (16, 17), (18, 19); (24, 25), (26, 27) and (28, 29). Normally in a month, 9 clusters of two days each can be obtained. From among the total number of landing centres in a zone, 9 centres are selected with replacement and allotted to the 9 cluster days selected as described earlier. These 9 days are evenly distributed among the strata in case of fishing zones with more than 1 landing centre. A landing centre-day, which is the PSU, comprises 24 hour duration from noon of the first day to the noon of the following day.

A landing centre-day has been divided into 3 periods as given in the infographic below. One field staff is usually provided to each zone. A field staff starts data collection from period 1 on each selected landing centre-day. He/She will be present throughout the periods 1 and 2 at the centres. The data on landings during period 3 (night landings) is usually collected from the landing centre by enquiry on the following day morning. The sum of the observations on the 3 periods contribute to the data for one landing centre-day (24hrs). Thus in a 10-day period, data from 3 centre-days are sampled and consequently in a month, 9 landing centre-days are sampled.

Period 1	1200 to 1800 hours on 1st day
Period 2	0600 to 1200 hours on 2nd day
Period 3	1800 hours to next morning 0600 hours

Selection of fishing units and recording of fish landings

It may not be practicable to record the catches of all boats landed during an observation period, if the number of boats/ craft is large. In this case, sampling of the boats/craft becomes essential. When the total number of boats landed is 15 or less, all the boats are enumerated for catch composition and other particulars. When the total number of boats exceeds 15, the procedure in the infographic is followed to sample the number of boats.

Less than or equal to 15 boats	• 100%
16 - 19	• First 10 and the balance 50%
20 - 29	• 1 in 2
30 - 39	• 1 in 3
40 - 49	• 1 in 4
50 - 59	• 1 in 5 and so on

From the boats, the catches are normally removed in baskets of standard volume. The weight of fish contained in these baskets being known, the total weight of the fish in each boat under observation is obtained. The procedure of selection of the landing centre-days and the boats landed on the selected day for single-centre zones is the same as in the case of a stratum in a multi-centre zone.

Estimation procedure

From the landings of the observed fishing units, the landings by all the units during the observation period are estimated. By adding the quantities landed during the two 6- hourly period and during the night (12-hours), the quantity landed for a day (24-hours) at the centre, that is the landings for each centre-day included in the sample, is estimated. From these, the monthly zonal landings are obtained. From the zonal estimates, district-wise, state-wise and all India landings are arrived. The corresponding sampling errors are also estimated.

Administration of the survey

Plan of operation

The survey staff is given 10-12 weeks training course immediately after recruitment and is posted at the survey centres. Each survey centre is housed in a 1-2 room accommodation and each centre is provided with literature connected with identification of fish, a reference collection of local fish species, crustaceans and molluscs, field notebooks and registers. The programme of work for the following month is carefully designed by the staff of Fishery Resources Assessment Division (FRAD) at the CMFRI headquarters. Generally, one field staff is allotted to each zone to collect the fish landing data. At the end of every month, the survey staff receives the programme of work for the next month by post, that includes the names of landing centres to be observed and details such as dates and time for observations at each landing centre. The field staff are instructed to send the data collected during every month to reach the Institute's headquarters (Kochi, Kerala) at least by the end of first week of the subsequent month.

Supervision of data collection

Surprise inspections are carried out by the supervisory staff of the Institute and the enumerators are inspected while at work in the field and their field notebooks and diaries are scrutinised. The estimated zonal landings are always compared with the previous year's survey figures, and if any variation which cannot be explained is observed, the technique of interpenetrating sub-samples is adopted to detect observational errors. Zonal workshops are held periodically to review the progress of work and update the sampling frame and to impart refresher courses to the field staff.

Errors due to non-response, their magnitude and control

Non-response occurs when the regular field staff is not available to observe the centre-day included in the sample. Usually, arrangements are made at the Headquarters/Research/Regional Centre to minimise the non-response.

Analysis of data

In the existing sampling methodology, the interest is to estimate gear-wise, species-wise landings for the state in a month, fishing effort according to different types of fishing boats and also in terms of man hours. The analysis is carried out at CMFRI headquarters. Before the data is processed for analysis it will be ensured that the data collection is made as per the approved schedule, by checking the appropriate proforma. The responsibilities and functions of staff at the headquarters are data coding, estimation and database management. The data analysis is computerised and estimates are made using the software developed by the FRAD. The processed data are again counter- checked for errors. When discrepancies are detected, the estimation procedure is scrutinised in detail.

Reference

M. Srinath, Somy Kuriakose and K. G. Mini, 2005. Methodology for the Estimation of Marine Fish Landings in India, CMFRI Special Publication No. 86, 57p.

***

3.4

Fish Sampling and Measurement[*]

The basic data required for stock assessment studies include (i) total catch, (ii) fishing effort, (iii) area and type of fishing, and (iv) biological characteristics of the catch (length/age, reproductive and feeding biology). These data can be collected either from commercial fisheries or from resource surveys. Whatever is the source of data, it is essential to ensure that (i) the data fully represent the stocks, and (ii) there is no systematic error or bias. Hence, the aim of data collection on a fishery should be to obtain samples that fully represent the population under investigation. Unbiased sampling requires adoption of proper tactics, and large and long-term investments on manpower and investment. In this Chapter, some basic aspects of fish sampling and length measurements are discussed.

Sampling programmes

Sampling commercial catch

For the purpose of sampling length composition data, samples may be collected either on board the fishing vessels, or at landing centres and wholesale fish markets. The most convenient place for sampling is the fish landing centre. Since number of boats land their catch within a short period of time, it is possible to examine and measure samples from several vessels within a short time.

For length measurement, the objective of selection is to collect random sample and, for this, suitable method should be applied depending upon the type of landings. Generally, sampling fish at selected landing centres is done in two stages. First, the boat from which the sample is proposed to be collected is chosen at random during the selected day. At the second stage, the catch of the selected boats is sampled. If the fish are landed in small numbers, all the fish should be sampled. If the fish are landed in large numbers and they are heaped in the landing centre, it is necessary to select only a small number as the sample for length measurement. Such a selection should be random and should not be biased. The selected sample is assumed to represent the population. For collecting a random sample, it is necessary to adopt a prescribed method of selection. One method of collecting random sample is to divide the heap into two equal halves and select one half at random. If the selected half is also large, again this half may be divided into two halves in the same way as before, and the process should be continued until a sample of desired numbers is obtained. The catches of large quality fishes, crustaceans and cephalopods are usually placed separately in containers/baskets. In such cases one or two containers chosen at random form a good sample. The length of all the fish in the sample should be measured. The procedure of sampling remains the same for one species or for mixed species. In the case of mixed species, the desired species may be sorted out first before further sub-sampling is done.

One of the important aspects in a sampling programme for length composition is to decide on the number of fish to be measured. The number to be measured depends on the length

---

[*] *E. Vivekanandan, National Consultant, World Bank/GEF Ocean Partnership Project, Bay of Bengal Programme Inter-governmental Organisation, 91, Saint Mary's Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.*

range of the species under consideration. If all fish landed are of the same size, measuring one fish is sufficient. On the other hand, if the length range is large, as in the case of tropical fisheries, a comparatively large sample will be required. For instance, if the size range is 50 to 300 mm and, if it is divided into 20 length groups, the sample size may have to be as high as 200 or more. Moreover, the length may vary considerably among the landings of different vessels, if the vessels had fished in different schools. Hence, it is advisable to sample many vessels rather than selecting the desired sample from a single vessel. Since the length composition of a stock may change with time due to growth, recruitment and migration, it is better to collect samples at regular intervals (every month, rather than collect a large sample on a single day). For instance, sampling 4800 specimens in a year, spread over collection of average of 200 specimens at fortnightly intervals is better than sampling the same number from just two collections of 6-month interval.

There are several practical considerations, which limit the sample size on a given day. Limited fund availability may be one of the major reasons for restrictions in collection or measurement of samples. Inadequate funds will limit the number of fish to be sampled. Inadequacies in sampling affect the precision of the estimates, especially in the case of high-valued shrimps and fishes like pomfrets and large-sized fishes like sharks, seerfishes and tunas.

Sampling catches from resource surveys

Samples are also collected from resource surveys using research vessels. These samples are useful for estimation of absolute abundance of length composition of stocks. In addition to length composition, these surveys provide information such as area of fishing, gear specification, fishing duration, depth of fishing, etc. The procedure for estimation from resource surveys is different from that for commercial catches. Commercial landings are visible and are observed after they are caught/landed. On the other hand, the natural fish population is neither visible nor directly measurable. Resource surveys provide an opportunity for sampling from selected fishing gear and fishing ground. For this, the gear and area of fishing should be carefully selected by considering the objective of sampling.

The main consideration should be the selection of suitable gears. No gear can catch all the species of fish and all the size groups of fish. For properly sampling the population in the sea, it may be necessary to use a set of different gears.

The other main consideration in sampling is the patchy distribution of fish in the sea. The distribution may be segregated by size groups, which may result in large variations among samples. This problem may be partially overcome by stratifying the entire area into smaller sub-areas. For demersal species, depth is the most important criterion. If variation among hauls is high, the number of samples should be high.

After fixing the areas, sub-areas and stations, sample hauls are taken from each station and the quantity and length measurements are recorded. Estimates of the number of fish of various length groups are obtained for each station and sub-area and later combined for the entire area. The precision of an estimate increases with increase in the number of sampling stations. The number of stations should be high if the variability in catches is high among stations. Moreover, an estimate derived from a single survey is not enough. As fish population is dynamic, it is necessary to repeat the survey in the same area. After completion of repeated surveys, the abundance for each area is estimated by multiplying the mean density in each sub-area by the total area under consideration by assuming that each haul catches all fish in the area covered by it, and that the density in the sampled area is a good measure of the overall density.

Methods of length measurement

Measurement of finfishes

The three length measures for fish are the total length, fork length and standard length (Fig. 1). Total length is measured from the top of either upper or lower jaw to the tip of the longest caudal fin rays. It may be made to the tip of the dorsal lobe or to the ventral lobe, or to the longer of the two or their average. For measurement of total length, the tail fin may be kept in the normal extended position, or the tips of one or both caudal lobes together in such a way that their outer edges are parallel to each other and to the axis. The International Council for the Exploration of the Sea (ICES) recommended the total length as the standard measurement of length. The ICES defined the total length as the maximum length by bringing the longest lobe of the caudal fin into the midline of the fish.

Fork length is measured from the tip of either upper or lower jaw to the cartilaginous tip of the shortest or medium caudal fin ray. In the case of tunas, the fork length is accepted as the standard dimension. The International Commission for the Conservation of Atlantic Tunas has confirmed fork length as the standard measure for the tunas.

Standard length is measured from the tip of the lower or upper jaw to the hypural bone (urostyle) or to the keel, or to the tip of the caudal peduncle or to the last scale. Since it is difficult to locate any of these last points of measurement when quick measurements are made, the standard length is often not measured uniformly and accurately and hence its use is not generally recommended.

In the case of rays and other dorsoventrally flattened fish, disc width rather than total length is generally used.

For measurements, it is essential to have a convenient measuring device. A standard type of measuring board consisting of a flat wooden or plastic platform with a scale centrally fitted on it with a headpiece at the zero end of the scale is useful in the field as well as in the laboratory. To measure a fish, its mouth is closed, placed flat in its right side, snout to the left on the measuring board. The snout is pressed down gently at the zero end of the measuring board and the body and tail straightened along the midline before the reading is taken from the scale.

The dimension of large fishes like the seerfishes and tunas is often measured from tip to tip along the body surface with a tape. This is a measurement of the curvature of the body, which will always be greater than the total length.

Measurement of crustaceans

The standard measurement of penaeid and nonpenaeid prawns and lobsters is the minimum length of the carapace from the inside of the eye socket to the posterior margin of the carapace (Fig. 1). In the case of the lobsters, the total length parallel to the midline is taken as the standard measurement. The standard measurement of the crabs is the maximum distance across the width of the carapace (including the spines, if present). All the measurements of crustaceans could be made with vernier caliper.

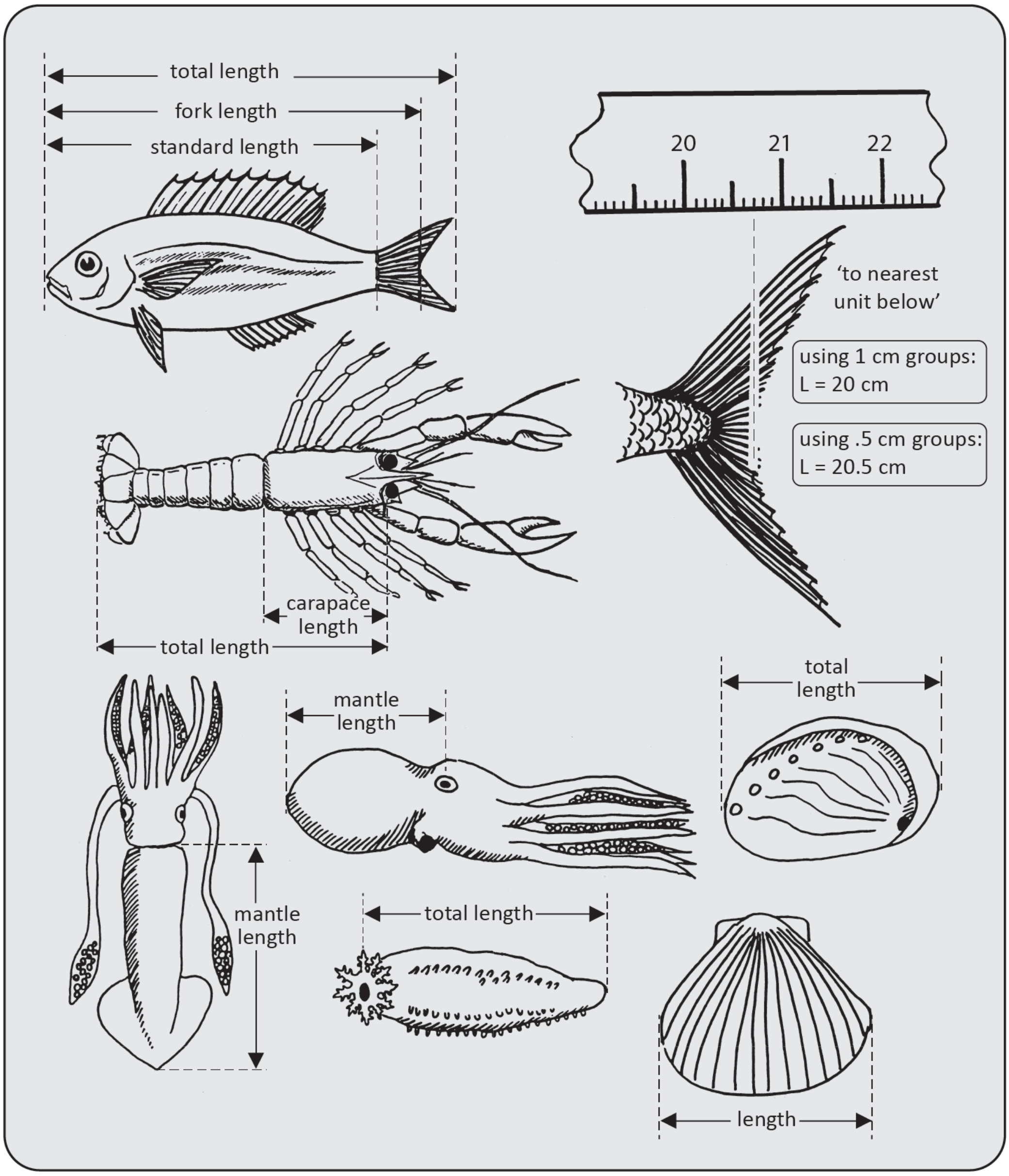


Fig. 1. Length measurements

Measurement of molluscs

The possible measurements of bivalves are the length, width and depth of the shell. In general, the length is measured since it is usually the longest dimension. The length of shell is defined as the greatest measure in an antero–posterior direction, approximately parallel with the axis of the hinge (Fig. 1). The standard measurement of gastropods is the maximum measurement from the tip of the whorl to the base of the shell. The standard measurement of cephalopods is the mantle length. However, because of body elasticity, weight is suggested as a more accurate and less variable index of size for the cephalopods.

***

Chapter 4

Growth Estimation

4.1

Estimation of Growth Parameters[*]

The data commonly used for fish stock assessment is the length frequency data collected periodically by sampling from commercial catches. The data so obtained will consist of fish of different age groups. The fish born on the same day (single spawning) is termed as a cohort and the fish of the same age will not have same length, rather it will vary with a mean and variance. If we make a histogram of their length, most of the fish will fall at the middle and it will have the well-known bell shape. The sample collected at a time will be a mixture of such bell shaped distributions corresponding to different age groups. If we are able to trace the length distributions of each cohort separately from its initial age up to its life span then we would be able to work out its growth and growth model parameters. As the sample collected from commercial catch will be a mixture of cohorts of different age groups, the problem reduces to resolution of individual components (known as normal distributions or Gaussian components) from the mixture.

Resolution of Gaussian components from polymodal distributions

The frequency distribution of length obtained from a sample of fish is usually skewed and polymodal. The modes corresponding to individual age groups are very useful in separating the different Gaussian components of which it is assumed to be composed of. Here the problem is to resolve a distribution into Gaussian components. Different procedures are available for resolution of a mixture into Gaussian components. These are probability paper method, parabola method and Bhattacharya's method. Among this the last method is the most popular.

Probability paper method: Decomposition of polymodal frequency distributions using probability paper method was introduced by Harding in 1949 and later modified by Cassie in 1950. This involves dissection of the distribution at points of inflexion of the probit plot, followed by correction for overlap of components. In this method, the cumulative percentages of the frequency distribution are first plotted against the mid-points of the classes on a probability graph paper and the points of inflexion are marked. Cumulative percentages of these points are the keys for separation of the components and each segment between them is due to separate distribution. Each of these components is then extracted by adjusting the original cumulative percentages within segments so that the total is 100. These adjusted values, if plotted on the same probability paper, will be linear. The means of each separated component are estimated from the actual frequencies falling in the corresponding region.

Parabola method: If the frequency distribution of random variable distributed as normal has *'y'* as the frequency for a class with mid-value *'x',* then we can express 'y' as:

$$y = N \int_{x-c/2}^{x+c/2} f(x)\, dx$$

where *f(x)* is the probability distribution of a normal random variable with mean μ, standard deviation σ the class interval and *N* the total frequency. An approximation for the relation is:

---

[*] *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

$$y \approx \frac{(N c e^{-(x-\mu)^2})/2\sigma^2}{\sqrt{2\pi\sigma}}$$

$$\ln(y) \approx \ln(\frac{Nc}{\sqrt{2\pi\sigma}}) - \frac{(x-\mu)^2}{2\sigma^2}$$

The above equation is of the form which is a quadratic, representing a parabola. The axis of symmetry of the above parabola will be at $x=\mu$. Hence, if we plot the natural logarithm of the class frequencies against the mid-values of the classes, we can represent the different peaks with different parabolas each corresponding to a normal distribution whose mean is the point where the axis of symmetry intersects the *x*-axis.

Bhattacharya's method: If $y(x)$ denotes the observed frequency of the class with *x* as its mid-value and *h* the class width, then:

$$y = \int_{x-h/2}^{x+h/2} \sum_{i=1}^{k} N_i f(x, \mu_i, \sigma_i)\, dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r f(x, \mu_r, \sigma_r)\, dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu_r)^2}\, dx$$

$$\ln(y) = \ln(\frac{hN_r}{\sigma_r \sqrt{2\pi}}) - \frac{h^2}{24\sigma^2} - \frac{\sigma_r^2 - (h^2/12)}{2\sigma_r^2} t^2$$

by ignoring terms with higher orders in *'h'* and:

$$t = \frac{(x-\mu_r)}{\sigma}$$

$$\Delta t^2 = 2h(x - \mu_r + \frac{h}{2}\sigma_r^2 \qquad \text{and}$$

$$\Delta \ln(y) = -h(\sigma_r^2 - h^2/12) - (x - \mu_r + h/2)/\sigma_r^4$$

That is, the graph of $\Delta\ln(y)$ against the mid-value of the class will be linear. If the 'x' intercept and the angle the line makes with the –ve direction of the x-axis are denoted, then the mean and standard deviation of the Gaussian component corresponding to this region are estimated first as a plot of $\Delta \log y(x)$ - $\log(y(x + h))$ - $\log(y(x))$ against *'x'*. Then the number of regions where the graph looks like straight lines with *–ve* –ve slope, indicate the number of components (under certain conditions). By connecting the points in the regions, straight lines can be filled for these regions. If θ_r is the angle it makes with the *x* axis and λ_r is the *x* intercept for the r^{th} region $r = 1,...,k$ for then the mean and variance of the r^{th} component is estimated as:

$$\mu_r = \hat{\lambda}_r + h/2$$

$$\hat{\sigma}_r^2 = (dh\,Cot\hat{\theta}_r / b) - h^2/12$$

where *'b'* and *'d'* denote the relative scales for *'x'* and $\Delta \log y(x)$ respectively. The proportions of the mixture can be estimated as:

$$\hat{p}_i = \hat{N}_i \Big/ \sum_{i=1}^{k} \hat{N}_i$$

where $\hat{N}_i$ is the total frequency of the i^{th} class and it is estimated by:

$$\hat{N}_r = \Sigma\, y(x) \,/\, \Sigma\, \hat{\mathrm{p}}_r \hat{N}_i$$

Here, the summation is restricted to the region under consideration and:

$$\hat{p}_{ri} = P(\frac{x + h/2 - \hat{\mu}_i}{\hat{\sigma}_i}) - P(\frac{x - h/2 - \hat{\mu}_i}{\hat{\sigma}_i})$$

where *P(x)* is the distribution function of standard normal variate.

Estimation of growth parameters

Once we have data on age and corresponding length obtained from the above procedure, we may use any one of the following methods as per the situation to estimate the growth parameters.

Gulland and Holt Plot: For small values of Δt, the required expression is:

$$\frac{\Delta L}{\Delta t} = K\,L_\infty - K\,\bar{L}_t \quad \text{where} \quad \Delta L = L_{t+\Delta t} - L_t \quad \text{and} \quad \bar{L}_t = \frac{(L_{t+\Delta t} + L_t)}{2}$$

By regressing $\frac{\Delta L}{\Delta t}$ on $\bar{L}_t$ (of the type $y = a + b\,x$) we can get estimates of the growth parameters as:

$$\hat{K} = -\hat{b} \quad \hat{L}_\infty = -\frac{\hat{a}}{\hat{b}}$$

Example: The first two columns of the following table pertain to the age and corresponding average length of fish of a cohort. The growth parameters can be estimated by calculations in the remaining columns and a followed regression. The steps followed are:

1. Generate column dL as the increment in length (difference of consecutive values of l(t)).
2. Generate column dt as the increment in age (difference of consecutive values of Age(t)).
3. Compute values in column dL/dt as the ratio of values in dL and dt).
4. Compute the mean length Lbar(t) as the average of consecutive values of L(t).

Age (t)	L(t)	dL	dt	dL/dt	Lbar(t)
1	25.7	10.3	1	10.3	30.85
2	36.0	6.9	1	6.9	39.45
3	42.9	4.6	1	4.6	45.20
4	47.5	3.2	1	3.2	49.10
5	50.7	2.1	1	2.1	51.75
6	52.8	1.4	1	1.4	53.50
7	54.2	-	-	-	-

Now regress the values in column dL/dt with the values in Lbar(t). That is, carryout regression analysis with values in column dL/dt as Y values and values in Lbar(t) as X values and obtain the regression coefficients *'a'* and *'b'*.

Regression statistics	
Multiple R	0.999922
R Square	0.999844
Adjusted R Square	0.999804
Standard Error	0.046844
Observations	6

	Coefficient	***Standard error***	***t Stat***	***P-value***
Intercept	22.36353	0.11182	199.9952	3.75E-09
Lbar(t)	-0.39163	0.00245	-159.872	9.18E-09

The estimates of coefficients in the regression model obtained through the regression analysis are a = 22.36353 and b = -0.39163 and the estimates of growth parameters are:

$\hat{K} = -\hat{b}$ = 0.39163 and $\hat{L}_\infty = -\frac{\hat{a}}{\hat{b}} = -\frac{22.36356}{-0.39163} = 57.1$

Ford-Walford Plot: The growth equation can be brought into the form:

$L_{t+\Delta t} = a + b\,L_t$ where $a = L_\infty\,(1-b)$ and $b = e^{-K\Delta t}$

When Δt is constant, we can get estimates of *a* and *b* by regressing $L_{t+\Delta t}$ on L_t and the estimates of growth parameters can be obtained as:

$$\hat{K} = -\hat{b} \quad \text{and} \quad \hat{L}_\infty = -\frac{\hat{a}}{\hat{b}}$$

Example: For the same set of data, the column L(t+1) is made with the next value of L(t). As per the Ford-Walford plot we regress the values in L(t+1) with values in L(t) and find the regression coefficients 'a' and 'b'.

Age (t)	**L(t)**	**L(t+1)**
1	25.7	36.0
2	36.0	42.9
3	42.9	47.5
4	47.5	50.7
5	50.7	52.8
6	52.8	54.2
7	54.2	-

Regression statistics	
Multiple R	0.999987
R Square	0.999974
Adjusted R Square	0.999968
Standard Error	0.039173
Observations	6

	Coefficient	***Standard error***	***t Stat***	***P-value***
Intercept	18.7018	0.074708	250.3308	1.53E-09
L(t)	0.672493	0.001713	392.5672	2.53E-10

The estimates of coefficients in the regression model obtained through the regression analysis are a = 18.7018 and b = 0.672493. Thus the estimates of growth parameters are:

$$\hat{K} = -\frac{\ln(\hat{b})}{\Delta t} = -\frac{\ln(0.672493)}{1} = 0.3968 \quad \text{and} \quad \hat{L}_{\infty} = \frac{\hat{a}}{(1-\hat{b})} = \frac{18.7018}{(1-0.672493)} = 57.1$$

Method of Chapman and Gulland: When Δt is constant, using the growth equation we can make the relation:

$$L_{t+\Delta t} - L_t = cL_{\infty} - cL_t \quad \text{where } c = 1 - e^{-K\Delta t}$$

Through a regression of $(L_{t+\Delta t} - L_t)$ on L_t we can arrive at a regression relation of the form and $y = a+bx$ using the estimates of coefficients of this regression equation we can estimate the growth parameters as:

$$\hat{L}_{\infty} = -\frac{\hat{a}}{\hat{b}} \quad \text{and} \quad \hat{K} = -\frac{\ln(1+b)}{\Delta t}$$

Example: For the given data first we generate a column with values L(t+1)-L(t) and regress these values on L(t) to obtain the constants a and b in the linear regression equation.

Regression statistics	
Multiple R	0.999945
R Square	0.999891
Adjusted R Square	0.999863
Standard Error	0.039173
Observations	6

	Coefficient	*Standard error*	*t Stat*	*P-value*
Intercept	18.7018	0.074708	250.33083	1.528E-09
Lbar(t)	-0.32751	0.001713	-191.182	4.49E-09

The estimates of ‘a’ and ‘b’ from the regression analysis are a = 18.7018 and b = -0.32751. The corresponding estimates of growth parameters are:

$$\hat{L}_{\infty} = -\frac{\hat{a}}{\hat{b}} = -\frac{18.7018}{-0.32751} = 57.1 \quad \text{and} \quad \hat{K} = -\frac{\ln(1+b)}{\Delta t} = \frac{-\ln(1-0.32751)}{1} = 0.3968$$

ELEFAN – Electronic Length Frequency Analysis

The first component ELEFAN-I in the system of ELEFAN is the program for estimation of growth parameters from length frequency data. It was first developed in 1978 and it consisted of (i) component for separation of samples into normally distributed components, (ii) estimation of growth parameters by generating the growth curve and (iii) minimizing the sum of squared deviations from the means of the component distributions. Later versions incorporated an algorithm which bypasses the sample separation step and fits the growth curve to peaks defined independently of any assumed underlying distribution.

- Data pre-processing: ELEFAN-I uses a simple high-pass filter to identify peaks and troughs in length frequency data. The high pass filter used is a running average over 5 classes which leads to the definition of peaks of the length frequency distribution that are above the corresponding moving average. Those below the corresponding running average are the troughs separating peaks.
- Steps involved in fitting of the growth curve in ELEFAN-I are:
    i. Calculate the maximum sum of points available in a set of length frequency samples. These are points which can be accumulated by one singe growth curve. It is termed as available sum of peaks (ASP).
    ii. Trace through the set of length frequency tables sequentially arranged in time for any arbitrary input of growth parameters L_∞ and *K*. A series of growth curves starting from the base of each of the peaks are then projected forward and backward in time to meet all other samples or the same sample again and again.
    iii. Accumulate points obtained by each growth curve when passing through the troughs separating peaks.
    iv. Select the curve which passes through most peaks and avoid most troughs and accumulate the largest number of points called 'Explained Sum of Peaks (ESP)'.
    v. Decrement or increment the values of L_∞ and *K,* until the ratio ESP/ASP reaches a maximum.

The growth model used in ELEFAN-I is the seasonally oscillating version of the generalized von Bertalanffy Growth Function (VBGF) of the form:

$$L_t = L_\infty[1-\exp(-KD(\mathrm{t}-t_0) + \frac{\mathrm{CKD}}{2\pi}\ Sin(2\pi(\mathrm{t}-t_s))]^{1/D}$$

where,

L_t is the predicted length at age t

L_∞ is the asymptotic length

K is the growth constant – stress factor by Pauly, 1981

D is another growth constant – termed as surface factor by Pauly, 1981

C is a factor that express the amplitude of the growth oscillations

t_0 is the age at which the fish would have had zero length

t_s sets the beginning of the sinusoidal growth oscillation with respect to t = 0

In ELEFAN-I, the model is used with two of the original parameters replaced (i) t_s with winter point WP and (ii) t_0 is described as a factor used to adjust a growth curve to an absolute age scale. Here a parameter "T0" is internally used to fulfil the role of t_0. Winter point WP designates the period of the year, expressed as a function of a year when growth is slowest. In northern hemisphere, WP is often found to be near 0.2 (February) while for the southern hemisphere, WP is often a value close to zero. The relation between WP and t_s is:

$$t_s + 0.5 = WP$$

When D = 1 and C = 0, the model will take the form of the normal VBGF used for fisheries research. When 0 < C < 1 growth oscillates seasonally, and when C > 1 growth oscillates strongly.

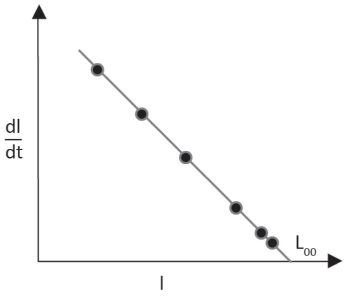
***

4.2

Von Bertalanffy Growth Parameters*

A growth curve is an empirical model of the evolution of a quantity over time. Some growth curves for certain biological systems display periods of ***exponential growth***. **A Gompertz curve,** named after Benjamin Gompertz, is a sigmoid function. It is a type of mathematical model, where growth in terms of weight, is slowest at the start and end of a time period. In biology, a growth model is a depiction of length or weight of animals as a function of age. In the case of fish populations, the study of growth is to determine the body size as a function of its age. The growth model developed by **von Bertalanffy** (1934) has been found to be suitable for the observed growth of most of the fish species. This model expresses length as a function of age of the animal.

Fish increases in length as they grow older but their growth rate which is the increment in length per unit time decreases as they grow old. When the rate of growth (dl/dt) is plotted against the length (l), in most cases it will look almost like a straight line with descending limb (negative slope). This line will cut the x-axis at a point where the rate of growth is zero. This is the point beyond which the fish will not grow further and the length of the fish at this point is known as the asymptotic length denoted by L_∞.

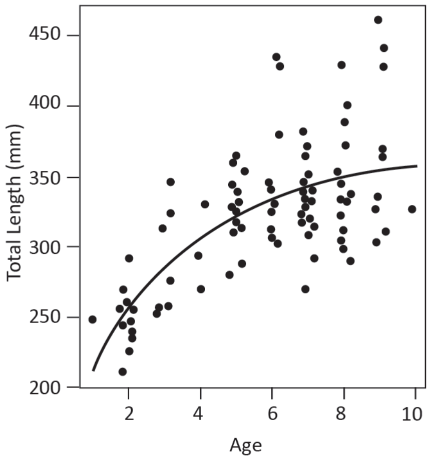


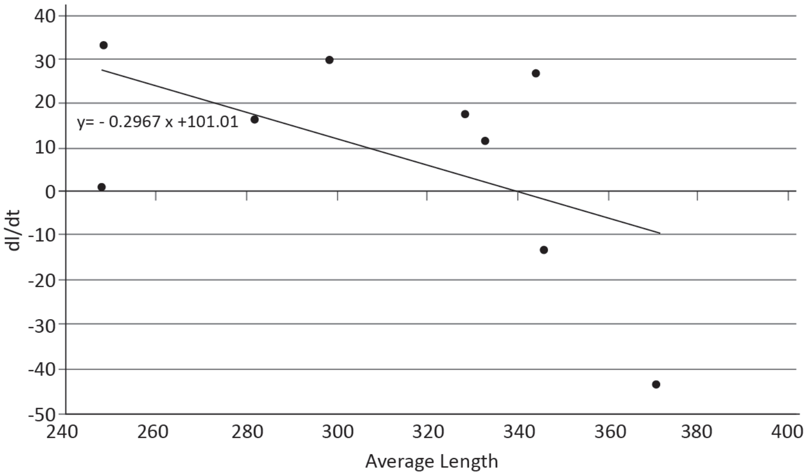
Example: Length (mm)-at-age of croakers. The figures show plots of the growth curve (upper panel) and growth rates (lower panel).

Avg Age	AvgLength
1	248.0
2	248.4
3	281.8
4	298.3
5	328.2
6	345.9
7	332.5
8	344.3
9	370.8
10	327.0

---

* *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*





To develop the growth model, the above phenomenon can be represented by means of a differential equation, as shown below:

$$\frac{dl}{dt} = K\,(L_\infty - l)$$

This can be rewritten as:

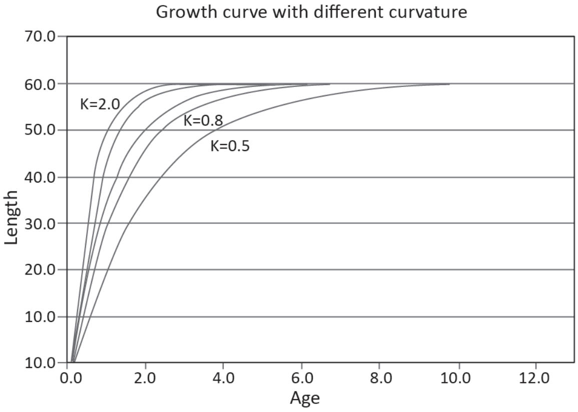
$$\frac{dl}{(L_\infty - l)} = K\,dt$$

The required growth model is then obtained by integrating the above differential equation to yield,

$$-\log(L_\infty - l) = K\,t + C$$

where *C* is a constant to be determined.

Expressing this equation for the length l we get,



$l = L_\infty - Ce^{-Kt}$

When $t=0$ the length ℓ also will be zero so that we get,

$0 = L_\infty - C$

Hence $C = L_\infty$ and we get the equation as:

$l = L_\infty(1 - e^{-Kt})$

But usually the length will be zero at a different point $t = t_o$ so that we get the solution for the constant as:

$C = L_\infty\, e^{K\, t_0}$

Hence the von Bertalanffy growth equation is obtained by substituting the above value for C as:

$l = L_\infty\, (1 - e^{-K\,(t - t_0)})$

Parameters of the model are *K*, L_∞ and t_o. Here *K* is the growth coefficient (termed as the curvature), L_∞ is known as the asymptotic length and t_o is the age at birth.

For different growth coefficients, the length-at-age changes as shown in the above plot.

***

4.3

Age Determination in Fishes Using Hard Parts[*]

Fisheries management relies on understanding the fish population dynamics by determining the biological parameters, including size at maturity, duration of spawning season, mortality estimates, age and growth. Accurate information on age of fish is an important pre-requisite for extracting precise information on growth, mortality, recruitment and other fundamental population parameters of fishes for stock assessment. The outcome of conventional age estimates using length frequency data depends upon the sample quality, selectivity of fishing gear, etc. The hard parts of the fishes grow with the fish and the growth process may leave some inscription on such parts and if that can be interpreted properly, we can get precise idea on growth. These inscriptions may result from either changes in the environment which the fish inhabits, or food availability, or physiological state of the fish. However, free swimming fishes always live in ideal conditions and do not leave any environment related markings in their skeletal structures. So interpretation of hard part inscriptions needs utmost care.

Ageing techniques

There are four approaches to age the fish.

i. Direct observation of fish in confinement or marking/tagging recapture technique

Direct observation of fish in confinement is the technique used by fish farmers. Tagging and marking experiments are conducted as the data collected are useful in estimating the population size, mortality rates and migration. Tagging does not enable individual fish to be aged unless the age of the fish at tagging is known. The method is very useful for fish living in areas where growth is continuous throughout the year. It is useful when large numbers of fish recaptured at annual intervals are available. However, cultivated or tagged fish seldom have the same growth rate as that of the wild or untagged fish. Tagging or marking of fish usually involves considerable time and recapturing is not assured.

ii. Injection (chemical marker) technique

Artificial time markers can be introduced into skeletal structures by injecting chemicals into the fish. The initial works were based on the use of lead acetate but this is toxic and tetracycline is now commonly used. It has the advantage of being an antibiotic drug, stable in solid form. Tetracycline is readily absorbed by vertebrates and deposited in bony structures where calcification is taking place. In teleost fish, tetracycline is laid down as a narrow ring timing the point of injection. The areas in which tetracycline is deposited in skeletal tissue appears fluorescent yellow under ultraviolet light, enabling them to be detected easily. However, this is not a popular technique.

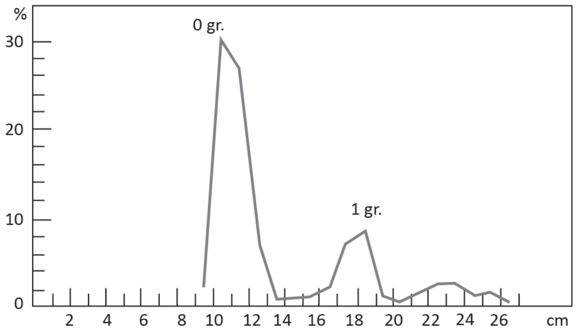
---

[*] *E.M. Abdussamad, Principal Scientist, Pelagic Fisheries Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

iii. Identification of cohorts based on length frequency data of fish

Length frequency data are used in various analytical, graphical and software assisted techniques to estimate the age, growth and other population parameters. The common methods employed are:

1. Petersen method



This is a simple method, fastest but most inaccurate method of ageing fishes. This method can be used only with species which have a restricted spawning season so that the fish bred in a single season can be identified as a single mode in a polymodal length distribution. The mode with the lowest value is identified as O-year group fish. Subsequent modes will be 1-year group, 2-year group fish and so on. The method can be very good for young fish but becomes increasingly less useful for older fish as the growth rate slows down and the modes merge. In practice, length-frequency distributions of fish caught over the shortest time period possible are plotted; the shorter the time period the more precisely the modes will be defined. A regular sequence of such length frequency distributions enables the progression of the modes to be followed.

2. Monthly modal progression analysis

Length frequency data collected at random from commercial and fishing surveys are used to estimate the age and growth of the fish.

3. Scatter diagram technique of monthly modal length

By plotting the monthly modal values of the length frequency data of fish as a scatter diagram, growth as well as the number of broods recruited per year can be estimated.

4. Bhattacharya method

This is a graphical method of splitting a composite distribution into separate normal distributions, *i.e.* when several age groups or cohorts of fish are represented in the same sample (For details consult FAO Fisheries Technical Paper No. 306.1 ,Rev.).

5. Probability paper/plot method

This method aims to resolve the normally distributed components of a length frequency distribution.

6. Age determination using hard parts of fish

Fishes grow continuously, but growth rate varies over time and season. These changes in growth rates may be reflected as zones or bands in the hard parts. By tracking down these inscriptions, the age of the fishes can be determined. Hard parts on which growth lines are found include scales, otoliths, opercular bones, spines, vertebrae, etc. During the slow growth phase, the rings/bands will be laid close together, whereas during fast growth phase they will be laid far apart.

Among skeletal structures, otoliths and scales are most widely used as they are easy to collect and store. The opercular bones of the head, pectoral and pelvic girdles, dorsal spine, etc. are also widely used.

Otoliths

There are three pairs of otoliths in teleost fishes. These are three-dimensional structures but do not necessarily grow at the same rate equally in all dimensions. If there is a pattern in the otolith it will be composed of a number of concentric shells with different radii. Depending on the amount of organic material in each shell or zone, its appearance will vary from extremely opaque to completely hyaline (transparent). For reading otoliths, it is preferable to identify and count the opaque zones, as characteristic growth patterns. Among the three, sagittal otoliths are generally used for age determination as they are the largest and easy to collect and process. They are located in the sacculus of the inner ear.

Scales

Scales vary in shape depending on the fish and body shape. Scales at the shoulder of the fish, between the head and the dorsal fin, are best suited for age determination. Scales are almost two-dimensional structures. The anterior part is formed of a series of sclerites, which should extend in a regular pattern from the centre of the scale. The structural discontinuities used for age determination result from irregularities in the pattern of the sclerites; they may be slightly distorted or they may be slightly closely spaced than the majority of the sclerites; usually the discontinuities are narrow and they are usually called ‘rings’.

Scales are thin structures and they need no preparation before viewing; the scales should be cleaned before they are stored. For reading, the slide with mounted scales is placed on the stage of a low-power microscope. The magnification used depends upon the size of the scale; in general, the lowest possible magnification is the best because it enables the whole scale pattern to be seen.

Finally, the age and growth estimates obtained under conventional methods can be cross checked and validated for correctness using the hard part inscription. However, detailed information on the biology and population dynamics of the fishes under question is an essential pre-requisite for any ageing work.

***

Chapter 5

Mortality Estimation

5.0

Mortality Estimation[*]

Fish as a natural resource follows most of the established behavior expected of by similar animals. The birth and ensuing recruitment, growth, reproduction and death, technically referred to as mortality are well defined phases of any animal's life and fishes are no exception to this. As is evident from its logic, fish populations increase in their abundance, popularly termed as biomass, by birth of animals or by growth apart from occasional immigrations. The loss of animals is mostly through death (mortality), which could occur due to ageing, natural mortality, or due to fishing; fishing mortality apart from the predation inflicted by larger animals in the sea. Hence mortality phenomenon happens to be the single most important cause of change in abundance of fish in any defined population or technically referred to as stock. By its sheer importance as the leveling force in face of animals with varying degrees of reproduction, mortality assumes an important position in the study of dynamics or fluctuations in the biomass of a given resource. Like its growth counterparts, mortality too has well laid conceptualization coupled with clearly defined procedures of measurement. Thus the phenomenon of rate of loss of animals in a particular population is a parameter to be estimated with the sampled animals. The measurements taken from the sampled individuals help an assessor to find out the composition of fish available at various ages and such information collected over a period of time will enable the observer to find out the rates at which fish of a particular age die due to natural and unnatural causes.

The inevitability of the mortality phenomenon can be understood by the fact that for a group of contemporarily hatched fish the number can only dwindle over time. The contemporaries or those individuals who were hatched almost at the same time are technically termed as cohorts. The phenomenon of mortality applies to each such group of cohorts and how they decline in number over time. To clearly delineate this process of decline in numbers, it is essential to follow the fate of the cohort. As mentioned earlier, cohort is a batch of fish of all of approximately the same age, and belonging to the same stock (Sparre and Venema, 1992). All fish of a cohort are assumed to have the same age at a given time so that they all attain the recruitment age at the same time. In the context of mortality, one is interested in the number of survivors from a cohort as a function of age. As mortality is split into natural and fishing induced ones, estimating the mortality entails the determination of total mortality (natural mortality + fishing mortality) first and then splitting this into natural and fishing mortalities, as appropriate.

The progress of a cohort over time is displayed Figure 1. (see page 80). In a cohort model, it is assumed that R individuals are recruited into the fishery at the age 't_r' (denoting age at recruitment). From this age, fish are exposed to some degree of natural mortality, M. After certain time these fishes are exposed to fishing at age 't_c' (age at first capture) denoted by F for fishing mortality. At some point (tmax), the older fish are not vulnerable to fishing.

---

[*] *J. Jayasankar, Principal Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

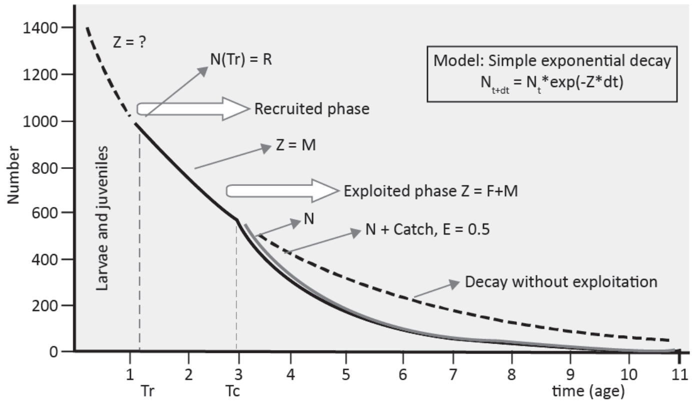


Fig. 1. Basic dynamics of the decay of a cohort

This setup assumes an all or none type of selection popularly referred to as knife edge selection, whereby at ‘t_c’ most of the age classes are vulnerable (Sparre and Venema, 1992).

Dynamics of a cohort

The dynamics of similarly aged fish of a stock are assumed to follow the model of natural decay, whereby the reduction in numbers due to total mortality is an exponential function of the number of cohorts at the beginning of the period. Notationally, the rate of change in numbers or number of losses or number of animals died in a small time is given by the following equation:

$$\frac{\Delta \mathrm{N(t)}}{\Delta t} = \mathrm{Z} * N\,(t)$$

where the deltas (Δ) indicate the change in numbers and a small interval of time, say one day or week, etc., Z is the coefficient of reduction or popularly known as rate of annual instantaneous mortality, usually scaled to account for one year. N(t) indicates the number of individuals alive at time t, preferably converted to years. This total mortality is supposed to be the arithmetic sum of natural mortality M and fishing mortality F. Notational depiction is as follows:

$$Z=M + F$$

A gentle mathematical juggling would yield the number of individuals alive at time t, which follows the time of recruitment of the cohorts into the fishery at Tr and could lead to an equation:

$$N(t)=N(Tr)*\exp(-Z(t-Tr)).$$

That is, the number of individuals available at the present time in years is a function of the difference between the time at recruitment and the present time and also the number of live individuals at the time of recruitment. Of course, this whole relationship assumes that the rate of instantaneous mortality Z remains constant throughout. As a corollary, it can be said that higher the value of Z, the faster would be the decline in numbers and the lower would be the maximum age.

As an example let us look at a case where the recruitment number N(Tr) of a cohort is 1,00,000 and total instantaneous mortality Z = 1.5 per year. Assuming Δ t as one day that is $1/365^{th}$ of a year then the number of survivors at different time intervals is given in Table 1.

Table 1. Number of survivors with advancing age

Age of cohorts (t years)	Time interval in years	Number of survivors (N(Tr)*exp(-Z*(t-Tr)))
tr	0	100000
tr+ one day	0.00274	99590
tr+0.2	0.2	74082
tr+0.3	0.3	63763
tr+0.4	0.4	54881
tr+0.4+one day	0.40274	54656
tr+1.5	1.5	10540
tr+3.0	3	1111
tr+5.0	5	55
tr+8.0	8	1

As it can be seen that the loss is very severe in the initial phases as compared to the last stage, like 8 years whereby the cohorts effectively vanish. The steepness depends upon the value of Z and it is better depicted in Figure 2.

As is evident, the decline is the steepest in the Z=2.5 case and it is the slowest in Z=0.5 case. The X axis indicates the time gap after Tr in years and the Y axis entries indicate the number of surviving animals.

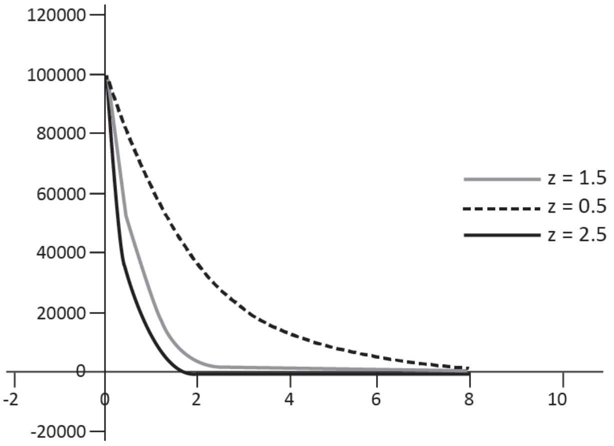


Fig. 2. Reduction in number of survivors with advancing age

After the animals obtain age of first capture, they are most vulnerable to fishing mortality, whereas up to the age of recruitment the decline in numbers is mostly due to predation or disease *i.e.*, natural mortality. Assuming that Z=F+M, the number of cohorts caught in a period from t1 to t2 in years is expressed as the function of total and fishing mortality as follows:

$$C(t1,t2)=F/Z*[(N(t1)-N(t2)]$$

Wherein N(t1) and N(t2) are the animals available at time periods t1 and t2. This equation is of extreme importance in fish stock assessment and is famously referred to as Baranov's equation or "Catch Equation". The fraction F/Z is also very important from assessment and management point of view and is popularly referred to as "Exploitation Rate". At the same period, the number of animals dying due to natural causes is:

$D(t1,t2) = M/Z*[N(t1)-N(t2)]$

The catch equation can be rewritten by involving the number of individuals at the beginning *i.e.*, t1 as follows:

$C(t1,t2) = N(t1)*F/Z*(1-exp(-Z(t2-t1)))$

One major assumption, which is the soul of this entire conceptualization is the fact that during the time interval (t1,t2) the situation at the ground is not fluctuating enough to influence the mortality rates, F and M. But criticisms are always possible on the count that natural mortality rates tend to differ with aging and younger fishes which are possibly smaller in size are less prone to fishing mortality as compared to their older counterparts.

Another conceptualization based on catch equation is the “Average number of survivors during the time period (t1,t2)”, which is given by:

$$\overline{N}\,(t1,t2) = N\,(t1) * \frac{1-e^{-z*(t2\text{-t}1)}}{Z * (t2-t1)}$$

Estimation of total instantaneous mortality (Z)

a) From catch rates

There are very many ways of estimating Z from the data collected from the fishery. One such method is based on catch rates or Catch Per Unit Effort (CPUE), which is the ratio of total quantity of fish caught to the total number of units of gear utilized to catch the same. When the fishes are caught with the same gear whose catchability coefficient (q) with respect to a particular resource is constant, the proportion of surviving members of the cohorts at two time periods (t1,t2) is equal to the ratio of the catch rates at the two time periods recorded by exploratory survey. That is:

$$\frac{N(\text{t}2)}{N(\text{t}1)} = \frac{CPUE\ (\text{t}1)}{CPUE\ (\text{t}2)}$$

A slight modification of the catch equation would lead to the following relationship when the number of cohorts available at the time limits *viz.,* N(t1) and N(t2) are known:

$$Z = \frac{1}{t2\ \text{-}\ t1} * \log(\frac{N(\text{t}1)}{N(\text{t}2)})$$

Using the previous two relationships, it can be derived:

$$Z = \frac{1}{t2\ \text{-}\ t1} * \log(\frac{CPUE\ (\text{t}1)}{CPUE\ (\text{t}2)})$$

When the data available pertains to commercial fisheries, where the time-series is on quarterly or annual basis, the equation used could be similar to the one described previously and the CPUE is calculated as the catch of cohort during the period (t1,t2) divided by the effort during that period. The catch rate can then be expressed as the product of average number of survivors in the period (t1,t2) and the catchability coefficient of the gear.

b) Heincke’s method

Assuming that mortality rate (Z) is constant throughout the life of an individual, the following equation holds based on certain algebraic norms:

$$Z = -\ln\,(\textstyle\sum_{t=1}^{\infty} N(t) / \sum_{t=0}^{\infty} N(t)$$

which is called the Heincke’s equation. In plain words, the mortality rate is the negative value of the ratio between the number of surviving individuals from age 1 to those surviving from age 0. Substituting CPUE at each year in the place of N(t)'s, this equation assuming that they are proportional the same reads as:

$$Z = -In \ \frac{CPUE(1) + CPUE(2) + CPUE\ (3\ and\ above)}{\mathrm{CPUE(0) + CPUE\ (1) + CPUE\ (2) + CPUE\ (3)}\ (3\ and\ above)}$$

c) Robson- Chapman method

Another estimate of Z is proposed by Robson and Chapman (Sparre & Venema, 1992) and the formula is:

$$Z = -In \ \frac{N(1) + 2*\ N(2) + 3*\ \mathrm{N}(3) + ...}{N(0) + 2N\ (2) + 3N(2) + 4(N3) + ...-1}$$

d) Linearised catch curve method

Ideally for estimating most of the parameters including the mortality rate, the type of data required is the number of sampled and raised animals belonging to a cohort at various age categories. However, in fishery sampling, age determination is a time and manpower consuming exercise and invariably aging is done by using the length of the animals sampled and their categories thereof. Here length is used as an alibi for age. Further, it is worth recalling that age and length are functionally linked through the Von Bertalanffy Growth Function (VBGF). Using the inversion of VBGF, length can be converted into age. The specific relationship is as follows:

$$\mathrm{t(L)Z} = \mathrm{t_o} - \frac{1}{\mathrm{K}} * \mathrm{In}\ (1- \frac{\mathrm{L}}{L_\infty})$$

where t(L) is the age at length L (cm or mm) and t_0, L_∞ and K are the classical VBGF parameters. Using this in the equation relating the logarithm of catch rate over a small time interval and the mid-time interval, which is as follows:

$$\mathrm{In}\ \frac{\mathrm{C(L_1,L_2)}}{\Delta t} = \mathrm{c} - Z * \mathrm{t}\left(\frac{\mathrm{L_1,L_2}}{2}\right)$$

which in turn can be rewritten using the catch and length information as:

$$\mathrm{In}\left(\frac{\mathrm{C(t,t+}\Delta t)}{\Delta t}\right) = \mathrm{c} - Z * (\mathrm{t}+\frac{\Delta t}{2})$$

Here the change in time:

$$\Delta \mathrm{t\ is\ given\ by}\ \frac{1}{\mathrm{K}} * \mathrm{In}\frac{L_\infty - L_1}{L_\infty - L_2}$$

Thus from this linear function, the total instantaneous rate of mortality can be estimated as the negative slope. It can be noted that in the previous equations 'c' is a term which is made of constant terms or in other words by terms which are not involving either time or length at different classes.

To put this linearised catch curve method into action, a plot of:

$$\mathrm{In}\ \frac{\mathrm{C}(L_1 - L_2)}{\Delta t} \quad \text{against} \quad \mathrm{t}\left(\frac{L_1 - L_2}{2}\right)$$

has to be made. Only the stable range of 't' values which are in the fully exploited range of the animal's life and which also is not close to t_∞ (age at maximum length of the animal) must be included for the computation of the coefficients of regression. This procedure is partially subjective, which must be given due care.

A worked out example of estimating total instantaneous mortality rate Z from length frequency data is given below.

The case is that of *Upeneus vittatus* from Manila Bay, Philippines (quoted in Sparre and Venema, 1992) and the length intervals and catch numbers of the pseudo cohorts is given in Table 3. The VBGF parameters are K=0.59 per year; L_∞ =23.1 cm; and t_0=0

To select the most appropriate portion of the length intervals a plot of:

$$\text{In}\ \frac{C(L_1-L_2)}{\Delta t} \qquad \text{against} \qquad t\left(\frac{L_1-L_2}{2}\right)$$

is made and in the above case it looks like this:

As is evident from the scatter, the first 7 observations and the last two observations do not follow the steady fall pattern and hence can be avoided (Fig. 3). Therefore, only the points in the mean time range 1.5 to 3.15 are considered for estimating the regression coefficients. In the present case, the estimated slope is -4.19433 and hence the estimated Z rate is 4.19.

Table 3. Catch (in numbers) at different length intervals

L1-L2	C(L1,L2)	t(L1)	Δ t	t(L1+L2)/2	ln(C(L1,L2)/Δt)
6-7	3	0.51	0.102	0.56	3.381395
7-8	143	0.612	0.109	0.665	7.179252
8-9	271	0.721	0.116	0.778	7.756284
9-10	318	0.837	0.125	0.898	7.841493
10-11	416	0.961	0.135	1.027	8.033166
11-12	488	1.096	0.146	1.168	8.114464
12-13	614	1.242	0.16	1.32	8.252576
13-14	613	1.402	0.177	1.488	8.14997
14-15	493	1.579	0.197	1.675	7.825061
15-16	278	1.776	0.223	1.884	7.128205
16-17	93	2	0.257	2.123	5.891279
17-18	73	2.257	0.303	2.402	5.484482
18-19	7	2.56	0.37	2.735	2.940162
19-20	2	2.93	0.474	3.151	1.439695
20-21	2	3.404	0.66	3.702	1.108663
21-22	0	4.064	1.096	4.525	-
22-23	1	5.16	4.064	6.188	-1.40217
23-24	1	-	-	-	-

e) The cumulated catch curve method

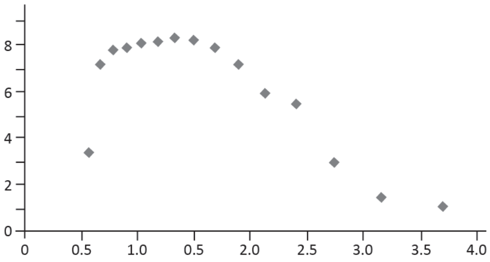


Fig. 3. Estimation of Z by catch curve method (x=Age in years; y=(*l* n c/Δt)

Another approach to estimate Z from length frequency data is the ‘Cumulated Catch Curve’ method propounded by Jones and Van Zalange. The main difference here is in the time range (t1,t2) and t2 is assumed to be very large to be near ∞ and that would lead the linearised catch curve equation to become:

LnC(t,∞)=d- Z*t, where C(t,∞) is called cumulated catch curve equation. Then the Jones and Van Zalinge equation for length converted catch curve would be:

$$\text{In}\ (\text{C}(L,\ L_{\infty})) = \text{a} + \frac{\text{z}}{\text{k}}\ \text{In}(L_{\infty} - L)$$

After selecting the appropriate portion of the scatter between:

$$\text{In}\ (\text{C}(L,\ L_{\infty})) \quad \text{and} \quad \text{In}(L_{\infty} - L).$$

From the slope, Z is estimated as slope * K. For the previous example, the cumulated catch curve approach is done as shown in Table 4 below:

Table 4. Cumulated catch curve approach

L1-L2	C(L1,L2)	C(L1,L_{∞})	ln(C(L1,L_{∞}))	ln(L_{∞}-L1)
6-7	3	3816	8.246958	2.839078
7-8	143	3813	8.246172	2.778819
8-9	271	3670	8.207947	2.714695
9-10	318	3399	8.131237	2.646175
10-11	416	3081	8.033009	2.572612
11-12	488	2665	7.887959	2.493205
12-13	614	2177	7.685703	2.406945
13-14	613	1563	7.354362	2.312535
14-15	493	950	6.856462	2.208274
15-16	278	457	6.124683	2.091864
16-17	93	179	5.187386	1.960095
17-18	73	86	4.454347	1.808289
18-19	7	13	2.564949	1.629241
19-20	2	6	1.791759	1.410987
20-21	2	4	1.386294	1.131402
21-22	0	2	0.693147	0.741937
22-23	1	2	0.693147	0.09531
23-24	1	1	0	-2.30259

The plot based on the last two columns is shown here. From the plot it can be seen that the most appropriate range to be used for estimating the slope is the x range from 1.41 to 2.31 and the corresponding slope value is 6.51. Hence the estimated Z rate = Slope*K= 6.51* 0.59= 3.84.

f) Beverton and Holt's Z-equation based on length data

Beverton and Holt (Sparre and Venema, 1992) have shown that there exists a functional relationship between Z and the average length of fish $\overline{L}$, which is given by

$$Z= K * \frac{L_\infty - L_2}{L - L^1}$$

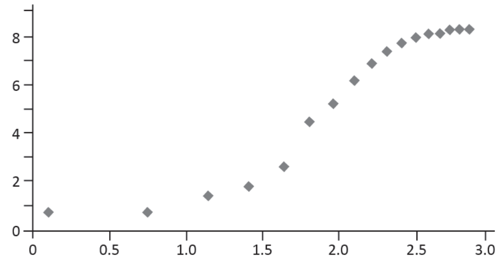


Fig. 4. Cumulated catch cure plot (x = l n (L_∞ - L_1)

where L' is some length for which all fish of that length and longer are under full exploitation and it is the lower limit of the class interval of lengths from which point full exploitation is presumed.

For example, if the VBGF parameters of a cohort are K=0.45 per year and L_∞ = 100 cm and if it is assumed that L'=45 cm then the Z estimates for the following data are given in Table 5.

Table 5. Z estimates

Length Group	Mid Length	N(1960)	N(1970)	N(1980)	N(1960)*Mid Length	N(1970)*Mid Length	N(1980)*Mid Length
45-50	47.5	256	268	212	12160	12730	10070
50-55	52.5	237	226	161	12442.5	11865	8452.5
55-60	57.5	211	180	116	12132.5	10350	6670
60-65	62.5	187	141	79	11687.5	8812.5	4937.5
65-70	67.5	161	105	52	10867.5	7087.5	3510
70-75	72.5	138	76	31	10005	5510	2247.5
75-80	77.5	113	50	17	8757.5	3875	1317.5
80-85	82.5	87	30	8	7177.5	2475	660
85-90	87.5	62	15	3	5425	1312.5	262.5
90-95	92.5	36	6	1	3330	555	92.5
95-100	97.5	12	1	0	1170	97.5	0
	Sum	1500	1098	680	95155	64670	38220
	$\overline{L}$				63.43667	58.898	56.20588

Where N(1960) indicates numbers caught in year 1960 and so on. The mean length here is a weighted average of the lengths detailed.

Based on these figures, the Z values for various years are as follows:

Z(1960)= $0.3 * \frac{100\text{-}63.44}{63.44\text{-}45} = 0.6$ per year

Z(1970)= $0.3 * \frac{100\text{-}58.90}{58.90\text{-}45} = 0.9$ per year

Z(1980)= $0.3 * \frac{100\text{-}56.21}{56.21\text{-}45} = 1.2$ per year

g) Power- Wetherall method

As a special application of the Beverton- Holt's Z- equation, it can be expressed that

$\overline{L}$=a + b*L' where Z/K=-(1+b)/b and L_∞=-a/b or alternatively

b=-K/(Z+K) and a=-b* L_∞

This means that plotting $\overline{L}$ -L' against L' gives the estimates of a and b and from them the parameters L_∞ and Z can be estimated.

h) Pauly's empirical equation for natural mortality estimation

Pauly (Sparre and Venema, 1992) made regression analysis to functionally link natural mortality M with VBGF parameters and climatic parameters and the empirical formula arrived by him is given below:

Rate of natural mortality per Year (M)= -0.0152-0.279*ln L_∞+0.6543*lnK+0.463*lnT

where T is the average annual temperature at the surface in degrees centigrade.

Table 6 gives the estimates of natural mortality for various combinations of T and VBGF parameters.

Table 6. Estimates of natural mortality

T=5°C					T=25°C			
L_∞ (cm)	K=0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0
10	0.24	0.7	1.1	1.7	0.51	1.5	2.3	3.6
80	0.14	0.38	0.6	1.0	0.29	0.8	1.3	2.0
200	0.10	0.30	0.47	0.7	0.22	0.6	1.0	1.6

Method of computing

The above discussed methods of estimating rates of mortalities can be implemented practically either by manual means (highly exhausting) or by using computer based spread sheets or by software, custom made for this purpose.

References

Sparre, P. and Venema, S.C. (1992) Introduction to Tropical Fish Stock Assessment. Part 1. Manual, FAO Fisheries Technical Paper No. 306.1, Rev. 1., Rome. 376 p.

***

Chapter 6

Recruitment and Selectivity

6.0

Recruitment and Selectivity*

It is possible to infer that the natural mortality coefficient does not remain constant between ages. There are evidences to suggest that the fishing mortality coefficient (F) too varies with age. There are three factors, which are involved with the variation of F with age. The first is the selective action of the meshes of the fishing gear and its consequent effect on the recruitment; the second is the probability of the capture and escape of fish; and the third is the heterogeneous fishing effort, *i.e*. exploitation by multifleet, each using different gear/ mesh size. These three factors, either independently or in combination, often introduce sudden changes in the fishing mortality when certain ages are reached. The interaction between fish growth (and consequent body size) and mesh size of the gear determines the recruitment of fish and gear selectivity. In this context, it is easy to distinguish between recruitment and selection. Recruitment is a characteristic of the fish, whereby the young fish previously inaccessible to the fishing gear, become vulnerable due to growth. Selection is a characteristic of fishing gear and the way it is operated.

The relationship between the proportion of fish retained and the length of the fish is referred to as selection curve. The trawls, for example, exploit all the fish which are larger than the length that the mesh can retain; the proportion of retention increases from 0 percent at the length zero and reaches 100 percent for the larger fish. On the other hand, the gillnets and hooks & line are selective for a certain length range, thus capturing only those size groups, which the mesh can retain and thereby excluding the capture of very small and very large fish. Therefore, mesh size of a gear determines the size at which the fish are caught. A change in the mesh size will have two effects. First will be a change in the size of fish, and hence in the age, at which they are first retained by the net; and the second will be a change in the fishing mortality. For instance, the probability of capture, and hence F is zero for *Saurida tumbil* below 80 mm length if the codend mesh size of the trawl net is 15 mm; for the larger codend mesh of 30 mm, the probability of capture and F are zero for fish of 120 mm and below.

Estimation of recruitment and gear selectivity serves different purposes related to fisheries management and fish stock assessment. It is an important tool for determining the optimum mesh size for the fishery, since gear selectivity is strongly related to the estimation of total mortality and prediction of future yields. Moreover, estimation of population densities, the variable selection probabilities for different sizes need to be assessed and corrected.

The lengths at which 50 percent of fish are retained (and 50% escape) are referred to as the 50 percent lengths of first capture, or simply as the length at first capture, ‘lc’. In other words, the ‘lc’ can be regarded as the average length at which a fish first becomes subject to exploitation. The selection factor (SF) is defined as the ratio of the 50 percent retention length (lc), to the mesh size (MS), *i.e.* $SF = lc / MS$. For example, for a fish with ‘lc’ of 120 mm in 30 mm mesh size, the $SF = 120/30 = 4.0$. If the SF has been determined for a given species and a particular design of the gear, it can be used to calculate the ‘lc’ for a given mesh size as $lc = SF * MS$. The mesh size for a given ‘lc’ can also be calculated as $MS = lc / SF$.

---

* *E. Vivekanandan, National Consultant, World Bank/GEF Ocean Partnership Project, Bay of Bengal Programme Inter-governmental Organisation, 91, Saint Mary’s Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.*

The length range l_{25} to l_{75} corresponding to 25 percent and 75 percent retainment, respectively is called the selection range. It is possible to determine approximate selection factors and selection range for a variety of species by taking into account their shape. Generally, the selection factor is influenced by the depth ratio of the fish. The depth ratio is the body length of fish (excluding caudal fin)/maximum depth of body. Since the body depth increases in proportion to the body length, the body length also could be used to determine the SF. From the first approximation on the relationship between body depth and SF, or between body length and SF, the selection factors for the other species could be determined. Since it is much easier to determine depth ratios and body length than to determine the SF for all species experimentally, the SF for all the species can be determined indirectly by following this method.

Estimation of selectivity by different gears

Estimation of trawlnet selection

Selectivity of trawl can be expressed as the proportion of fish at each length entering the net, which are retained. As young fish grow and are recruited into the selection range, they suffer only a low fishing mortality. As they grow larger, their chances of escaping from the net becomes less, until they are too large and cannot not escape at all. At this stage they are subjected to maximum fishing mortality. The selection probability in trawl type gears takes S-shaped sigmoid selection curve, which is referred to as a selection ogive.

The fine meshed end of the trawl net where the catch is collected is called as the codend. The mesh size of the codend determines, to a large extent, the selectivity of the trawl. The mesh size (MS) is usually defined as the length of the stretched whole mesh. There are three methods to determine the selectivity of trawls by conducting experimental trawling, *viz.*, covered codend method, alternate haul method and trouser codend method. In the covered codend method, the quantity and size of fish that escapes through the meshes of the codend can be determined by covering the codend with a much larger bag with finer meshes. The selectivity of the gear can be determined by comparing the size of the fish in the codend with those of the fish in the cover. The trawl selection ogive can be expressed in the form of a straight line by the following equation:

$$\ln(1/S_L-1) = S_1-S_2 * L \qquad \text{(Eq. 1)}$$

where S_L = number of fish of length L in the codend / number of fish of length L in the codend and in the cover; L is the mid-length of the length group; and S_1 and S_2 are a and b, respectively in the regression equation.

The length at which 50 percent of the fish are retained in the codend, L50% = S_1/S_2. In the example given in Table 1, 46-50 cm is the length group at which 50 percent of fish are retained in the codend mesh size of 1.5 cm. The selection factor SF can be calculated as follows: SF = 48 (mid-length)/1.5 = 32. If the mesh size is increased to 3.0 cm, the L 50% would be higher than 48 cm (let us assume 60 cm). This implies that if the codend mesh size is 1.5 cm, fish of length 45 cm and above are retained by the net; and, if the mesh size is 3.0 cm, fish of length 60 cm and above are retained.

Table 1. Hypothetical example showing the numbers caught in the codend and in the cover

Length group (cm)	Codend	Cover	Percent retained
21-25	25	1800	1.3
26-30	31	950	3.2
31-35	42	450	8.5
36-40	56	250	18.3
41-45	78	125	38.4
46-50	82	82	50.0
51-55	96	55	63.6
56-60	84	28	75.0
61-65	72	12	85.7
66-70	60	2	96.8
71-75	35	0	100.0
76-80	22	0	100.0

In the alternate haul and trouser codend methods, two codends with different mesh size are used. In the alternate haul method a series of hauls are made, alternating the two different mesh-sized codends. In the trouser codend method, both the codends are fitted side by side. An example showing the numbers of the ribbonfish, *Trichiurus lepturus* caught per 10 hrs of alternate hauling using 1.5 cm and 3.0 cm codend mesh size is shown in Table 2 (on page 92). The C_2/C_1 values in this example shows that the ratios do not tend to be constant for the larger fish. This situation normally happens in practice probably because gear efficiency may be affected by the codend mesh size. When this happens, it is necessary to adjust the C_2/C_1 values by considering that any gear efficiency effect applies equally to all lengths and can be measured directly by the ratio of catches for the largest fish. In this example, it is considered that the length group 66-70 cm and above are retained 100 percent in the larger mesh size (3.0 cm). By using the calculation procedure given in Table 2, it can be noted that the values of C_2/C_1 are approximately constant for fish of 66-70 cm length group and greater.

Estimation of gillnet type selection

Gillnets are the most selective gears used in commercial fishing. They are passive gears, *i.e.*, the fish have to swim into the net to get caught. The length composition of the gillnet catch will be largely determined by the selectivity of the net. The selection curve for gillnets is usually bell-shaped and takes a descending slope on the right hand side. Small fish can pass through the meshes and large fishes avoid getting caught because their body is larger than the meshes. A larger mesh may not exert the same mortality on the optimum length Lm (the top of the bell-shaped selection curve) of fish, as does a net with smaller mesh. Since active swimming is a prerequisite to encounter gillnet, and the larger fish are expected to be more

Table 2. Number of the ribbonfish, *T. lepturus* caught per 10 trawling hours by using 1.5 cm and 3.0 cm mesh codends

Length group (cm)	1.5 cm C_1	3.0 cm C_2	C_2 / C_1	Adjusted values** of C_2 / C_1
21-25	25	1	0.04	0.03
26-30	31	6	0.19	0.14
31-35	42	24	0.57	0.41
36-40	56	50	0.89	0.64
41-45	78	75	0.96	0.69
46-50	82	97	1.18	0.85
51-55	96	120	1.25	0.90
56-60	84	112	1.33	0.96
61-65	72	101	1.40	1.01
66-70*	60	83	1.38	0.99
71-75*	35	49	1.40	1.01
76-80*	22	30	1.36	0.98

* Length groups with 100% retention for the larger mesh size (3.0 cm)
Total numbers caught in the last 3 length groups:
for codend C_1 = 117; for codend C_2 = 162
$C_1 / C_2 = 117 / 162 = 0.72$
**C_2 / C_1 * 0.72

active and move over longer distances, their probability of encounter will be larger. Therefore, the peak in the selection curve of a larger mesh will be higher. These effects make estimation of gillnet selection more difficult than estimating trawl selection.

In 1963, Holt suggested a method for determining the selection curve of gillnet by using two gillnets with different mesh sizes, MS_1 and MS_2. The two nets are set to fish in the same area at the same time and the numbers caught in each length group are observed. It is assumed that the probability functions of the two selection curves have the same standard deviation. As in the case of trawl selection, it is also assumed that Lm is proportional to the mesh size. Hence Lm = SF * MS. In this situation, the logarithms of ratios of the catches by the two nets should be linearly related to fish length (L), according to the following relationship:

y = a + bL, where y = logarithm of ratio of catches at length L.

From the parameters a and b, it is possible to estimate the SF from the relationship:

$$SF = - 2 a / b (MS_1 + MS_2) \qquad \text{(Eq. 2).}$$

The lengths of maximum retention (L1 and L2) for the two nets are:

$$L_1 = SF*MS_1; \; L_2 = SF *MS_2.$$

If the natural logarithms of C2 / C1 ratios are plotted against the mid-length, L given in Table 3 (see page 93), the result is a straight line with a slope of 0.1081 and an intercept of – 6.0944.

Table 3. Hypothetical example showing the numbers (C_1 and C_2) caught per unit fishing time by gill nets with different mesh sizes (10 cm and 12 cm)

Length group (cm)	Mid-length (cm) (X)	10 cm C_1	12 cm C_2	C_2 / C_1	ln C_2 / C_1 (Y)
31-35	33	8	1	0.13	-2.04
36-40	38	21	3	0.14	-1.97
41-45	43	48	10	0.21	-1.56
46-50	48	65	15	0.23	-1.47
51-55	53	81	36	0.44	-0.82
56-60	58	56	84	1.65	0.50
61-65	63	21	54	2.57	0.94
66-70	68	15	48	3.20	1.16
71-75	73	5	38	7.60	2.03
76-80	78	2	20	10.00	2.30

$b = 0.1081$; $a = - 6.0944$

The estimate of selection factor is then:

$$SF = (2 * 0.6944)/(0.1081 * (10 + 12)) = 5.13.$$

The lengths corresponding to the maximum retention lengths are:

For gear 1 = $L_1 = 5.13 * 10 =$ 51.3 cm

For gear 2 = $L_2 = 5.13 * 12 =$ 61.6 cm.

The standard deviation:

$$SD = ((L_2 - L_1) / b)^{0.5} = ((61.6 – 51.3)/0.1081))^{0.5} = 9.8 \text{ cm}$$

When the selection curve of gillnets has been determined, population analysis is simple. For this, two selection lengths, ‘lc’, the length at which the fish enters the selection range, and ‘ld’, the length at which they grow out of it, are needed.

Gillnet selectivity is affected by several factors; such as mesh size, elasticity of the twine, hanging ratio, shape, size and behaviour of fish and the method of fishing.

Selection by other gears

Like the trawl (active gear) and gillnet (passive gear), other types of gears are also to certain extent selective. However, the selection properties for other types of gears have not been fully understood. For hooks & line fisheries, some authors report that the selection curve is bell-shaped as in the case of gillnets. A given size of hook will efficiently capture fish that are big enough to get the hook into their mouth, yet not so big that the hook cannot hold it.

In the case of traps, the selection depends on the size and shape of the gear and may vary with the characteristics of the population. The selection of funnel-shaped traps may resemble the retainment in the codend of a trawl, and hence, S-shaped selection curve. Small traps may accommodate only one individual territorial fish of particular size. In large traps, presence of large predators will keep small prey fish from entering the gear. If smaller prey first occupies the trap, it would attract large predators, which would predate the prey before the trap is hauled.

The pattern of selection can often be changed by making adjustments to the gear.

Resultant effect of recruitment and gear selection

In cases where there is mass spawning by fish school, such as those of small pelagics, the recruitment in terms of length can be quite sharp. In the extreme case, the recruitment approaches knife-edge selection at a threshold length 'lr'. Fish below the length 'lr' suffer no fishing mortality at all but on reaching 'lr', they are suddenly exposed to high fishing mortality. If selection and recruitment are occurring over the same range, then the effective selection, *i.e.*, the proportion P_1 of the full fishing mortality to which the fish of a given size in the stock are exposed, will be as follows:

$$P_1 = r_1 * S_1 \qquad \text{(Eq. 3)}$$

where r_1 = proportion recruited, and S_1 = proportion of those entering the net which are retained. Thus, a resultant curve is obtained as the product of the recruitment and mesh selection curves. The resultant curve expresses the effective entry of fish to the catch and the resultant mean selection length can be determined from the curve.

Other methods of measuring selection curves

Estimation of selection ogive from a catch curve

In the linearised catch curve method to estimate mortality, the left hand side of the curve is discarded because the juvenile fish are not fully exploited (Refer Chapter 5). To estimate how many fish are not exploited at each age, and how many would have been present in the population, the straight line, which is used to estimate the total mortality coefficient, should be extrapolated. The differences between the expected and actual numbers should give the ogive (curve) resulting from the combined effect of recruitment and mesh selection. In this method, it is assumed that the total mortality rate is constant for all ages. Hence, the results obtained from this method should be treated with reservations. Nevertheless, since the calculations are easily performed, the method has gained significant popularity.

Estimation of selection curves from cohort analysis

The cohort analysis produces estimates of F values for different length groups (Refer Chapter 8). These F values can be used for determination of gear selectivity and recruitment curve. They used the following equation to any gear or combination of gears combined with recruitment curve:

$$S(i) = F(i) / MAX(F(j)) \qquad \text{(Eq. 4)}$$

where F(i) is the fishing mortality for size group ‘i’, and MAX (F(j)) is the maximum value of F among all size groups. This equation can be applied to gears without mesh, such as hooks & line.

Computer programmes for estimation of selectivity and recruitment

ELEFAN II, which is a collection of auxillary routines, estimates the selection pattern given the probability of capture for each length class using a logistic transformation. This routine estimates L_{25}, L_{50} and L_{75}. This is one of the few routines in the ELEFAN package that may be used without the need to access the ELEFAN data file. The data needed to run this routine is the smallest midlength, largest midlength, class interval and the probability of capture for each length class.

A recruitment pattern is obtained by projecting a set of length frequency data backward into a one-year time axis. Recruitment patterns can be used to infer the number of recruitment pulses occurring per year. However, if the parameter t_o is not given, the exact time of recruitment cannot be determined. A routine is built into the programme, which automatically fits a recruitment pattern with one or two normal distributions when the recruitment explained by one or two normal curve ranges between 90 percent and 110 percent of the observed annual recruitment. The routine functions in the same manner as the catch curve routine, for which ELEFAN data file and growth parameters ($L\infty$, K, C, and WP) must be provided.

In the LFSA package, the routine CCURVE (length-converted catch curve analysis) estimates the gear selection ogive from the catch curve.

In the FiSAT package, the routine ASSESS has the sub-routines for estimation of recruitment pattern and probabilities of capture by different mesh sizes of trawls and gillnets. The sub-routine on prediction also provides estimates of knife-edge selection and selection curves.

***

Chapter 7

Exploratory Survey for Biomass Estimation

7.0

Exploratory Survey for Biomass Estimation[*]

Estimates of biomass and annual yield can be derived from bottom trawl surveys, especially for monitoring demersal fish stocks. But the estimation of total biomass from this based on catch per unit effort estimates involves some crucial assumptions. The mean catch per unit area is an index of the stock abundance. This is on the assumption that it is proportional to the abundance. Using swept area method this index of stock abundance can be converted into an absolute measure of biomass.

The objectives of bottom trawl survey are:

- Estimation of the total biomass and catch rates.
- Estimation of biomass of selected species.
- Collection of biological data such as length frequency data for estimation of growth and mortality parameters.
- Collection of environmental data.

The bottom trawl is a conical net bag with wide mouth fitted with weights on the ground rope and floats on the head rope. The net is kept open by tow otter boards which are wooden or iron structures towed by the warps attached forward of their centre so that they tend to diverge. These may be very long and sweep the sea bed over a wide area. They frighten the fish towards the advancing net and increases its effectiveness. The shape of the net varies depending on the kinds of fish targeted and the types of bottom. The ground rope is fitted with roller gear so that the trawl can be used on stony bottom without any damage. The tail end of the gear from which the captured fish are removed is called the codend where most of the size selection takes place. In order to obtain a representative sample of all the size ranges of the species the mesh size should be relatively small at the codend.

For estimation of stock sizes a completely randomized design or a stratified random sampling design is preferred and in most cases stratified sampling design is preferred. Strata are constructed in accordance with the density distribution of the fish so that areas with high/medium/low densities are separated. For stratification some prior information is required which is obtained in a first survey following simple random sampling design and the variability obtained is used for stratification. The distribution of hauls within strata should be random taking into account the practical difficulties. The number of hauls possible in a given period can be calculated as:

Number of hauls per day = T / (t2+t3+t4)

where T is the number of hours available per day, t2 is the duration of one haul, t3 is the time used for shooting and hauling the trawl and t4 is the average time taken to cover distance between stations. It is important to standardize the length of the haul throughout the survey, since the catchability of species and sizes often depends on the duration of haul. Following are the important points to be remembered while recording data from a trawl survey:

---

[*] *T.V. Sathianandan, Principal Scientist & Head of Division, and* Grinson George, *Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

- The objective of the survey determines the data items to be recorded, *e.g.* biomass estimation, length frequency analysis, mortality estimation.
- Data items include specification of gear, haul duration, position at start and end of haul, wire length, wing spread, bottom type, depth, etc.
- Catch record should include total weight, species composition, length frequencies for selected species.
- Data should be well organized to facilitate processing.
- There should be a log summarizing the whole cruise.
- There should be fishing log that provides information on vessel's position, time of start, end of haul gear rigging, etc. Summary information on catch should also be recorded in the fishing log.
- Detailed information on catch in terms of length, weight, sex, maturity stage, etc. for each specimen should be recorded along with length frequency distributions.

Swept area method

Trawl sweeps a well defined path, the area of which is the length of the path times the width of the trawl which $D = v \times t$ is called the swept area. It is estimated as:

$$a = D \times h \times X2$$

Where, 'v' is the velocity of the trawl over the ground when trawling, 'h' is the length of the head-rope, 't' is the time spent for trawling and X2 is the fraction of the head-rope length, 'h', which is equal to the width of the path swept by the trawl and the wing spread is $h \times X2$. Different values of X2 in use are 0.4 to 0.6 for Southeast Asian bottom trawls, 0.5 as a compromise suggested by Pauly and 0.6 in the Caribbean suggested by Klima. Catch per unit area estimated by dividing the catch by the swept area is used for the estimation of biomass. When exact positions of the start and end of the haul are available, the distance covered in nautical miles is estimated as:

$$D = 60 \times \sqrt{(Lat1 - Lat2)^2 + (Lon1 - Lon2)^2\, Cos^2\,(0.5(Lat1 - Lat2))}$$

where Lat1, Lat2 are the latitude at start and end of haul in degrees, Lon1,Lon2 are longitude at start and end of the haul in degrees. When the velocity of the vessel and its course together with direction and speed of the current are available, then the distance covered per hour is calculated as:

$$D = \sqrt{VS^2 + CS^2 + 2VS \times CS \times Cos(dirV - dirC)}$$

where VS is the velocity of the vessel in knots (nautical miles per hour), CS is the velocity of current in knots, dirV is the course of vessel in degrees and dirC is the direction of current in degrees.

If cw is the catch in weight of a haul and 't' the time spent in hauling (in hours), the cw/t is the catch in weight per hour. If 'a' is the swept area then a/t is the swept area per hour. Then the catch per unit of area is obtained as:

$$CPUA = \frac{cw/t}{a/t} = \frac{cw}{a} \; kg/\, nm^2$$

If X1 is the fraction of the biomass in the effective path swept by trawl, which is actually retained in the gear and $\overline{cw/a}$ is the mean catch per unit area of all hauls, then an estimate of the average biomass per unit area is:

$$\bar{b} = \frac{\left(\overline{cw/a}\right)}{X1} \; kg / nm^2$$

Let A nm^2 be the total area under investigation, then the estimate of total biomass for this area is obtained as:

$$B = \frac{\left(\overline{cw/a}\right)A}{X1} \; kg$$

An example of biomass estimate from commercial trawl data off Saurashtra coast in western India is given here. A trawler (overall length: 17.5 m) conducted fishery survey during 1985-1989. During the 5-year period, the survey was conducted in eighty-eight 10’ squares between the latitude zones 20°N and 70°E (off Veraval) and 23°N 68°E (off Jakhau) at depth range of 12 to 70 m. The area of each 10’ square in the survey area was considered as 326.6 km^2.

The total area considered (A) for the survey was estimated as (326.6 * 88) 28,740.8 km^2.

The area swept (a) by the gear during one hour of trawling was calculated considering the trawling speed (v) as 2.5 knots/h (= 4.3 km/h), the headrope length (h) of the trawlnet as 24 m, and X^2 as 0.5. The area swept was calculated as 0.052 km^2/h for the entire period of the survey.

The biomass was calculated by pooling the catch from each 10’ square during the 5-year period. The total catch was 205.2 t and the CPUE was 43.9 kg/h.

Biomass = (43.9 * 28740.8)/(0.052 * 0.5) = 48528 t

Density = Biomass/Area considered

= 48528 / 28740.8 = 1.688 t/km^2.

Precision of the estimate of biomass in the swept area method can be achieved by increasing the number of hauls. Another way of increasing precision is to apply stratified sampling by considering depth and bottom type. Suitable stratification may improve precision for the same number of hauls.

However, estimation of biomass and density from the CPUE involves several crucial assumptions, such as (i) the CPUE is proportional to the biomass abundance, and (ii) the proportion of retainment in gear, etc. It has been observed that for some stocks, the observed CPUE is only related to stock size, and in such cases, there may be no CPUE data that are satisfactory. For example, the CPUE from purseseine fisheries for shoaling pelagics may lead to erroneous estimates. For pelagic trawling, avoidance can be very great. Moreover, survey by any gear provides an estimate of only the target stocks of that gear and not the total biomass of the considered area. Due to this reason, the biomass is usually underestimated by the swept area method. This method has its main application to gears hauling non-selectively along the seabed.

For a meaningful estimate, surveys have to be conducted for several weeks every year, for which the cost could be high.

***

Chapter 8

Virtual Population Analysis

8.1

Maximum Sustainable Yield (MSY)[*]

The key objective of fisheries management is that removals should be as large as possible but sustainable in the long-term, that is, the fishery catch should be equal to the maximum sustainable yield, the MSY. In other words, the MSY for a given fish stock means the highest possible annual catch that can be sustained over time, by keeping the stock at the level producing maximum growth. The MSY refers to a hypothetical equilibrium state between the exploited population and the fishing activity.

Recruitment overfishing occurs when a stock level is low, the amount of fish that can be harvested is limited because of lack of adults, which in turn produces too few fish that are able to replenish those that have been removed. Growth overfishing is where the recruits to a fishery are caught before they reach the size that would produce the maximum yield per recruit. At the other extreme, when a stock is excessively large, the yield is restricted; the growth of the stock slows down due to competition for food, cannibalism or the limitations in the carrying capacity of the environment. In between these two states lies a stock size at which the sustainable catch is at the highest level – MSY.

When discussing MSY, it is essential to distinguish between three closely related concepts: MSY, BMSY and FMSY. The effective implementation of the overall MSY goal depends on a clear understanding of the relationships between these three concepts.

MSY is supported by a stable population size known as BMSY (= “biomass MSY”). Consequently, “reaching MSY” means rebuilding fish populations to the BMSY level, in order to be able to support the level of annual catches known as MSY. While MSY is the catch and BMSY is the population size (= biomass), the fishing mortality is the catch rate (resulting from dividing MSY by BMSY). When the fish population is at BMSY and the fishing fleet is catching an annual amount of fish equal to MSY, then the resulting fishing mortality is FMSY.

Up to a certain level, the yield increases with increase in stock biomass (Fig. 1) and fishing mortality (or fishing effort) (Fig. 2; see page 102).

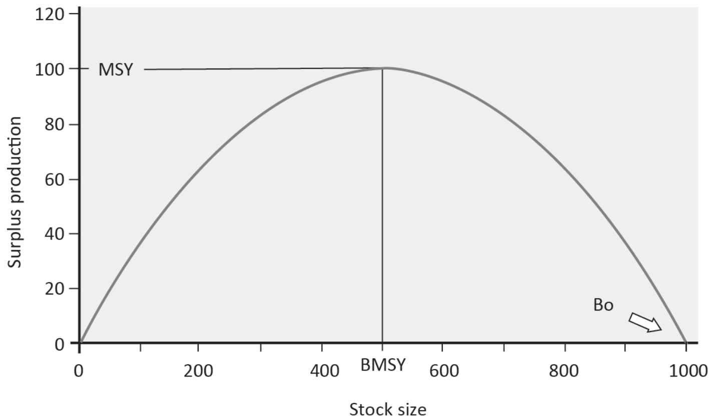


Fig. 1. A typical catch curve in relation to stock biomass; Bo = Biomass of virgin stock

---

[*] *E. Vivekanandan, National Consultant, World Bank/GEF Ocean Partnership Project, Bay of Bengal Programme Inter-governmental Organisation, 91, Saint Mary’s Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.*

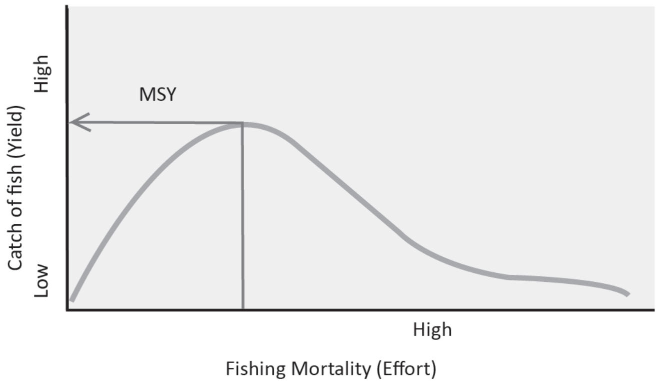


Fig. 2. A typical catch curve in relation to fishing mortality (see also Fig. 1 in Session 1.1)

Population growth

The key assumption behind the MSY is that populations of organisms grow and replace themselves – that is, they are renewable resources. Additionally, it is assumed that because the growth rates, survival rates, and reproductive rates increase when harvesting reduces population density, they produce a surplus of biomass that can be harvested.

Another assumption of renewable resource harvesting is that populations of organisms do not continue to grow indefinitely; they reach an equilibrium population size, which occurs when the number of individuals matches the resources available to the population. At this equilibrium population size, called the carrying capacity, the population remains at a stable size.

The logistic model is a function that is used to describe bounded population growth under the previous two assumptions. The logistic function is bounded at both extremes: when there are no individuals to reproduce, and when there is an equilibrium number of individuals (*i.e.*, at carrying capacity).

The equation describing logistic growth is:

$$N_t = \frac{K}{1 + \frac{K-N_0}{N_0} e^{-rt}}$$

Where,

Nt = Population size at time t

K = Carrying capacity of the population

N_0 = The population size at time zero

r = the intrinsic rate of population increase (the rate at which the population grows when it is very small)

From the logistic function, the population size at any point can be calculated if Nt, K and N_0 are known.

MSY estimation principle

The simplest way to model harvesting is to modify the logistic equation so that a certain number of individuals is continuously removed.

$$\frac{\mathrm{dN}}{\mathrm{dt}} = rN\left(1 - \frac{N}{K}\right) - H$$

Where H represents the number of individuals being removed from the population – that is, the harvesting rate. When H is constant, the population will be at equilibrium when the number of individuals being removed is equal to the population growth rate (Fig. 3). The equilibrium population size under a particular harvesting regime can be found when the population is not growing – that is, when dN/dt = 0. This occurs when the population growth rate is the same as the harvest rate (H):

$$rN\left(1 - \frac{N}{K}\right) = H$$

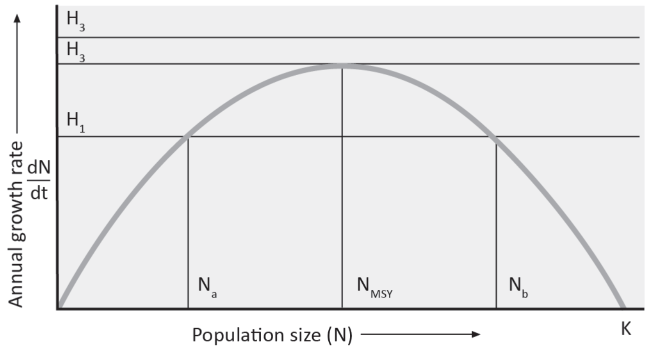


Fig. 3. Varying population growth rate with population density

Figure 3 shows how growth rate varies with population density. For low densities (far from carrying capacity), there is little addition (or "recruitment") to the population, simply because there are few organisms to give birth. At high densities, though there is intense competition for resources, and growth rate is again low because the death rate is high. In between these two extremes, the population growth rate rises to a maximum value (NMSY). This maximum point represents the maximum number of individuals that can be added to a population by natural processes. If more individuals than this are removed from the population, the population is at risk for decline to extinction.

Figure 3 also shows several possible values for the harvesting rate, H. At H1, there are two possible population equilibrium points: a low population size (Na) and a high one (Nb). At H2, a slightly higher harvest rate, however there is only one equilibrium point (at NMSY), which is the population size that produces the maximum growth rate. With logistic growth, this point, called the maximum sustainable yield, is where the population size is half the carrying capacity (or N = (K/2). The maximum sustainable yield is the largest yield that can be taken from a population at equilibrium. In Figure 3, if H is higher than H2, the harvesting would exceed the population's capacity to replace itself at any population size (H3 in Fig. 3).

Because harvesting rate is higher than the population growth rate at all values of N, this rate of harvesting is not sustainable.

Implications of MSY model

Starting to harvest a previously unharvested population will always lead to a decrease in the population size. That is, it is impossible for a harvested population to remain at its original carrying capacity. Instead, the population will either stabilize at a new lower equilibrium size or, if the harvesting rate is too high, decline to zero.

The reason why populations can be sustainably harvested is that they exhibit a density-dependent response. This means that at any population size below K, the population is producing a surplus yield that is available for harvesting without reducing population size. Density dependence is the regulator process that allows the population to return to equilibrium after a perturbation. The logistic equation assumes that density dependence takes the form of negative feedback.

If a constant number of individuals is harvested from a population at a level greater than the MSY, the population will decline to extinction. Harvesting below the MSY level leads to a stable equilibrium population if the starting population is above the unstable equilibrium population size.

Methods of estimating MSY

The two best known and commonly used methods for estimating the MSY are the surplus production models, namely the Schaefer model and the Fox model. These models need time series data on catch and fishing effort and are based on the assumption that the biomass of fish in the sea is proportional to the catch per unit effort. These two models are described in Section 10 of this manual.

In spite of the advances made in the field of fishery sciences and the establishment by most fishery research institutions of data bases for stock assessment purposes, there are still situations where no catch and effort time series exist. In such cases estimation of potential yield is usually based on rough estimates of standing stocks (*e.g.* from resource surveys) and the general knowledge of some biological characteristics of the species. Two equations have been proposed in the past to estimate the potential yield of unexploited fish stocks by Gulland in 1971 and exploited stocks by Cadima in 1977 for which only estimates of overall biomass and natural mortality are available.

Gulland proposed the following estimator of the MSY of a virgin stock when estimates of the natural mortality rate (M) and the biomass of the virgin stock (Bo) are available:

MSY = 0.5 MBo

This estimator has been extensively used with variable success. It is based on the observation that in the Schaefer production model, the biomass at MSY is equal to half the biomass in the virgin state (BMSY=0.5Bo). It is also based on the assumption that the fishing mortality (FMsy) at MSY is roughly equal to M. It follows that MSY =FMsyBMsy can be replaced by MSY =M0.5Bo.

Considering that the ratio between Bo and BMsY and between M and FMsy could be different for different species groups, Gulland proposed the following generalized form of the above equation:

MSY = xMBo, and suggested that x could be estimated from the Beverton and Holt yield tables.

It should be noted that Gulland's equations are not applicable when significant exploitation is already underway and the virgin biomass cannot be estimated.

Uses of MSY

MSY has been especially influential in the management of renewable biological resources such as commercially important fish. In fisheries terms, the MSY is the largest average catch that can be captured from a stock under existing environmental conditions. The MSY aims at a balance between too much and too little harvest to keep the population at some intermediate abundance with a maximum replacement rate.

Relating to MSY, the maximum economic yield (MEY) is the level of catch that provides the maximum net economic benefits or profits to society. The MEY is usually less than MSY.

Limitations of MSY approach

Although it is widely practiced, the MSY has come under heavy criticism by ecologists and others from both theoretical and practical reasons. The concept of maximum sustainable yield is not always easy to apply in practice. Estimation problems arise due to poor assumptions in some models and lack of reliability of the data. Biologists, for example, do not always have enough data to make a clear determination of the population's size and growth rate. Calculating the point at which a population begins to slow from competition is also very difficult. The concept of MSY also tends to treat all individuals in the population as identical, thereby ignoring all aspects of population structure such as size or age classes and their differential rates of growth, survival, and reproduction.

As a management goal, the static interpretation of MSY (*i.e.*, MSY as a fixed catch that can be taken year after year) is generally not appropriate because it ignores the fact that fish populations undergo natural fluctuations (*i.e.*, MSY treats the environment as unvarying) in abundance and will usually ultimately become severely depleted under a constant-catch strategy. Thus, most fisheries scientists now interpret MSY in a more dynamic sense as the maximum average yield obtained by applying a specific harvesting strategy to a fluctuating resource.

***

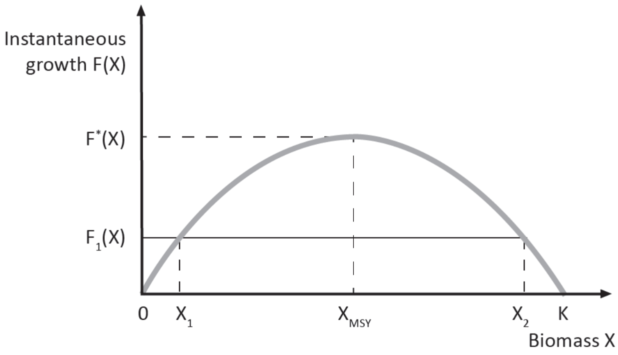
8.2

Maximum Economic Yield (MEY)*

Fishery resources are renewable natural resource but are not inexhaustible. They become extinct if the rate of harvest or exploitation is higher than the rate of regeneration or reproduction. Here the size of the stock (population) depends on the biological, economic and social considerations. Fisheries come under 'Common Property Resource (CPR)', due to which a comprehensive management measure becomes difficult for implementation. "In an open access regime like fishery, negative externalities are many, which implies that uncontrolled fishery will bound to end up in what is called tragedy of commons, (Grafton *et al.*, 2006)".

Sustainable fisheries yield

The sustainable yield in fishing, commonly referred to as "Maximum Sustainable Yield (MSY), is a biological phenomenon. MSY means that level of fish catch or yield that can be harvested from a given system is in perpetuity without affecting the stock of the system (or the sea). In other words, a catch level is said to be sustainable whenever it equals the growth rate of the population since it can be maintained for ever. As long as the population size remains constant, the growth rate will remain constant as well.



Economics of fisheries management

Economics play a vital role in fisheries management. In the earlier stages, fisheries management focused on controlling the effort to maintain fish stocks. The common assumption is that if the control measures are strictly implemented, further increase in effort is prevented and thus a sustainable harvest can be expected. But by 1970 it was found that such measures failed to control the fishing effort and capacity as the fishers substituted from regulated to unregulated inputs (Wilen, 1979) and further remedies suggested also failed to prevent the increase in fishing effort (Townsend, 1990).

"An economic perspective of fisheries management is that marine resources should not only be managed sustainably but also in a way that they contribute to and provide net benefits for the nation as a whole. Indeed the economists argue that sustainable and economically profitable fishery is complimentary. A level of harvest that maximizes the sustainable returns from fishing is often at a stock size that is greater than that which would maximize the overall yield from a fishery. Moreover, if there are other costs associated with fishing like habitat damage or environment loss, etc., the economic optimum level of harvest that accounts for these costs would be even less, and the desirable fish stock even larger. In other words,

---

** R. Narayanakumar, Principal Scientist & Head of Division, Socio-Economic Evaluation & Technology Transfer Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

a fishery that is economically optimum in the long run is also likely to be an ecologically sustainable fishery" (Grafton *et al.,* 2006).

Maximum Economic Yield (MEY) is realized at that level of effort in which the sustainable net return from the fishery is maximum. The difference between the total revenue (TR) and the total cost (TC) is maximum. This difference is also referred to as resource rent.

Total revenue (*TR*) = Price (*P*) × Catch (*H*)

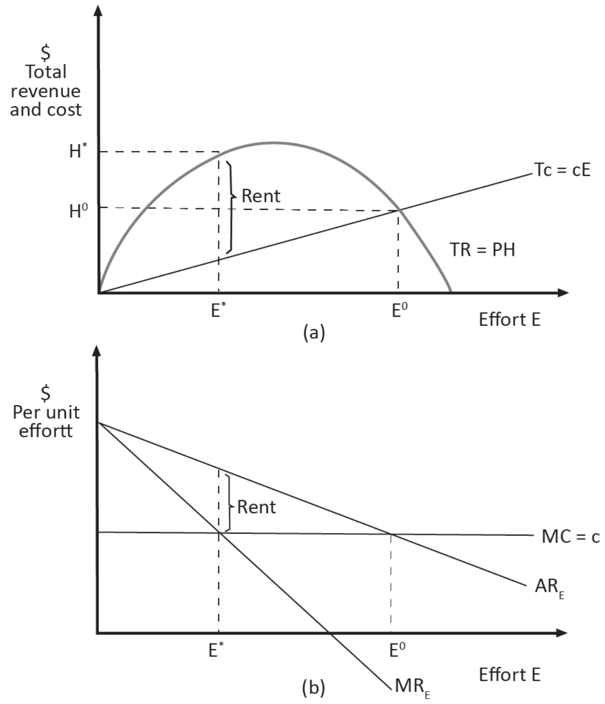
TC = Unit cost (*c*) × Effort

Rent = TR – TC

The resource rent is maximized at the point E*. Here:

MEY is left of *MSY*

- Optimal harvest (*H**) is less than the *MSY* harvest
- But rent is larger than at *MSY*



The marginal analysis can show that the MEY occurs at the point where MC =MR. It is observed that for marginal unit of effort, marginal rent is = 0 and average rent >1.

The point E* is that effort level at which the MEY occurs. At this point of effort only the difference between the total revenue from fishing and total cost of fishing is the maximum. This difference is also referred to as resource rent.

The goal of traditional fisheries management is to achieve *MSY.* However, the economists aim for MEY in contrast to MSY. AT MEY, compared to MSY, the fish catch is lower and fishing profit is higher. Correspondingly, the fishing effort is lower and the fish stock is higher. Thus MEY is where more fish is conserved and economic is the friend of conservation (Dixon, 2005; Grafton *et al.,* 2006).

MEY is affected by the changes in price of fish and the costs of fishing. When the price of fish increases, the total revenue curve shifts upward at all effort levels, leaving the intercepts unchanged and the point of MEY moves closer to MSY but never beyond MSY so long as the cost of fishing increases with effort. On the other hand, if the cost of fishing increases, the total cost curve moves upward to the left, thus the new point of MEY is to the left of the previous MEY. This will lower the optimal fishing effort (E*) because with a more costly harvest, it pays more to have larger stocks from which to catch. In total, a fall in fish price or an increase in cost of fishing will lead to lower harvest with less fishing effort and a larger stock size in order to maximize the economic profits (Grafton *et al.,* 2006).

Estimation of Maximum Economic Yield

The MEY and the effort to harvest this MEY (f_{mey}) can be estimated following Devaraj and Smita (1988) as mentioned below:

$p = a – by$…………………………………………………………….(1)

where,

p is the unit weight of fish

y is the annual yield

The average price per unit weight of fish (p) is generally a monotonically decreasing function of annual yield (y).

The profit is obtained as a difference between total revenue (TR) and total cost (TC), *i.e.*,

$P = TR – TC = (p-c)y$ ………………………..……………………(2)

where 'c' is the cost of harvesting one unit weight of fish. From this, a cost function will be fit from the data collected. Then the MEY and f_{mey} will be obtained using the following equations:

$MEY = (a – c) / 2b$ …………………………………………………..(3)

$f_{mey} = [a +/- (a^2 – 4\ b\ MEY)]^{1/2} / 2b$ ……………………………(4)

where, a = intercept; b, c =regression coefficients.

From $f_{mey,}$ the optimum fleet size is obtained by dividing 'b' by the average annual fishing days. Based on this the excess capacity and thus the capital investment (over and above the optimum fleet size) can be worked out.

Beyond Maximum Economic Yield

While estimating the MEY, three assumptions have been made: (i) zero discount rate; (ii) cost of fishing is a simple linear function of stock size; and (iii) fishing costs rise proportionately with effort.

The discount rate is that interest rate at which future catches are valued today. If the discount rate is very high, the MEY will correspond to a bionomic equilibrium (Clark, 1990), because it will be profitable to harvest the stock today itself if the loss of future net returns are very heavily discounted. The fisheries with commonly used discount rates, MEY will be conservationist and generate stock sizes that are larger than those associated with MSY. This implies that maximizing economic viability of fisheries is compatible with economic sustainability of the fisheries.

The cost of fishing is an important component that decides the MEY. Generally, the cost of fishing increases with a decrease in stock size at an increasing rate. This is the characteristics of the fishing practice. Under such a situation, it will be desirable to have a catch and effort level further to the left of the bionomic equilibrium.

In case of multi-species fishery, the estimation of MEY becomes complicated in many ways due to factors like biological interactions, apportioning of the cost of fishing, value of target versus by-catches and splitting the effort and related aspects. Though all these factors can be accommodated in the model, the computation of species-wise MEY becomes very difficult.

The concept of uncertainty in fisheries makes the process of estimating MEY more complicated. The sources of uncertainties include lack of complete biological data to calculate the stock-recruitment relationship, inability to accurately measure the actual catch and effort of fishers and the current size of the fish stock. Price of fish and the precise cost of fishing are also the other sources of uncertainties facing the fishery. Such aspects call for a different approach to manage fisheries.

Despite such assumptions, the MEY is a good target reference point for fisheries management. Because MEY ensures that the stock levels in many fisheries are larger than those associated with the traditional MSY target, it needs to be ensured that major inputs like fuel and labour are utilized efficiently so as to maximize the profit (Grafton *et al.,* 2006). If the resources are used beyond the MEY target, it will result in excess fishing capacity, lower returns and thus lower profits. Hence it pays rich dividend to follow the MEY as an important component for aiming at a sustainable fishery.

References

Clarke, C.W., 1990. Mathematical bioeconomics: the Optimal Management of Renewable Resources, *John Wiley & Sons, New York*.

Devaraj, M. and P. Smita. 1988. *Economic performance of mechanised trawlers in the State of Kerala, India. Fisheries Research*, 6 (3): 271-286.

Grafton, R.Q., J. Kirekly, T. Kompas and D. Squires, 2006. *Economics of Fisheries Management*. Ashgate Publishing Company, England, 160p.

John A. Dixon**,** 2005. Fisheries and Aquatic Resources. Caspian EVE 2005/UNDP and WBI.

Townsend, R.E., 1990. Entry restrictions in the Fishery: A Survey of the Evidence, *Land Economics,* 66:359-378.

Wilen, J., 1979. Fishermen behaviour and the design of efficient fisheries regulation programmes. *Journal of Fisheries Research Board of Canada*, 36: 855-858.

***

8.3

Cohort Analysis[*]

Cohort analysis is a general terminology for analytical techniques used in the estimation of mortality and population size using catch-at-age data. Different methods are available under the class of cohort analysis namely 'Virtual Population Analysis (VPA)' by Fry, Gulland and Cohort Analysis by Pope and also multispecies VPA. In these models, we trace the decline in abundance of cohorts as they age and pass through the fishery. The decline in cohort abundance is thus used to determine mortality rates. Originally, the method was developed to trace the population size over the life space of a cohort and later devised to estimate fishing mortality rates, current abundance and recruitment size. The cohort analysis concept was first developed by Derzhavin (1922) with his work on sturgeon. In 1949, Fry termed Derzavin's model as Virtual Population Model and applied it to trace historical population size. In 1965, Gulland attempted calculations for estimating historical population sizes starting with the oldest age group and worked backwards with a guess of fishing mortality rate F termed as terminal F.

The basic concept behind cohort models is that the size of a cohort when it first enters the fishery, known as recruitment size, can be approximately calculated using catches of that cohort across all the period it is in the fishery. The historical population size is known as virtual population. The term 'virtual' is used to distinguish something that is merely conceptual from something that has physical reality. It is otherwise referred to as cohort analysis because each cohort is analyzed and traced separately. These methods are now used to estimate cohort size, historic patterns of fishery mortality, age structure and recruitment size using data on numbers of fish harvested. The methods were initially developed for age-structured data and later extended for length-based data. The basic theory behind age-structured analysis is that the abundance of fish of age (t+1), which are the survivors from the year before, equals the abundance of fish of age t reduced by fishing and natural mortalities.

Gulland's VPA model

It is based on two basic equations. The first one is the exponential decay model given by:

$$N_{t+1} = N_t \, e^{-(F_t + M_t)} \qquad (1)$$

and the second equation is Baranov's catch equation given by

$$C_t = \frac{F_t}{N_t + M_t} \, N_t \left(1 - e^{-(F_t + M_t)}\right) \qquad (2)$$

Here N_t refers to the number of fish of age *t* in the year class, F_t and M_t are instantaneous rates of fishing and natural mortalities for fish of age *t* in the year class (constant within a year) and C_t is the catch in number of fish of age *t* in the year class.

The assumptions are (i) fishing takes place continuously through out the year and

(ii) fishing and natural mortality rates are constant within a year.

---

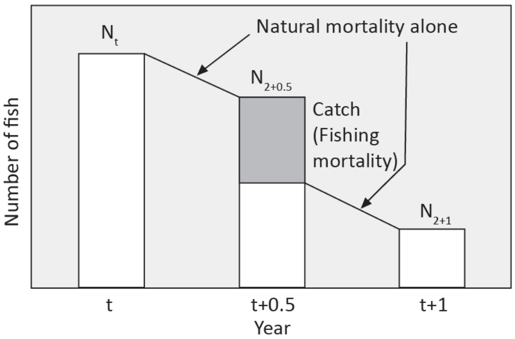
[*] *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

By dividing equation (1) with equation (2) we get the following expression:

$$\frac{N_{t+1}}{C_t} = \frac{Z_t\, e^{-Z_t}}{F_t\left(1-e^{-Z_t}\right)} \qquad (3)$$

Here $Z_t = F_t + M_t$ is the instantaneous rate of total mortality. Using the above three equations, starting from the terminal class, if N_{t+1}, C_t and M_t are known, we can solve for F_t from (3) and then compute N_t using (1). This procedure can be repeated to obtain F_{t-1} from (3) and then compute N_{t-1} using (1) knowing the values of N_t, C_{t-1} and M_{t-1}. Thus we can calculate these values recursively backward for all the age classes. To solve for F_t from (3), which is a non-linear equation in $F_{t,}$ iterative procedures have to be used and the calculations are not straight forward.

Pope's Cohort Analysis



Pope (1972) simplified Gulland's VPA model to avoid the iterative calculation procedure. In Pope's method, an year is broken into two parts with catch assumed to have occurred instantly in the middle of the year (say t+0.5). The natural mortality operates throughout the year so that by t+0.5 half of it (M/2) operated on the population size and the remaining half (M/2) is operated between t+0.5 and t+1. Accordingly, if N_t, $N_{t+0.5}$ and N_{t+1} are the population sizes at year t, t+0.5 and t+1 respectively then we have the following relation based on the exponential decay model (1).

$$N_{t+0.5} = N_t\, e^{-\left(\frac{M_t}{2}\right)} - C_t$$

$$N_{t+1} = N_{t+0.5}\, e^{-\left(\frac{M_t}{2}\right)}$$
$$= \left(N_t\, e^{-\left(\frac{M_t}{2}\right)} - C_t\right) e^{-\left(\frac{M_t}{2}\right)} \qquad (4)$$

The above equation can be rearranged to obtain an expression for backward calculation as:

$$N_t = \left(N_{t+1}\, e^{\frac{M_t}{2}} + C_t\right) e^{\frac{M_t}{2}} \qquad (5)$$

Thus starting from the terminal year class, if N_{t+1}, C_t and M_t are known using (5) population size, N_t can be calculated backward. Once N_t is available, corresponding fishing mortality rate can be calculated by rearranging (1) as:

$$F_t = \text{In}\left(\frac{N_t}{N_{t+1}}\right) - M_t$$

By rearranging (2) we can get the expression for the population size as:

$$N_t = \frac{C_t}{\left(\frac{F_t}{Z_t}\right)\left(1 - e^{-Z_t}\right)} \qquad (6)$$

If for the terminal class, the terminal $\left(\frac{F_t}{Z_t}\right)$ is known, say R (terminal F/Z), then the population size for the terminal class can be estimated as:

$$N_t = \frac{C_t}{R\left(1 - e^{-\left(\frac{M}{1-R}\right)}\right)}$$

Length-based VPA

In length-based virtual population analysis, in the derivation of Pope's model, instead of age classes we have length classes and hence we have to consider the mid-point of length classes. In place of t, t+0.5, t+1 we have N_t, $L_{t+0.5}$, L_{t+1} and when converted into age using inverse von Bertalanffy growth equation, if $dt = t_2 - t_1$ is the age difference corresponding to lengths L_t and L_{t+1} we will have the following equation corresponding to (5):

$$N_t = \left(N_{t+1}\, e^{\frac{M_t dt}{2}} + C_t\right) e^{\frac{M_t dt}{2}}$$

The approach for the length-based cohort analysis is basically the same as for the length-converted catch curve (Section 5). The arithmatic involved is the back calculation from an assumed value of F for the largest length group, because the F value for young fish takes a non-linear form in catches, population numbers, etc. It is assumed that the number of fish in all length classes caught in one year reflects that of a single cohort during its entire life span. This approach can be applied very widely in many commercial fisheries. This approach is valuable due to the existence of (i) large number of length groups in the fishery; (ii) long series of length composition data; and (iii) a complex and variable fishery in which F is likely to vary with age (length) and year.

The basic procedure is simple, but laborious if repeated for several cohorts and also for several assumptions about the value of F for the largest length group and about M. It is well suited to computer handling. Though this procedure allows estimates of F for each length group separately, the accuracy of the final estimates for a fishery can be increased by considering that different cohorts are likely to experience similar changes in F from year to year (due to changes in the number of vessels), and for different length groups (due to selectivity of the gears). On the other hand, if there are changes only in the number of vessels, the selectivity may not change. If there are changes in the type of gear used, or in fishing grounds, resulting in differences in the sizes of fish caught between the types of gear or fishing grounds, then the selectivity may not be constant from year to year.

The VPA or cohort analysis is used to determine (i) the number of fish that must have been present in the sea, based on the numbers in the catch; and (ii) the fishing effort that must have been spent on each length group to catch the numbers.

Computer programmes on VPA

ELEFAN III

ELEFAN III exclusively deals with the VPA. It requires in addition to length frequency data, the monthly catch data. It is necessary that the entire catch from the whole stock must be used, not the catch of a single gear type. This is because, the catch-at-length data representing only part of the fishery is generally not proportional to the total catch.

A routine is available in ELEFAN III, which allows users to store monthly catches and coefficients of length-weight relationship. The routine is in the form of data entry, data editing and viewing and printing.

ELEFAN III incorporates three types of VPA, *i.e.*, VPA I, VPA II and VPA III. VPA I estimates the standing stock (in numbers) and fishing mortalities by time interval (month, year, etc) for any given cohort. The following input is necessary to run VPA I: (i) a file identifier independent of the files stored in disk; (ii) the number of periods for which catches are available; (iii) M and Ft (terminal fishing mortality) estimates (on an annual basis); and (iv) catches by ages starting with the youngest fish. After providing the inputs, the results are displayed graphically. The user has the option to repeat the analysis by changing the values of M and /or Ft.

VPA II is used to estimate mean standing stock for a stable age distribution, as can be simulated by combining data for several years. It makes use of either catch-at-length data or length frequency data saved in ELEFAN 0. In either type of data, the programme requires the mean annual catch data (in tonnes) represented by the samples and constants for the length-weight relationship. For length frequency data, in addition to the input requirements, monthly catch (in tonnes) should also be entered to allow conversion of the data from length frequency to catch-at-length type of data. The growth parameters (L_∞ and K), M and Ft should be entered. After the requirements are satisfied, the computer estimates the steady-state biomass for each length class. The results in the biomass and other outputs such as the estimated Fi (fishing mortality in the i^{th} period), catches and population are displayed graphically on the screen.

VPA III provides estimates of monthly length-wise standing stock and fishing mortality by segregating cohorts through catch-at-length data by means of a set of growth parameters. This approach assumes that little exchange occurs between the monthly cohort, which is true especially for short-lived groups, such as anchovies and penaeid prawns. VPA III incorporates the features of both VPA I and VPA II. A set of monthly length frequency and catch data, L_{∞}, K, M and Ft is the required input. Using the parameters given, the programme will initialize the array that stores the results for computed monthly summaries. Once the above routine has been completed, the user will be given the opportunity to reanalyse separate cohorts, one after another. The results generated in each cohort analysis will update the array that stores the monthly summaries. The results are displayed graphically.

The programme LCOHOR in the LFSA package can execute Jones' length-based cohort analysis. FiSAT analyses the VPA in three forms, *viz.*, age structured VPA, length-structured VPA and length/age VPA in the routine ASSESS. In addition to the ELEFAN, LFSA and FiSAT packages, a package of microcomputer programmes, ANACO (Analysis of Cohorts) performs the VPA calculations. The ANACO also offers a number of additional options such as sensitivity analysis.

***

Chapter 9

Prediction Models

9.1

Beverton and Holt’s Yield Per Recruit Model*

Introduction

Beverton and Holt’s yield-per-recruit model is one of the commonly applied methods for providing management advice for fish populations. The yield-per-recruit model is widely used mainly due to its relative simplicity and the fact that the model inputs like natural mortality, age/length at capture and weight-at-age are readily available for most species, or reasonable guesses can be made by inference with similar species.

The basic yield-per-recruit model divides the population into several growth groups to represent individual variability in growth. All animals in a growth group grow according to the same growth curve but the growth curve differs among growth groups. This approach to modelling individual variability in growth was chosen because it is simpler than attempting to develop a size-structured or an age- and size-structured population dynamics model. The extended yield-per-recruit model considers the dynamics of the population by age within each of ‘n’ growth groups.

Assumptions

The yield per recruit model developed by Beverton and Holt in 1957 is a “steady state model”. A model that describes the state of the stock and yield when the fishing pattern has been the same over a long period so that all recruited fish alive are exposed to fishing is termed as a steady state model. The Beverton and Holt’s model makes the following assumptions:

- Recruitment is constant.
- All fish of a cohort are born on the same day.
- Recruitment and selection are knife-edged.
- The fishing and natural mortalities are constant through out the phase after recruitment.
- There is a complete mixing within the stock.
- Growth in weight is isometric. That means b=3 in $W_t=aL_t^b$

Yield-per-recruit model

This model describes the effects of alteration of two parameters that can be controlled by man- the F, and the age or length at first capture. As mentioned earlier in this book, overfishing may occur as growth overfishing or recruitment overfishing. Growth overfishing occurs when young fish are caught before they grow to a reasonable size. Thus, the problem is to estimate the most suitable length (or age) at first capture and suggest the mesh size, which would allow the young fish to escape. The right mesh size optimizes the yield that could be obtained from a given number of recruits. Therefore, calculation of yield from a given recruitment, known as yield-per-recruit, is a basic element in the assessment of fish stocks.

** K.G. Mini, Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India & E. Vivekanandan, National Consultant, World Bank/GEF Ocean Partnership Project, Bay of Bengal Programme Inter-governmental Organisation, 91, Saint Mary’s Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.*

The assumed life history of a cohort in the Beverton and Holt model, as shown in Figure 1 is as follows:

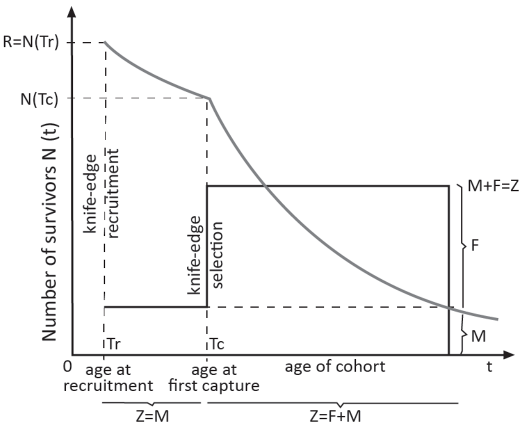


Fig. 1. The life history of a cohort as assumed in the Beverton and Holt model

1) At age T_r all fish belonging to a given cohort recruit to the fishing grounds at the same time: "knife-edge recruitment".

2) From age T_r to age T_c the cohort is not exposed to any fishing mortality. In that period they suffer only from natural mortality, M, this is assumed to remain constant throughout the life span of the cohort.

3) At age T_c, the "age-at-first-capture", the cohort is assumed to be suddenly exposed to full fishing mortality, F, which is assumed to remain constant for the rest of the cohort's life. The catch from the cohort is therefore assumed to be zero until the cohort has attained the age T_c .

The number of survivors at age T_r is the recruitment to the fishery

$$R = N\,T_r \qquad ————(1)$$

The number of survivors at age T_c, the age at first capture is

$$N(Tc) = R\,e^{-M(T_c-T_r)} \qquad ————(2)$$

The number of survivors of the recruited cohort at age t is

$$N_t = N(Tc)e^{(-(M+F)\,(t-T_c))}$$

$$N_t = Re^{(-(M)(T_c-T_r)-(M+F)\,(t-T_c))} \qquad ————(3)$$

The fraction of the recruits surviving up to age t is given by

$$N_t/R = e^{(-(M)(T_c-T_r)-(M+F)(t-T_c))} \qquad ————(4)$$

This means that Eq. (4) gives the number of fish at time t per recruit. This also means the fraction of each fish that is recruited to the fishery.

The numbers caught between a very small interval $(t,t+\Delta t)$ is given by

$$C\,(t,t+\Delta t) = \Delta t\,F\,N\,(t) \qquad ———— (5)$$

The above equation gives the number of fish caught from a cohort, in the time period from 't' to $t + \Delta t$ when Δt is small. To obtain the corresponding yield in weight, this number should be multiplied by the individual weight of a fish. If Δ is small, then the body weight of a fish will remain approximately constant during the time period from 't' to, $t + \Delta t$ and the yield becomes

$$Y\,(t,t+\Delta t) = \Delta t\,F\,N\,(t)\,w(t) \qquad ———— (6)$$

where w_t is the body weight of a t years old fish.

The yield per recruit for the time period from 't' to $t + \Delta t$ is divided by the number of recruits, R

$$\frac{Y(t, t+\Delta t)}{R} = \Delta t\, F \frac{N(t)}{R} w(t) \qquad ———— (7)$$

where N_t/R is defined by Eq. (4).

The above equation is the Beverton and Holt model for a short time period. To obtain the total yield per recruit for the entire life span of the cohort, Y/R, all the small contributions defined by Eq.(7) have to be added up. The expression for the total yield per recruit is:

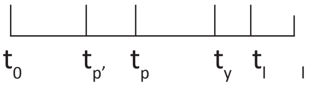
$$\frac{Y}{R} = F\, e^{(-M(T_c - T_r))} W_\infty \left[\frac{1}{Z} - \frac{3S}{Z+K} + \frac{3S^2}{Z+2K} - \frac{S^3}{Z+3K} \right] \qquad ———— (8)$$

where:

$S = e^{-K(T_c - t_o)}$
K = von Bertalanffy growth parameter
t_o = von Bertalanffy growth parameter
T_c = age at first capture
T_r = age at recruitment
W_∞ = asymptotic body weight
F = fishing mortality
M = natural mortality
Z = F + M, total mortality

The model allows us to calculate Y/R with varying inputs of the different parameters, such as F and t_c and then assess which effect the various input values have on the yield-per-recruit of the species under investigation. It is important to note that t_c and F are the two parameters over which the fishery managers have control. Fishing mortality F is proportional to effort and t_c is a function of gear selectivity, which in turn is related to mesh size. Hence Y/R can be considered as a function of F and t_c, and often Y/R values are calculated for varying inputs of F and plotted for finding optimum value of F. This curve is known as yield per recruit curve and it often has a maximum that corresponds to the Maximum Sustainable Yield (MSY). This maximum changes as the value of 't_c' used is changed. By varying *F* and 't_c' simultaneously we can obtain a combination of F and 't_c' which gives the highest value for MSY.

Before proceeding with calculations, the size (or age) of fish and its relation to capture should be properly understood. Considering the age of fish from 't_0' (age of fish when the length is 0) to 't_l' (age at $L_{¥}$), the fishery may intercept the fish at different stages of its life in the following way:



where $t_{p'}$ = age at which encountered by the gear; t_p and t_p = age of exploitation; t_y = optimum age of exploitation; t_l = maximum age represented in the fishery. Hence, from t_p to t_l = exploitable life span; from t_0 to l = life span; t_0 = t_p = unexploited phase (only natural mortality occurs); and t_p to t_l = exploited phase (both natural and fishing mortalities occur). Usually t_p = t_p.

The theory behind the computations used in estimating yield per recruit and the optimal size at first capture was first developed in 1956 by Beverton and Holt. The calculation of yield per recruit commences by considering the fate of number (R) of recruits entering the fishery at some age (t_p); in a short interval of time (t) the numbers and weight of fish caught can be obtained for summing or integrating, over the period that the group of fish is exposed to the fishery. In other words, the number of fish (or weight of fish) caught from the age of recruitment ‘t_p’, up to some limiting age, ‘t_l’.

For performing the calculations, the following two steps are required: (i) estimations on (a) von Bertalanffy growth parameters, (b) fishing mortality coefficient, which may vary with age, (c) natural mortality coefficient, supposedly constant, and (d) average weight of an individual fish of age ‘t’; and (ii) calculation of yield per recruit by making use of the parameters above.

The yield-per-recruit analysis is not without its weaknesses. Prime amongst these are the assumptions that population is in steady state and that recruitment is independent of the size of the spawning stock. The last of these assumptions is only needed if the results of the analysis are used to make inferences about likely yields (rather than levels of yield-per recruit). Recent analysis of stock and recruitment data sets for many species reveal that the assumption that recruitment is independent of spawning stock size is probably invalid for many, if not most stocks. Other weakness of the yield-per-recruit approach, as it is usually applied is, that it completely ignores the dynamics of movement and migration and assumes that length is related deterministically to age.

Computer programmes on prediction models

ELEFAN II, which is a collection of auxillary routines, estimates the relative yield and biomass-per-recruit. The routine estimates the optimal level of exploitation rate (E) using the yield-per-recruit model of Beverton and Holt, modified by Pauly and Soriano (1986). The Y/R can be computed in the following two ways: (i) assuming knife-edge selection, wherein Y/R is a function of M/K, E and C (mean length at first capture/L_∞); and (ii) using probabilities of capture that are gradually changing. The inputs required for (ii) are the smallest mid-length, largest mid-length, class interval, L_∞, M/K and the probability of capture for each mid-length. The routine also includes a plot of relative biomass/recruit against exploitation rate (E).

The LFSA package has two programmes, the TBYR and MIXFISH. The TBYR uses a special version of the Thompson and Bell yield and stock prediction model for the single stock, single fishery situation. The TBYR takes its starting point in the stock numbers by length group calculated by LCOHOR and converts them into age groups. Since conversion from length group to age group is problematic for short-lived species, this programme should be used only for long-lived species (5 years or more).

The other programme in the LFSA is MIXFISH, which is a length-based Thompson and Bell model with option for analysis of a mixed fishery. It is similar to the TBYR, but without the conversion of the length into age groups. It can be used for long-lived as well as short-lived species. Although designed for analysis of a mixed fishery, the MIXFISH contains the single species case as an option. It contains an option for mesh assessment and produces an output showing the total yield for various combinations of effort and L_{50}%.

The FiSAT package contains the Y/R and B/R analyses and knife-edge selection and selection curves. It also contains the Thompson and Bell yield and stock prediction for single/ multispecies fisheries.

Packages such as ANALEN and BEAM also can be used for analysing the prediction models. In the ANALEN, the programme ANAJON calculates the yield-per-recruit and the mature biomass-per-recruit for different levels of exploitation. It also provides short-term and long-term forecasts of catches. The BEAM package has four programmes, BEAM 1, BEAM 2, BEAM 3 and BEAM 4. The first two are simple bioeconomic modeling of artisanal and industrial sequential shrimp fisheries based on Thompson and Bell model and a simple input-output microeconomic model. BEAM 1 simulates results by age groups and BEAM 2 uses standard commercial categories. BEAM 3 is a stochastic model that can handle up to four species (or both sexes of two species) and many fleets operating sequentially or simultaneously. BEAM 4 allows for more realistic simulations and has high data requirements.

Reference

Sparre, P., and S. Venema. 1992. Introduction to tropical fish stock assessment. Part 1. Manual. FAO Fish. Tech. Paper 306/1 rev 1. FAO, Rome.

***

9.2

Thompson and Bell Yield and Biomass Prediction Model*

In 1934, Thompson and Bell developed a model to predict catch and stock size for a given fishing pattern. The model developed by them is an age structured model. This model is used to predict the effects of changes in fishing effort on future yields. The length-based version of this model usually takes necessary inputs, the number of recruits and fishing mortality rates for each length classes, from a length-based cohort analysis. When recruitment details are not available, the model can still be used to get relative figures (in the form of per 1000 recruits). An important aspect of this model is that it allows bio-economic analysis, if value of the catch is provided as the input. The main four input parameters required for this model are: (i) The main input is the range of F values for each length group. The range of F values should be related to the real situation in the fishery, which may be obtained from the VPA analysis or from any other method such as the length-based catch curve method; (ii) Another important input parameter is the number of recruits, which also may be obtained from the VPA (or cohort analysis); (iii) The model also requires weight of individual fish in each length group; and (iv) For economic analysis the model requires the price per kg of each length group. The output parameters are the prediction of catch in numbers, total deaths in numbers, the mean biomass and yield for a combination of different F and M values. The effects of changes in F on the yield, average biomass and value of the catch can be calculated.

The model makes predictions of catch in numbers, total number of deaths, yield, mean biomass and value for each age group. Changes in effort can be introduced by multiplying the F-array by a suitable factor and corresponding predictions can be obtained. Thompson and Bell model can be used to study the effect of certain management measures (such as closed seasons, increase in fishing effort, etc.) on the yield, biomass and value. The outputs from Thompson and Bell analysis are for each length class the stock size, the catch in numbers, the yield in weight, the biomass and the time required to grow from the lower limit to the upper limit of the length group. Finally, the total of the catch, yield and mean biomass are also obtained.

The length converted Thompson and Bell analysis use the F - array (fishing mortality rates for each length class) estimated through cohort analysis as the reference F - array and assesses the effect of raising or reducing the F - array by a certain factor. The prediction made by length converted Thompson and Bell analysis is a prediction of the average long-term catches assuming recruitment to remain constant.

Calculations involved in length-based Thompson and Bell predictions

Inputs for the analysis consists of:

- Fishing mortalities for different length groups
- Number of fish in the smallest length group (recruitment size)
- The growth parameters
- Natural mortality rate(s)

---

* *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

- Length-weight relationship parameters
- Average price per kg for different length groups
- The i^{th} length class is $(L_i - L_{i+1})$
- Calculate the total mortality sequence as (*x* is the multiplier used to raise or reduce the fishing mortality rates sequence, *x = 1* for the current level of exploitation)

$$Z_i = M + x\,F_i$$

- Calculate population size for successive classes

$$H_i = \left[\frac{L_\infty - L_i}{L_\infty - L_{i+1}}\right]^{\frac{M}{2K}}$$

$$N_{i+1} = N_i \left(\frac{\frac{1}{H_i} - x\,\frac{F_i}{Z_i}}{H_i - x\,\frac{F_i}{Z_i}}\right)$$

- Calculate Catch for each class

$$\mathrm{C}_i = [N_i - N_{i+1}]x \left(\frac{F_i}{Z_i}\right)$$

$$\overline{w}_i = a \left[\frac{L_i + L_{i+1}}{2}\right]^b$$

- Calculate average weight for each length class

$$Y_i = C_i\,\overline{w}_i$$

- Calculate yield for different classes

$$V_i = Y_i\,\bar{v}_i$$

- Calculate value for each class

$$\overline{\mathrm{N}}_i = \frac{N_i - N_{i+1}}{Z_i \Delta \mathrm{t}_i}$$

- Calculate average population size for each class ($\Delta t_i = t_2 - t_1$ where t_1 and t_2 are the age corresponding to lower and upper limits of the i_{th} class obtained using the inverse VBGF)

$$\overline{B}_i = \overline{N}_i\,\overline{w}_i\,\Delta \mathrm{t}_i$$

- Calculate average biomass for each classes

$$\mathrm{Y} = \sum_i Y_i$$

- Calculate total yield and average biomass combing all the cohorts as:

$$\overline{\mathrm{B}} = \frac{\sum_i B_i \Delta t_i}{\sum_i \Delta t_i}$$

- To increase or decrease effort, the factor *x* is suitably changed.
- Calculate the above quantities for different values of x = 0.2, 0.4, 0.6 etc.
- The total yield, average biomass and total value are then plotted against different values of *x* to arrive at MSY and MEY levels.

The assumption in this method is that the stock remains in a steady state and all parameters, including recruitment remain constant.

***

Chapter 10

Macro-analytical Models

10.0

Macro Analytical Models*

Production models are classified into two major groups, namely Macro/Global/Synthetic models and Micro Analytical Models. Macro models are based on quite simple equations where both the state of the population and fishing activity are each described by a single variable. These models take into account only the interrelationship between observable inputs such as fishing effort and observable output, which is the yield obtained from the fishery. Surplus production models are macro analytical models.

Surplus production models are an important approach to the study of harvested population dynamics. In surplus production models, the stock is considered as a single unit of biomass and modeling is not based on any age structure, length structure or dynamics of the population in terms of growth and mortalities. Instead, in these models the entire stock, the fishing effort and the total yield obtained from the stock are studied and a relationship between these are established without considering any micro-level details such as growth, mortality, age at first capture, mesh size effect, etc. The objective here is to obtain optimum levels of effort, which gives the maximum yield that can be sustained over a long period. These models do not demand much data for the analysis and for this reason these models are more popular. When reasonable estimates are available for the yield and corresponding fishing efforts over a period of time, these models can be used for obtaining optimum levels of effort and corresponding yield estimates.

Change in biomass depends on recruitment, growth and mortality. This can be represented by the following equation:

$$B_{t+1} = B_t + R_t + G_t - Z_t$$

where B_t is the biomass at time t, R_t is the weight of the new recruits into the fishery, G_t is the total increase in the weight of the animals due to growth and Z_t is the weight of the animals died during the period. Then production is given by:

$$P_t = B_{t+1} - B_t = R_t + G_t - Z_t$$

The population is in equilibrium when production is zero.

When

$P_t = 0$, population is in equilibrium

$P_t > 0$, population is in surplus

$P_t < 0$, population is in depletion

The population may collapse when P_t goes beyond some values. Here biomass is a point time concept and yield or production is a period concept.

---

* *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

At given time *t*, under fishing activity f_t and population state B_t, the change in B_t is assumed to depend on population state and fishing activity. Hence the equation used commonly to define surplus production models is:

$$\frac{dB_t}{dt} g\,(f_{t,}\ B_t)$$

Different versions of this model are given by different workers, such as:

1. Pella and Tomlinson

$$\frac{dB_t}{dt} = r\,B_t\left[1-\left(\frac{B_t}{B_o}\right)^{m-1}\right] - \mathrm{q} f_t\,B_t$$

2. Graham Schaefer's model

$$\frac{dB_t}{dt} = r\,B_t\left[1-\left(\frac{B_t}{B_o}\right)^{2}\right] - \mathrm{q} f_t\,B_t$$

3. Exponential model

$$\frac{dB_t}{dt} = r\,B_t\left[1-1\mathrm{n}\left(\frac{B_t}{B_o}\right)\right] - \mathrm{q} f_t\,B_t$$

Here B_0, *'r'*, *'m'* and *'q'* are parameters of the model which have to be estimated using data on yield and fishing effort.

In surplus production model, the rate of increase in biomass is taken as a function of biomass itself so that the relative change is given by the equation:

$$\frac{1}{B_t}\,\frac{dB_t}{dt} = f_t\,(B_t) - F_t \quad \text{where } F_t = q\,f_t$$

and F_t is the reduction in biomass due to fishing. When the production is surplus, the relative change in biomass will be positive and it will be zero when the population is in the state of equilibrium and hence $f\,(B_t) = F_t$ at equilibrium.

Graham-Schaefer Model: In this model, the first order differential equation is used to describe the rate of change of stock biomass B_t due to production. In the absence of fishing, the rate of change in the biomass is assumed to be a function of current population size only.

That is,

$$\frac{dB_t}{dt} = rB_t - \frac{r}{K}\,B_t^2$$

where B_t is the biomass at time *t*, *K* is the carrying capacity beyond which the population cannot grow and *'r'* is the intrinsic rate of increase in stock per unit time. When fishing mortality is added to this model it becomes,

$$\frac{dB_t}{dt} = (r - F_t)\,\frac{r}{K}\,B_t^2$$

$$= \alpha_t\,B_t - \beta\,B_t^2$$

where $\alpha_t = (r - F_t)$, $\beta = \frac{r}{K}$ and F_t is the instantaneous rate of fishing mortality.

For a short period $(r = h, t = h + \delta)$ during which the instantaneous rate of fishing mortality F_t is constant, the solution of the differential equation gives:

$$B_{h+\delta} = \begin{cases} \dfrac{\alpha_h B_h e^{\alpha_h \delta}}{\alpha_h + \beta B_h e^{\alpha_h \delta - 1}} & when\ \alpha_h \neq 0 \\ \dfrac{B_h}{1 + \beta\ \delta\ B_h} & when\ \alpha_h = 0 \end{cases}$$

and yield during the same period denoted by Y_h is:

$$Y_h = \int_{t=h}^{t=h+\delta} F_t\ B_t\ dt$$

and solution of this integral yields:

$$Y_h = \begin{cases} \dfrac{F_h}{\beta} \ln[1 - \dfrac{\beta\ B_h (1 - e^{\alpha_h \delta})}{\alpha_h}] & when \alpha_h \neq 0 \\ \dfrac{F_h}{\beta} \ln[1 + \delta\ \beta\ B_h] & when \alpha_h = 0 \end{cases}$$

The estimated average biomass during this short period $(r = h, t = h + \delta)$ is given by:

$$\bar{B}_h = \frac{Y_h}{F_h}$$

The surplus production during this period $(r = h, t = h + \delta)$ is:

$$P_h = B_{h+\delta} - B_h + Y_h$$

When yield is equal to surplus production, the population is in equilibrium.

Parameter estimation

It is assumed that the yield Y_t at equal time periods *t=1,…,T* are available. The following notations and assumptions are made for estimation purpose.

Parameters to be estimated are *'r'*, *'K'*, *'q'* and the initial biomass B_1.

B_t :	Population biomass at start of time *t*
Y_t :	Yield in biomass during time *t*
P_t :	Surplus production during time *t*
F_t :	Fishing mortality rate during time *t*,assumed to be proportional to fishing effort rate.
f_t :	Fishing effort rate during time *t*

$$F_t = q f_t$$

$$\alpha_t = r - F_t$$

Algorithm for estimation

The estimation procedure is by minimizing an objective function. With some starting guess estimates of the parameters compute the initial biomass and project through time estimating the yield for each time point *t=1,…,T*. The procedure is then iterative leading to the general function minimization procedure with the function to be minimized is:

$$f(r,K,q,B_1) = \sum_{t=1}^{T} [\log(Yt) - \log(\hat{Y}t)]^2$$

where Y_t is the actual yield and $\hat{Y}_t$ is the corresponding yield estimated according to the model. Fishing mortality can also be estimated from recorded yield using the equation:

$$F_t = \begin{cases} \dfrac{\beta\, Y_t}{\ln[\dfrac{\beta\, B_t e^{\alpha_t} - 1}{\alpha_t} + 1]} & when\ \alpha_t \neq 0 \\ \dfrac{\beta Y_t}{\ln(1+\beta B_t)} & when\ \alpha_t = 0 \end{cases}$$

Pella and Tomlinson's Model: One problem with the Graham-Schaefer model is that the maximum sustainable yield *MSY* always occurs when the biomass is half the carrying capacity *K*. This is a direct consequence of the parabolic relationship between $\frac{dB_t}{dt}$ and B_t, which in turn follows from the linear relationship between per capita productivity and population size. Pella and Tomlinson (1969) proposed an alteration to the model for which uncouples B_{MSY} from *K*.

One form of this model is given by $\dfrac{dB_t}{dt} = \begin{cases} a\, B_t^n - b\, B_t & for\, 0 < n \le 1 \\ b\, B_t - a\, B_t^n & for\, n > 1 \end{cases}$

Simple forms

1. The simple representation of Schaefer model is:

$$(Y_t / f_t) = a + b f_t$$

For this model, the catch per unit effort is considered as a linear function of effort and the linear relationship has negative slope and positive intercept. Under this model the catch per unit effort will be maximum when:

$$f_t = \frac{-a}{b}$$

The maximum sustainable yield (*MSY*) for the model is:

$$MSY = \frac{-a^2}{4b}$$

and the corresponding effort is:

$$f\,MSY = \frac{-a^2}{2b}$$

When we have time-series data on catch and effort by a linear regression of catch per unit effort (Y_t / f_t) (CPUE) on effort f_t , we can estimate the coefficients *'a'* and *'b'* and calculate MSY using this estimates.

2. In the model suggested by Fox, exponential relationship between *CPUE* and effort is assumed. The model is given by:

$Y_t / f_t = \mathrm{e}^{c+d} f_t$ or equivalently $\text{In}\,(Y_t / f_t) = c + d f_t$

This function will have maximum value for the yield when:

$$f_t = \frac{-1}{d}$$

and the maximum value of yield (*MSY*) is given by:

$$MSY = \frac{-1}{d}\, e^{\,c-1}$$

Using time-series data on catch and effort through a linear regression of logarithm of catch per unit effort $\text{In}\,(Y_t / f_t)$ on effort f_t, we can estimate the coefficients *'c'* and *'d'* and calculate MSY using this estimate.

***

Chapter 11

Productivity-based Biomass Estimates

11.0

Productivity-based Biomass Estimates*

Global biogeochemical cycle

Fixation of inorganic carbon to organic carbon in the ocean is driven purely by phytoplankton. Phytoplankton carbon fixation plays an important role in maintaining the quasi-steady state level of atmospheric CO_2. Relative contribution of marine primary productivity to global photosynthetic production is between 10 and 50 percent. The magnitude ranges from 20 to 55 Gt of C/ year (Ryther, 1969; Walsh, 1984; Martin 1992). Ocean-atmospheric coupled climate models predict changes in the ocean circulation and hypothesize that changes in the ocean circulation will stimulate phytoplankton biomass production in the nutrient depleted areas in the open ocean (Roemmich and Wunsch, 1985). The effect on atmospheric CO_2 is uncertain because the relationship between the enhanced primary production and air-sea exchange of CO_2 is not understood. The challenge is to study the magnitude and variability of primary productivity, its time scales and changes in atmospheric forcing and upscale it into secondary and tertiary productivity.

Relevance to Northern Indian Ocean (NIO)

The Northern Indian Ocean (NIO) comprises a unique variety of biogeochemical provinces, including eutrophic, oligotrophic, upwelling, and oxygen-depleted zones, all within an area of relatively small geographic extent. This reflects the pronounced semi-annual reversals in regional winds (the seasonal monsoons) that make this region a focus for intense study. Previously published sea-air flux estimates indicate that the NIO could account for 12-52 percent and 0.1-133 percent respectively of the known oceanic sources of N_2O and CH_4. Even though the uncertainties are large, particularly for CH_4, the Arabian Sea/NWIO could be an important contributor to the marine CH_4 source and a dominant global source of atmospheric N_2O. The atmospheric inventories of N_2O and CH_4 are currently both increasing by about 0.3 percent per year. Both gases are strongly active and together account for 18 percent of enhanced greenhouse forcing. N_2O is implicated in the generation of stratospheric NO_2, which influences stratospheric O_3 levels, and CH_4 participates in the photochemistry or tropospheric O_3 and OH and in the formation of stratospheric H_2O. Existing estimates of oceanic N_2O and CH_4 sources are dominated by disproportionately large contributions from biologically productive areas such as the NIO. However, because the spatial and temporal coverage of such regions remains limited, our global estimates remain unsatisfactory. Future climatically induced modifications to the upwelling and circulation characteristics of the NIO and other regions experiencing strong wind-driven upwelling may have profound effects on future biogas emissions from the oceans.

Estimation and integration of PP

Integrated *in-situ* column primary production (PP) can be estimated and computed at biome level using *in-situ* and satellite (SRS) remote sensing data by adopting suitable mixed layer PP model. SRS methods can be applied for computing primary productivity to integrate at biome level.

---

* *Grinson George, Senior Scientist, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

Chlorophyll is an important indicator of the quality of aquatic ecosystems that are amenable to *in situ* and space borne measurement. This property can be retrieved from ocean colour data after removal of the atmospheric signal from the detected radiance. Phytoplankton blooms (indicated by rapid increase in chlorophyll concentration) and spurts in primary productivity are important for maintaining the marine organisms at higher tropic levels, but when associated with eutrophication and harmful algal blooms, as noticed in the coastal waters of India, such events are directly linked (negatively) to the quality of water. Another important measure of water quality in the coastal environment is the suspended sediment load. Together with chlorophyll concentration they determine water light penetration, and light available for photosynthesis. Optical instruments such as spectral radiometers are able to monitor changes in chlorophyll and suspended sediment load in real time. Furthermore, such measurements can form the basis of local algorithms for application in remote sensing, allowing the results to be extrapolated to the entire study area through remote sensing. Optical methods for monitoring water quality and productivity have been established in other marine environments, for example in the USA. In India, a start in this direction has been established and operationalized by the SATCORE programme of ESSO-INCOIS.

Marine resources, especially fishery resources, have a strikingly important place of prominence in the biodiversity map of the earth. Their dynamics have very important influence; both direct as well as derived, on the wealth, health and eco-balance of many a maritime nations. Context to the aforementioned issue in the region can never be overstated with a prominent chunk of future requirement of socio-economic and nutritional sustenance centered in the marine sector. Towards establishing a scientifically deduced relationship between the marine environment and the resource availability on a realistic basis, there is a need for a focused application of established easy-to-surveil oceanic, geophysical and physicochemical parameters and their direct or latent influence upon plankton, which happens to be the self-replenishing source of food and nutrition for fishery resources. The spatio-temporal fluctuation of plankton richness, which can be remotely sensed, has long been established as a major factor in predicting resource richness in general and congregation and catchable availability in particular. Taking cue from these established models, paradigms can be designed to predict the resource availability from easy-to-observe parameters after a thorough validation of the prediction scenarios juxtaposed with the estimated catch attributable to various fishing grounds. The change in the pattern of fishing, period of absence and the composition of fish caught per haul, when analyzed for a range of geo-spatial expanses would help refining and augmenting the existing paradigms resulting in a comprehensive prediction algorithm. Further, such models would come in handy in the assessment of marine resource potentials and their periodic revalidation on a homogenous platform with a proper measure of confidence interval.

A simple exercise to estimate biomass from primary productivity for conceptualizing the idea

Authors	**Estimated 1^0 productivity**	**Extrapolated fish production**	**Remarks**
Riley, 1945 in Rabinowitch,, 1945	375 kg C/km^2 annually= 3.75 tonnes/ha	**15.5 million tonnes** (Indian Ocean)	8 times higher than terrestrial productivity
Steeman Nielsen and Jensen, 1957 Galathea expedition	40% for respiration from net productivity averages globally 1.2-1.5*10^6 tons	**2 million tonnes** (Indian Ocean)	Average annual production of hydrosphere similar to terrestrial productivity
Steeman Nielsen and Jensen, 1957	Eutrophic area productivity high	0.2-0.3% of fixed carbon as fish removed annually	High level of efforts in coastal waters with active fishery
Rhyther, 1959	Seasonal maxima also addressed	**3 million tonnes** (Indian Ocean)	Sea twice as productive as land
Schaefer, 1965	1.9*10^6 tons of organic carbon for all seas as average	200*10^6 tonnes for world oceans **40 million tonnes** (Indian Ocean)	Fish production 0.03% of potential
Raghuprasad *et al.*, 1969	Compilation of all the above	100 million tonnes (world oceans) **20 million tonnes** (Indian Ocean)	0.4% of potential harvested

(All the estimates were based on primary production – Organic carbon biomass generated by the producers)

Calculation of potential estimates of fishery from primary productivity estimates for Indian Ocean basin scale (Raghuprasad *et al.*, 1969)

Average annual productivity of Indian Ocean (Anton Brunn survey)	:	3*10^9 tonnes of Carbon = 0.35 hC/m^2
Respiration requirement	:	40% of organic production
Average net production	:	0.24 gC/m^2/day (Western Indian Ocean) 0.19 gC/m^2/day (Eastern Indian Ocean)
Area	:	29*10^6 km^2 (Western Indian Ocean) : 22*10^6 km^2 (Eastern Indian Ocean)
Net production of carbon	:	2.3*10^9 (Western Indian Ocean) 1.6*10^9 km^2 (Eastern Indian Ocean)
Total fish yield (0.03% of net production	:	12.6 million tons In 1967 the production was 2.1 million tons. A six fold increase in catch is possible as per the potential estimated

Estimates based on ecological efficiency

- Estimates of potential yield on annual basis is calculated and the potential biomass at the safest level (@10% ecological efficiency level)
- 23 million tons of fish from Western Indian Ocean and
- 15 million tons from Eastern Indian Ocean
- Total of 38 million tons possible from the entire Indian Ocean

Estimation of potential fish yield from zooplankton biomass

Zooplankton biomass estimated for Western Indian Ocean = $3.25*10^8$ tonnes

Zooplankton biomass estimated for Eastern Indian Ocean = $1.94*10^8$ tonnes

At 10 % ecological efficiency level

Theoretical estimate from carbon production for Western Indian Ocean = $2.3*10^9$ tonnes

Theoretical estimate from carbon production for Eastern Indian Ocean = $1.6*10^9$ tonnes

Potential fish biomass estimated for Western Indian Ocean = 18 million tonnes

Potential fish biomass estimated for Eastern Indian Ocean = 11 million tonnes

Total fish biomass estimated for Indian Ocean = **29 million tonnes**

References

Martin, J. H. 1992. Iron as a Limiting Factor in Oceanic Productivity, In: Primary Productivity and Biogeochemical Cycles in the Sea, P. G. Falkowski & A. D. Woodhead (eds.), pp.123-137, Plenum Press, ISBN 0306441926, New York, United States.

Rabinowitch E. I. 1945. Photosynthesis and related processes, Vol. I. Wiley Interscience, NY.

Raghu Prasad, R., S.K. Banerji and P.V. Ramachandran Nair. 1969. A quantitative assessment of the potential fishery resources of the Indian Ocean and adjoining seas. Indian J. Anim. Sci., 40: 73 – 98.

Roemmich, D. and C. Wunsch. 1985. Two transatlantic sections: Meridional circulation and heat flux in the subtropical North Atlantic Ocean. Deep-Sea Res. 32: 619–664.

Ryther, J. H., 1959. Potential productivity of the sea. Science, 130 : 602-608. Ryther, J.H. 1969. Photosynthesis and fish production in the sea. Science, 166: 72-76.

Schaefer, M.B. 1965. The potential harvest from the sea. Trans. Am. Fish. Soc. 9i(2) : 123- 28.

Steemann Nielsen, E. and A. Jensen. 1957. Primary oceanic production, the autotrophic production of organic matter in the oceans. Galathea Report 1, 47 -135.

Walsh, J. J. 1984. The role of the ocean biota in accelerated global ecological cycles: a temporal view. *Bioscience,* 34, 499-507.

***

Chapter 12

Recent Approaches in Stock Assessment

12.1

New Methods of Fish Stock Assessment[*]

Introduction

With the advancement of computer programmes, several new models are being used for fish stock assessment in recent years. These models attempt to address the complexities in fish stock dynamics, including integrating climatic and environmental data into the model. Improvements in stock assessment is necessitated by increasing demands for improved information to support fishery conservation and management efforts. Thus stock assessments remain an area of active research. Some of the recent approaches in stock assessment are outlined below:

1. Multi-species Surplus Production Model

This is a multivariate version of single species surplus production model. Here, the annual surplus production (ASP) is calculated for each stock as:

$$\mathrm{ASP}_{j,t} = B_{j,t+1} - B_{j,t} + \delta_j C_{j,t}$$

where $B_{j,t}$ is the estimated “adult” biomass of stock *j* at the beginning of year *‘t’*, $C_{j,t}$ the catch of stock *‘j’* during year *‘t’*, and $ä_j$ is a stock-specific correction factor that accounts for growth and mortality that would have taken place between the time the catch was taken and the beginning of year *t* + 1. Assuming an additive error structure for annual surplus production, the estimating equations take the form of a multiple linear regression for the Graham-Schaefer model and a non-linear regression for the Pella-Tomlinson model (Quinn and Deriso, 1999 - Quantitative Fish Dynamics. Qxford University Press. New York):

Graham-Schaefer : $ASP_t = \alpha\,\bar{B}_t + \beta\,\bar{B}_t^2 + \varepsilon_t$

Pella-Tomlinson : $ASP_t = \alpha\,\bar{B}_t + \beta\,\bar{B}_t^{\nu} + \varepsilon_t$

where á, â, and í are model parameters and ε_t are model residuals that are assumed to be normally distributed.

2. Multi-species Virtual Population Analysis (MSVPA) model

Multi-species virtual population analysis is an extension of the VPA model for simultaneous analysis of data for more than one species that incorporates the predator stomach content data into the virtual population model. In MSPVA, through a recursive algorithm, the fishing mortality at different age, recruitment, stock size, suitability coefficients and predation mortality are calculated based on catch-at-age data, predator ration and predator diet information. MSVPA allows the estimation of vital population rates used in the management of fishery resources. An additional advantage of the model is the estimation of the predation mortalities produced by predators on prey species and the annual consumption of prey by predators. The MSVPA input data includes the catch-at-age data, percent of maturity-at-age, weight-at-age, terminal fishing mortalities, predator stomach content data and residual mortalities.

---

[*] *T.V. Sathianandan, Principal Scientist & Head of Division, Fishery Resources Assessment Division, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

3. Dynamic multi-species models:

These models consider the functional relationships among individual species in a fished system. They build upon single-species theory to understand the dynamics of multi-species fisheries. These models account for interactions among selected species (often exploited fish species) but do not address the ecosystem as a whole. Dynamic multi-species models consider predator–prey interactions and evaluate the interactions between a subset of the species in the ecosystem. They do not model competitive interactions explicitly, but often include constraints such as conservation of total system biomass, or constant input of food from outside the model, which results in changes in abundance of one species indirectly affecting the abundance of species with which it shares prey. Example: Virtual Population Analysis (VPA) models allowing for cannibalism, multispecies VPA (MSVPA) and statistical assessment models (SAM; single-species with predation).

4. OSMOSE model

OSMOSE is a multi-species/single species model for fish species. The model assumes predation based on spatial co-occurrence and size and represents fish grouped into school characterized by their size, weight, age, taxonomy and geographical location. The processes considered in the fish life cycle are growth, explicit predation, natural and starvation mortalities, reproduction, migration and fishing mortality distinct for each species. OSMOSE has been first applied to the Benguela upwelling ecosystem for which 12 fish species have been specified, from small pelagic fish to large demersal species. The model needs basic parameters that are often available for a wide range of species. For output, a variety of size-based and species-based ecological indicators can be simulated and converted to *in situ* survey and catch data at the species and community level. The model can be calibrated to observe biomass dynamics.

5. Atlantis

Atlantis is an ecosystem model that considers all the components of marine ecosystems, namely biophysical, economic and social. It is a deterministic biogeochemical ecosystem model with its overall structure based around the Management Strategy Evaluation (MSE) approach. There are sub-models (or modules) for each of the major steps in the adaptive management cycle. Deterministic biophysical sub-model is at the core of the model, coarsely spatially-resolved in three dimensions, which tracks nutrient flows through the main biological groups in the system. The primary ecological processes considered in the model are consumption, production, waste production, migration, predation, recruitment, habitat dependency, and mortality. The trophic resolution is typically at the functional group level. The physical environment is represented via a set of polygons matched to the major geographical and bioregional features of the simulated marine system. The biological model components are replicated in each depth layer of each of these polygons.

Atlantis also includes a detailed exploitation sub-model. This model is focused on the dynamics of fishing fleets and also deals with the impact of pollution, coastal development, environmental (*e.g.* climate) change. It allows for multiple fleets, each with its own characteristics of gear selectivity, habitat association, targeting, effort allocation and management structures. It includes explicit handling of economics, compliance decisions, exploratory fishing and other complicated real world concerns.

The sampling and assessment sub-model in Atlantis is designed to generate sector dependent and independent data with realistic levels of uncertainty measurements. These simulated data are based on the outputs from the biophysical and exploitation sub-models, using user-specified monitoring scheme. The data are then fed into the same assessment models used in the real world, and the output of these is input to a management sub-model. This last sub-model is a set of decision rules and management actions, which can be drawn from an extensive list of fishery management instruments such as gear restrictions, days at sea, quotas, spatial and temporal zoning, discarding restrictions, size limits, bycatch mitigation, and biomass reference points.

6. Size spectrum model

Charles Elton introduced the “pyramid of numbers” in the late 1920s, but this remarkable insight into body-size dependent patterns in natural communities was in fallow until the theory of the biomass size spectrum was introduced by aquatic ecologists in the mid-1960s. They noticed that the summed biomass concentration of individual aquatic organisms was roughly constant across equal logarithmic intervals of body size from bacteria to the largest predators. These observations formed the basis for a theory of aquatic ecosystems, based on the body size of individual organisms, that revealed new insights into constraints on the structure of biological communities. Size spectrum is the distribution of biomass/abundance as a function of individual mass or size. The shape of this function resembles a power function and biomass size spectrums are represented using power functions. Spatial and temporal variability in the community structure can be observed in the shape of biomass size spectra.

7. Stock synthesis

In the history of fish stock assessment two different approaches have dominated. One using time series of an indicator of stock abundance (standardized catch rate as a proxy for stock abundance) along with time-series of fish catch (Schaefer, 1954 surplus production model). These models provide inference about current and target fish stock abundance and the maximum sustainable yield. The second approach depends on a time-series of detailed fish catch-at-age data in order to reconstruct the virtual abundance of each annual cohort that had been fished (Pope, 1972 – Virtual Population Analysis, VPA). In the last two decades there has been development of a third approach known as Integrated Analysis (IA) that takes a more inclusive approach to modeling fish population dynamics, utilizing a wide range of available data. Stock Synthesis (SS), implementation of IA, began during the early 1980s. Synthesis is a term used for development of a new product that is more than blend of its dissimilar parts. In fish stock assessments, different kinds of data can provide complementary information about the fish stock, but one source may not be sufficient in itself to provide a complete picture of the stock’s abundance and the impact of fishing on the stock. Stock synthesis inherently blends the population estimation paradigm of VPA with the population productivity paradigm of biomass dynamics models. The observations that can be included in SS are CPUE, effort, survey abundance, discards, length, age, weight composition data and tag-recapture data. It has the capability to use time-series of environmental and ecosystem factors to influence the population dynamics and observation processes over time. Three stages of SS assessment approaches are – initial development (basic concept), re-development as a generalized model and development of the computer code in ADMB (Automatic Differentiation Model Builder).

The new stock assessment approaches are broadening the types of information being considered and incorporated into the assessments. In recent years, the capacity to evaluate the influence of ecosystem and socio-economic dynamics is increasing. These new tools will facilitate sustainable fisheries and will provide a framework for exploring the social, economic, and ecological trade-offs in sampling and management strategies.

***

12.2

Genetic Stock Characterization of Fish using Molecular Markers[*]

Introduction

Accurate Identification of genetic resources is necessary for detecting new species and varieties for products of commercial value. Fish, as a group, apart from their economic value from a biodiversity viewpoint, have the highest species diversity among all vertebrate taxa. They exhibit enormous diversity in size, shape, biology and in the habitats they occupy. In terms of habitat diversity, fishes live in almost all conceivable aquatic habitats, ranging from Antarctic waters to desert springs. Of the 62,305 species of vertebrates recognized world-over, 34,090 (nearly 52 %) are valid fish species; a great majority of them (97 %) are bony fishes and the remaining (3 %) are cartilaginous (sharks and rays) and jawless fishes (lampreys and hagfishes). Further, on an average, 300 new fish species are described each year, and global surveys indicate that there could well be at least 5,000 species more to be discovered.

Loss of biodiversity is one of the greatest challenges facing modern society. This environmental crisis is increasingly evidenced by the loss or deterioration of genetic resources and habitats, as well as recent attempts to highlight and address the issue at the highest international levels. Appropriate conservation efforts for protection of the natural biological wealth warrant right attention for their sustainable utilization and for posterity. Public concern for biodiversity conservation has risen in the last 50 years and led to national and inter-national policies, legislation, and actions to conserve biodiversity, notably the Convention on Biological Diversity (CBD). To conserve and sustainably utilize the bioresources of the country and for maintaining sovereignty over them, several nations enacted the Biological Diversity Act (BDA). This encompasses guidelines to address a wide range of issues related to the utilization of bioresources and information within the country as well as by other countries.

Management of Fish Genetic Resources

The objective of management (documentation + conservation + sustainable utilization) of species and their habitats is to maintain the genetic identity and integrity of the species in their natural habitat as well as a genetically sustainable fishery. Hence, documentation of genetic variation and diversity is of vital significance to evolve conservation strategies with long-term impact. Genetic resources can be viewed as genetic differences at three hierarchical levels of organization, *viz.*, species, populations and individuals. At the highest level, species consist of '**populations'** or **'genetic stocks'** that are reproductively isolated from populations of other species. Each species harbours a unique set of genetic material and therefore, conservation may aim at a specific species, which requires sound knowledge about its biology, biogeography and within species (inter-populational) level genetic diversity. At the population level of organization, the identification of **discrete genetic breeding units** (usually called a '**stock**' in fisheries biology; this is roughly equivalent to a **'population'** or **'genetic stock'** to

---

[*] *A. Gopalakrishnan, Director, ICAR- Central Marine Fisheries Research Institute, Kochi - 682 018, Kerala, India.*

a geneticist) has been a major theme in fisheries research. The definition of a stock can vary, as the motivations of fishery managers may be influenced by political, economic or biological mandates. **Finally, the largest store of genetic variability in most species exists as genetic differences among individuals within a population**. Hence, the goal of pre-serving genetic variability in a population coincides with the goal of maintaining large ecologically sound natural populations. A fundamental need is to define distinct entities that range from individuals to species to ecosystems and beyond.

Population/Genetic Stock Identification (GSI)

Assessment of genetic variability is important for the management of wild genetic resources of fish. Most species are composed of **populations**, also called **genetic stocks**, between which limited gene flow occurs. These populations maintain their genetic makeup or characteristics distinct from other populations of the same species because of genetic variation within the species. This differentiation depends upon forces such as migration, mutation, selection, and genetic drift, which act on the species/population during its evolution. If such units are overfished, it is unlikely that population sizes will recover because of migration, and hence a collapse of the fishery may occur. Therefore, with the loss of a genetic stock, a species also loses the animals that are adapted to a particular habitat through evolution. Moreover, interbreeding of non-native fish stocks/species with a different make up tends to reduce the genetic variation that naturally exists between genetic stocks. In other words, the different natural genetic identities available for a species in different habitats are lost.

A fundamental problem for fisheries management is the identification of populations/stock of a species and this idea has been brought together with the definition of stock for management. The term stock has been used in various management contexts with little or no genetic content. Several approaches have been advocated to solve this problem.

- Ihssen *et al*. (1981) defines a stock as '*an intra-specific group of randomly mating individuals with temporal or spatial integrity'.*
- Larkin (1972) defined a stock as *'a population of organisms which share a common gene pool, is sufficiently discrete to warrant consideration as a self-perpetuating system which can be managed'.*
- In fishery management*, a unit of stock is normally regarded as a group of fish exploited in a specific area or by a specific method*.

If fishery managers are to include genetic considerations in their decisions, they will need information on the biological differences between discrete local groups of a species and they will need to understand the genetic and ecological processes that influence discreteness. Thus, the implementation of management strategies based on molecular genetic data can have indirect benefits for population biodiversity, as the main objective of such management plans is to avoid population crashes, which in turn benefits the maintenance of population genetic diversity.

Molecular Genetic Markers: The primary objective of the **genetic stock identification (GSI)** in fish is to assess the distribution and pattern of genetic variability at intra-as well as interspecific population levels. The first priority for such research is identification of appropriate **molecular genetic markers** to assess genetic diversity. Fish stock identification was initially based solely upon morphological and meristic differences. Because these characters can be influenced by the environment, their variations may not have a genetic basis, and hence do not necessarily provide information on genetic and evolutionary relationships. In the 1950s, dissatisfaction with performance of phenotypic methods for stock identification encouraged early exploration of genetic markers. The markers developed have spurred development of statistical algorithms and revolutionized the analytical power necessary to explore genetic diversity among populations. Methods that take advantage of naturally occurring genetic markers have attracted a good deal of attention because application of physical tags is very labour intensive, and biological markers, such as scale-patterns, can vary dramatically from year to year. The first GSI methods using **soluble proteins** and **gene products** such as **allozymes** (enzymes at cellular level) for estimating the contributions of two or more salmon stocks to a mixed harvest were developed in the late 1970s. Since then, the rapidly expanding availability of highly variable genetic markers and refinements in statistical analyses have considerably increased the ability to analyze the stock structure of different fish species; but this has also led to the genetic 'marker wars' among fish geneticists during the past several decades. For many years, allozymes were the universal workhorse genetic makers, and they made many valuable contributions to basic and applied conservation and management. Around 1980, the first applications of **mitochondrial DNA (mtDNA)** analysis to natural populations were published, and gradually, it replaced allozymes and provided answers to key management questions regarding stock structure. The development of DNA amplification using the **polymerase chain reaction (PCR)** technique has opened up possibility of examining genetic changes in populations over the past 100-years or more even using archive material. In PCR reaction, a DNA sequence can be amplified many thousand folds to provide sufficient product for restriction analysis or direct sequencing. Once appropriate primers are available, large number of individuals can be assayed quickly thus facilitating large population screening for variability. Portions of the mtDNA such as, the ATPase 6 and 8 and hypervariable trans-membrane segments of cytochrome *b* (Cyt*b*) that evolve exceptionally rapidly have been used for high-resolution analysis of genetic stock structure in fish. Although mtDNA has indeed provided a wealth of new insights, it is not a solution and has some limitations with respect to fishery management (*e.g.,* it is maternally inherited, so provides information only about female migration or gene flow, and it is only a single marker and hence has much less power than a full suite of nuclear markers).

In the 1990s, **microsatellites** (**Short Tandem Repeats**— **STRs** or **Simple Sequence Repeats— SSRs**) muscled aside mtDNA and these highly variable **co-dominant markers** have provided greatly increased power and opened up exciting new opportunities (*e.g.* parentage analysis and individual assignments) that were generally not feasible with allozymes or mtDNA. Microsatellites are repeated DNA sequences having a unit length of 2-6 base pairs tandemly repeated minimum 6 times usually; maximum several times at each locus. They are found in all prokaryote and eukaryote genomes investigated to date. Individual alleles at a locus differ in the number of tandem repeats of the unit sequence owing to gain or loss of one or more repeats and they as such can be differentiated by electrophoresis according to their size.

There are four types of microsatellites:

1. Perfect: Perfect tandem repeat sequences.
2. Imperfect: Tandem repeat sequences with intervening sequences.
3. Compound: More than one kind of repeats, adjacent ones.
4. Complex: More than one kind of repeats, with intermediary sequences.

Based on the number of base pairs in a repeat unit, microsatellites can be again classified into *mono* (*e.g.* C or A), *di* (*e.g.* CA), *tri* (*e.g.* CCA), *tetra* (*e.g.* GATA), *penta* (*e.g.* CGATA) and *hexa* (*e.g.* ATGGCA) repeat unit microsatellites. Microsatellites that are used in stock identification studies typically contain di- (AC)n, tri-(ACC)n, or tetra-nucleotide (GATA)n repeats. The most common ones are dinucleotide repeats. Tetra-nucleotide microsatellites are gradually replacing dinucleotide loci as the preferred genetic marker for stock analysis. Microsatellite loci are abundant in all eukaryote genomes and it has been estimated that there are from 10^3 to 10^5 microsatellite loci dispersed at 7- to 10-100 kilobase pair (kb) intervals in the eukaryotic genome. Fish genomes may contain more microsatellite loci than most other invertebrate and vertebrate taxa. Mapping studies suggest more or less even distributions of microsatellites throughout genomes, although they are somewhat rarer within coding sequences.

Several features of STR render them invaluable for examining fish population structure. Microsatellites are codominant in nature and inherited in Mendelian fashion, revealing polymorphic amplification products from all individuals in a population. They contain information, which are directly related to the effective number of alleles at each locus. PCR for microsatellites can be automated for identifying simple sequences repeat polymorphism. Small amount of samples of blood or alcohol preserved tissue is adequate for analyzing them. Because they are highly variable in nature, abundant variants are ensured for characterization of populations. However, sample size in excess of 50 may be required to represent the genotype frequencies. The microsatellites are non-coding and therefore variations are independent of natural selection. These properties make microsatellites ideal genetic markers for defining population genetic diversity and distance measures. Because most STR loci are unlinked and inherited independently, the greater the number of loci screened, the greater the likelihood of selecting loci that reveal significant allelic frequency differences among populations and more statistical power is gained in quantifying the extent of genetic differentiation among populations. Additionally, analysis of a larger number of loci may provide a more accurate picture of the evolutionary history of the genetic stocks.

Analysis of microsatellite polymorphisms is a PCR-based approach in which oligonucleotide primers are designed based on unique single-copy sequences flanking the microsatellite repeats. DNA extracted from tissue samples are subjected to PCR reactions. PCR primer pairs are selected such that PCR products are of small molecular size (usually <350bp), providing relative ease in amplification from low-quality DNAs and also allowing for distinguishing small differences in the molecular size of alleles among individuals by using polyacrylamide-gel electrophoresis or automated DNA sequencers. Ideally, each individual shows a single (homozygote) or two-band (heterozygote) DNA pattern, with one band inherited from each parent. Polymorphic alleles at a locus are usually characterized by their molecular sizes. For dinucleotide repeats, these will differ by two base units. Based on the STR allele frequency data, powerful statistical tests are employed to arrive at a decision whether the genetic stocks of a species are significantly different from one another.

However, the field now seems poised to shift towards another type of marker, **single-nucleotide polymorphisms (SNPs)**. Like allozymes, SNPs are generally diallelic, so each marker has less power than a single microsatellite locus. They occur in vast numbers throughout the genome; therefore, eventually large overall increases in power are possible. Furthermore, once developed, SNPs can be assayed more reliably and cheaply than microsatellites, which could be a considerable advantage in large-scale fishery management applications. However, development of sufficient numbers of SNP markers will be neither easy nor cheap, and analytical issues such as minimizing ascertainment bias remain to be resolved. Despite growing competition from new genotyping and sequencing techniques and latest class of markers, the use of the versatile and cost-effective microsatellites continues to increase, boosted by successive technical advances. Next-generation sequencing (NGS) technologies and the rise of commercial services allow the identification of large numbers of microsatellite loci at reduced cost in non-model species. As a result, more stringent selection of loci is possible, thereby further enhancing multiplex quality and efficiency. Numerous examples also exist where microsatellite analysis is used for fish population analysis and management of Pacific salmon (Fisheries and Oceans Canada website: http://www.pac.dfo-mpo.gc.ca/science/facilities-installations/pbs-sbp/mgl-lgm/proj/index eng.htm online.) and also for cod where microsatellites have even been used as evidence in a court cases against a fishermen claiming a false origin of his catch. Use of 20-25 polymorphic microsatellite loci (preferably tetra-nucleotide repeats) and 70–100 individuals from each population has become the standard and scientifically accepted protocol for population genetic analysis of fish along with information on biology and morphometry (*TRUSS*) data. Sequence information of mitochondrial complete ATPase 6/8 and Cytb genes of at least 20 individuals per population are also often generated along with this.

Genetic stock structure in fish

Distinct population structure has been observed in many fish species across the world indicating that propagation-assisted restoration programmes must be stock-specific to replenish declining populations. Generally, between populations of marine and freshwater species, marked differences exist in the level of genetic differentiation and genetic diversity, with marine species generally exhibiting lower levels of inter-population differentiation and greater genetic diversity. This is mainly due to the higher effective population sizes and/or higher inter-population migration rates in marine environments compared with freshwater. In addition, marine fishes and invertebrates are generally broadcast spawners and hence have large potential for movement between areas by larval drift in currents. In addition, adults of many species are capable of making long distance migrations. Early genetic studies of commercially important marine fishes using allozymes and proteins indicated that they generally had moderate levels of gene diversity and little population sub-division, often covering over several hundred kilometers. However, unexpected fine-scale population sub-structuring and deep genetic lineages have been observed in recent studies with high-resolution markers in many fishes, which calls for further in-depth integrated approaches of molecular genetics with life-history traits. This will prove whether the variability is due to isolation or adaptations to particular marine habitats or as a result of non-genetic factors such as large reproductive variation among families. Regular monitoring of populations is also essential to enable a distinction between normal population-size fluctuations and those severe enough to warrant conservation measures.

The greatest genetic threats in the marine ecosystem are the extinction of genetically unique sub-populations and loss of genetic diversity primarily through overfishing and climate change. Illegal, unreported and unregulated (IUU) fishing also contributes to this condition, and thus poses a severe threat to marine ecosystems. Controlling for compliance and enforcing fishing regulations is hampered by difficulties in identifying the geographical origin of fish and fish products, at point of landing and further down the food supply chain. Presently, there are no validated genetic methods for identifying the geographical origin of marine fish and investigate commercial fraud. '**FishPop-Trace**'(https://fishpoptrace.jrc.ec.europa.eu/web/fishpop trace) is an international project, funded by the European Union (EU) framework programme (FP7), aiming to generate forensically validated reference panels of SNP markers for geographical origin assignment in four commercially important fish species, cod (*Gadus morhua*), hake (*Merluccius merluccius*), herring (*Clupea harengus*) and common sole (*Solea solea*). SNP markers are selected and these are subsequently genotyped across populations to provide high resolution data to analyze genetic variation. These markers are validated to be used as tags for traceability and enforcement applications leading to a reduction in IUU fishing and conservation of remaining marine resources.

For a successful stocking programme such as sea ranching of endangered seahorse or sacred chank, genetic structure of the original wild population must be determined before any new fish are released into the waters. This information can be used to develop hatchery guidelines for breeding fish for stocking purposes. By ensuring that the stocked population is having the same genetic make-up as the wild population, re-integration of the stocked fish will likely be more successful and deviations from the original genetic structure will be minimal.

Integrating population genetics data into marine fisheries management

Maintaining the maximum level of genetic variations in fish stocks is vital for the preservation of genetic resources. Therefore, excessive loss of genetic variability should be avoided for sustainable management of resources. Application of molecular marker techniques to a number of species has shown that these methods can provide information on genetic stock structure that can be of direct management relevance. However, such information has not always been incorporated into fishery management and policy decisions in several countries. The complex problem requires agreement among scientists, governmental organizations and policy makers to define and implement policies on the sustainable management of these natural resources. Numerous factors (as mentioned below) have contributed to the imperfect integration of genetic data into management of aquatic species.

The fish stock assessment teams generally include quantitative fishery biologists and statisticians. In appropriate situations, the teams should be expanded to include geneticists as well as field biologists. It is always better that fish geneticists fully understand the complexities of the management process so that genetic information can be packaged in the most effective manner, and importance of GSI can be portrayed effectively for the policy makers. Also the managers involved in monitoring of fishery resources should acknowledge that GSI can provide valuable management information. Scientists, managers and policy makers could work together more effectively to foster productive dialogue to link statutory definitions and management or conservation goals.

It is difficult to develop an ideal sampling design for a genetic study without understanding the details of the life history of the target species and physical processes in the aquatic ecosystem. Genetic data can be integrated with other types of biological and oceanographical information. The sampling design of genetic studies does not always match the geographical regions to which management controls are applied. This can rarely result in discrepancy between biological and genetic management units. Implementing GSI over a broad geographical area requires extensive efforts to collect baseline data for populations from different coasts and to standardize laboratory procedures so that comparable data can be obtained by different laboratories. This requires funds, broad collaboration among laboratories and a willingness to share unpublished data.

Most fish geneticists are unfortunately, not exposed to the techniques of statistical model and decision analysis that form the basis for modern stock assessment science. Equally, managers and assessment biologists similarly would benefit from a greater literacy regarding the genetic principles that can profoundly affect the aquatic living resources for which they share stewardship responsibility. Therefore, it might be necessary to develop brief **integrated training courses** to equip geneticists and managers to work on assessment teams.

The purpose of stock assessment in fisheries is to provide timely and appropriate scientific advice on fisheries management for sustained production. Though there are few multi-species models, the assessments are almost mostly conducted for single species, whereas in reality, stocks are influenced by multi-species interactions. In addition, gears mostly harvest many species at a time, leading to difficulty in implementation of the management measures derived from single species stock assessment. Due to the lack of adequate and efficient models for multi-species interactions, stock assessments will generally continue to be based on single species models. Although the main approach in population genetic studies of natural populations still involves collecting individuals from two or more geographical locations and considering them as putative populations, *landscape genetics/seascape genetics*—the study of spatial genetic patterns in continuously distributed species—is rapidly evolving and the methods are beginning to be applied especially to marine species as well. These studies are expected to provide important insights into biological processes leading to effective multi-species stock assessment and management of marine ecosystems. However, considerable dialogue between geneticists, stock assessment scientists and managers, as well as creative thinking on both sides are required to develop effective ways to integrate these insights into fisheries management.

In conclusion, fish genetic stock diversity conservation requires preservation of as much variation as possible at all taxonomic levels and concerted efforts by integrating capture, and culture fisheries and environmental programmes using latest technological innovations. The genetic tools will provide innovative means in the future and are an assuring approach for food security of the world and in reducing the fishing pressure on natural resources. Genetic data need to be integrated with other types of biological and oceanographical information for understanding the details of the life history of the target species and physical processes in the marine ecosystem. Although better monitoring of biodiversity, better assessment of risk and a more strategic approach to conserving biodiversity are all essential components to successful risk management, an equally important need is the open dialogue among geneticists, quantitative fishery biologists, statisticians, conservationists and planners that would help sustainable management of stocks of the world's amazingly rich assemblage of fishes.

References

Ayyappan S., J.K. Jena and A. Gopalakrishnan, 2014. Molecular Tools for Sustainable Management of Aquatic Germplasm Resources of India. *Agricultural Research* 3(1): 1-21.

Ihssen P.E., H.E. Booke H.E., J.M. Casselman, J.M. McGlade, N.R. Payne and F.M. Utter, 1981. Stock identification: materials and methods. *Can J. Fish. Aquat. Sci.,* 38:1838–1855.

Larkin P.A., 1972. The stock concept and management of Pacific salmon. *In*: R.C. Simon R.C. and P.A. Larkin (Eds), The stock concept in Pacific salmon, H.R. MacMillan Lectures in fisheries. University of British Columbia, Vancouver, 231 p.

***

Glossary

Age/Length composition: The proportion of fish of different ages/lengths in the stock or in the catch.

Allee effect: A situation in which fertilization rates may decrease in a small population, causing fewer recruits to be born per reproductively mature individual. The result is that it will take longer for the population to recover from small stock sizes.

Asymptotic length (L_∞): The maximum size that fish of a given species could reach if they lived forever. Sometimes it is mistaken for the largest observed size for the species.

Bayesian approach: A method of incorporating pre-existing knowledge into a model by defining the relative levels of certainty about the values of each parameter. Bayesian methods are particularly useful for making decision analyses.

Biological reference points: A benchmark against which the abundance of the stock or the fishing mortality rate can be measured in order to determine its status.

Biomass: The total weight of all the fish in the stock.

Biomass-at-MSY (BMSY): A biological reference point. It is the long-term average biomass value expected if fishing at FMSY.

Carrying capacity: The maximum upper limit to the size of a stock, determined by the availability of prey, the presence of predators, or other limitations of the habitat.

Catch-per-unit effort: The number or biomass of fish caught as a function of fishing effort.

Cohort: A group of fish born in the same year.

Catchability: The fraction of the stock which is caught by a standardized (effective) unit of effort. It is also used as the constant of proportionality that relates effective effort to fishing mortality ($q \times f = F$) or as the constant of proportionality that relates an index of abundance to absolute stock size ($I = q \times N$).

Catch curve: A graph showing the logarithm of catch in number of fish, plotted against fish age. Assuming equilibrium conditions, the descending limb of a catch curve can be used to estimate total mortality.

Catch rate or catch per unit effort: The amount of catch that is taken per unit of fishing effort (*e.g.*, number or weight of fish per hour in a trawl).

Cohort analysis: A simplified VPA algorithm based on an approximation that assumes that, in a given time period, all fishing takes place instantaneously in the middle of the time period.

Density-dependent growth rate: Growth rate model that assumes the growth rate for a stock is directly related to how close the stock is to reaching its carrying capacity.

Exploitation rate: The proportion of a population at the beginning of a given time period that is caught during that time period (usually expressed on a yearly basis). For example, if 100,000 fish were caught during the year from a population of 500,000 fish alive at the beginning of the year, the annual exploitation rate would be 0.20.

Exploitation ratio: The ratio of fish caught to total mortality (= F/Z).

Fishing effort: A measure of the intensity of fishing operations. How effort is defined depends on the type of fishery (gear) and often on the type of information available. For example, trawl effort can be expressed as hours; for longline fisheries, number of hooks.

Fishing mortality rate (F): The part of the total mortality rate that is caused by fishing. Fishing mortality is usually expressed as an instantaneous rate, and can range from 0 per year (for no fishing) to high values such as 1.0 or more per year. Fishing mortality should reflect all deaths in the stock that are due to fishing, not just those fish that are actually landed.

FMSY: A biological reference point. It is the fishing mortality rate which, if applied constantly, would result in Maximum Sustainable Yield (MSY).

Fully exploited stock: The term means that the stock is not being overexploited nor underexploited. This can be interpreted in an equilibrium yield sense as fishing at FMSY, or in a yield-per recruit sense as fishing at Fmax.

Gear selectivity: The probability that a fish of a certain length or age will be captured by a given gear.

Goodness of fit: Shows how well a given model fits the available data; a number of formal statistical approaches are applied to models to examine goodness of fit.

Growth overfishing: An action that occurs when mortality rates are outpacing growth rates in terms of the overall weight or biomass of the stock. It occurs when the fishing mortality rate is above Fmax. This means that individual fish are caught before they have a chance to reach their maximum growth potential.

Growth rate: (i) Intrinsic growth rate: A value that quantifies how much a population can grow between successive time periods. The intrinsic growth rate is often estimated with production models and plays an important role in evaluating the sustainability of different harvest levels. (ii) Individual growth rate: A value that quantifies how fast the average individual in the population grows in size or in weight.

Index of abundance: Numerical value used to demonstrate the trend in relative abundance over time.

Instantaneous mortality rate (Z): Rate at which fish are dying at any given moment. For mathematical purposes, the instantaneous mortality rate is typically converted into an annual rate.

Integrated analysis: Refers to stock assessment methodologies that attempt to integrate multiple sources of data into a single estimation framework. For example, an integrated assessment can attempt to fit observations on total landings by fleet, size samples of landings, discard estimates, size samples of discards, standardized CPUE by fleet, fishery-independent surveys, and tagging records on movement, growth and recoveries.

Intrinsic growth rate: Also called density-independent growth rate, a growth rate that does not depend on the abundance, or density, of individuals in the stock. The stock size will change at a constant rate regardless of the size of the stock.

L50: (i) The length at which a fish has a 50 percent probability of being retained by a gear if it encounters that gear. (ii) The median length of a fish stock at maturity; half of the mature individuals in a population first attain maturity at a length longer than L50 and half at a length shorter than L50.

Least squares: A statistical criterion for the estimation of parameters in regression. Least squares means that the method aims to minimise the sum of squared differences between the observations and the predictions from a model.

Limit reference point: A benchmark that should not be exceeded with any significant probability according to a given set of management objectives.

von Bertalanffy growth model: A model for incorporating growth directly into the stock assessment model by assuming that growth occurs most quickly at the youngest ages, slows gradually as the individual gets older, and eventually levels off. The size at which the individual levels off is referred to as L_{∞} ("L-infinity").

Maximum Economic Yield (MEY): Conceptually similar to Maximum Sustainable Yield, except that the objective is to maximize long-term profits.

Maximum Sustainable Yield (MSY): The largest average catch that can be continuously taken from a stock under constant environmental conditions.

Natural mortality rate (M): The part of the total mortality rate that is due to causes other than fishing (*e.g.,* predation, disease, cannibalism, and perhaps increasingly, environmental degradation such as pollution). These many causes of death are usually lumped together for convenience, because they are difficult to separate quantitatively.

Nominal: Refers to quantities as they are reported, before any analyses or transformations. Nominal catch is the sum of catches that have been reported as round weight or, equivalently, the landings (nominal catches do not include such measures as unreported dead discards). Nominal effort pertains to measures of fishing effort or vessel carrying capacity that have not been standardized.

Optimum yield (OY): The amount of catch that will provide the greatest overall long-term benefit to society. The OY takes into account the biology inherent in maximum sustainable yield, as well as economics and the attitudes of the public towards risk and environmental protection.

Otolith: The ear bone of a fish, which can be analyzed to estimate the age of fish like rings are used in trees.

Overfishing: Action that occurs when the fishing mortality rate exceeds a specific threshold. Overfishing may occur as (a) Growth overfishing; (b) Recruitment overfishing, or (c) Ecosystem overfishing.

Population: A group of individual fish of the same species located in a given area.

Population size (usually denoted by *N*): The number of individual organisms in a population.

Population dynamics: Describes the ways in which a given population grows and shrinks over time, as controlled by birth, death, and migration.

Production model: A population model that describes, using simple functions, how the population biomass changes from year to year (or, how biomass changes in equilibrium as a function of fishing mortality). The simplest production functions aggregate all of the biological characteristics of growth, natural mortality and reproduction into a simple, deterministic model using three or four parameters. Production models are primarily used in simple data situations, where total catch and effort data are available but age-structured information are either unavailable or deemed to be less reliable.

Recruitment: A characteristic of the fish, whereby the young fish, previously inaccessible to the fishing gear, become vulnerable due to growth.

Recruitment age/length: The age/length at which fish are considered to be recruited to the fishery. In stock assessments, this is usually the youngest age group considered in the analyses, typically age 0 or 1.

Recruitment overfishing: Occurs when the parent stock is reduced, by fishing to the extent that not enough young fish are produced to ensure that the stock will maintain itself.

Recruits: Offspring that have survived long enough to be counted as part of the stock.

Selection factor: The ratio of 50% retention length (lc), to the mesh size.

Selectivity: The probability of fish being retained in a fishing gear as a function of the length of the fish.

Spawning stock biomass (SSB): The total weight of the reproductively mature individuals in a stock (usually males and females combined, but sometimes female SSB, alone, is used). This quantity depends on the abundance of year classes, the exploitation pattern, the rate of growth, both fishing and natural mortality rates, the onset of sexual maturity, and environmental conditions.

Standardisation: Refers to the action that have been adjusted to be directly comparable to a unit that is defined as the “standard” one. Nominal CPUE is standardized to remove the effect of factors which are known not to be related to abundance. This means that the effects of factors such as vessel size or spatial availability, which clearly affect CPUE, are removed, *e.g.,* by adjusting all observations to the “standard vessel”.

Stock: Stock is a biological unit of one species forming a group of similar ecological characteristics and, as a unit, is the subject of assessment and management. However, there are many uncertainties in defining spatial and temporal geographical boundaries for such biological units that are 100 percent compatible with established data collection and geopolitical systems. For this reason, the term stock is often synonym with assessment/management unit. Unlike fish population, a stock is defined as much by management concerns (such as jurisdictional boundaries or harvesting location) as by biology.

Stock assessment: An evaluation of the past, present and future status of the stock that includes a range of life history characteristics for a species, such as the geographical boundaries of the population and the stock; information on age, growth, natural mortality, sexual maturity and reproduction, feeding habits and habitat preferences; and the fisheries pressures affecting the species.

Stock-recruitment relationship: A function that describes how recruitment varies with changes in the reproductive output (or biomass) of the parental stock. The stock-recruitment relationship is particularly important for the understanding of the sustainability of alternative harvesting regimes.

Surplus production: The amount of biomass produced by the stock (through growth and recruitment) over and above that is required to maintain the total stock biomass constant between consecutive time periods.

Terminal F: Refers to fishing mortality values in the last year for which data are available in an assessment.

Total mortality rate (Z): The sum of natural and fishing mortality rates.

Virgin: Refers to an unfished condition of the stock in an equilibrium sense. For instance, Virgin Biomass is equivalent to the stock´s carrying capacity.

Virgin biomass (Bo): A biological reference point. It is the long-term average biomass value expected in the absence of fishing mortality. In production models, B_0 is also known as carrying capacity.

Virtual Population Analysis (VPA): An age or length-structured population dynamics model, usually backward-projecting, that assumes that an exact value is known for the number of fish caught.

Yield: Catch in weight.

Yield-per-recruit: The expected lifetime yield for the average recruit. For a given exploitation pattern, rate of growth, and natural mortality, an equilibrium value of Y/R can be calculated for each level of F. Y/R analyses play an important role in advice for management, particularly as it relates to minimum size controls.

***

Suggested Reading

Alagaraja, K., 1984. Simple methods for estimation of parameters for assessing exploited fish stocks. *Indian J. Fish*. 31, 177-208.

Beverton, R.J.H. and S.J. Holt, 1956. A review of methods for estimating mortality rates in exploited populations, with special reference to sources of bias in catch sampling. *Cons. Int. Explor. Mer*. 140, 67-83.

Beverton, R.J.H. and S.J. Holt, 1957. On the dynamics of exploited fish populations. *UK Min. Agri. and Fish., Fish. Invest.* 19, 533 p.

Beverton, R.J.H. and S.J. Holt, 1966. Manual of methods for fish stock assessment. II. Tables of yield functions. *FAO Fish. Tech. Pap*. 38, 67 p.

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

Caddy, J.F., 1989. Marine invertebrate fisheries: their assessment and management. John Wiley & Son, Interscience, New York, 752 p.

Cadima, E.L., 2003. Fish Stock Assessment Manual. *FAO Fisheries Technical Pap*. 393, 161 p.

CMFRI, 2014. Fish Stock Assessment and Management. Training Manual. Central Marine Fisheries Research Institute, Kochi, 150 p.

Cooper, A.B., 2006. A Guide to Fisheries Stock Assessment- From Data to Recommendations. Sea Grant College Program, NOAA, USA, 44 p.

Csirke, J., 1988. Small shoaling pelagic fish stocks. In “Fish Population Dynamics” (J.A. Gulland, Ed.), John Wiley & Sons, Chichester, 271-400.

Cushing, D.H., 1973. Production in the Indian Ocean and the transfer from the primary to the secondary level. In “Biology of the Indian Ocean” (B. Zeitzschel, Ed.), Springer-Verlag, Berlin, 475-486.

Devaraj, M., 1983. Fish Population Dynamics. Course Manual. CIFE Bulletin, 3, 98 p.

FAO, 1978. Some scientific problems of multispecies fisheries. *FAO Fish. Tech. Pap*. 181, 42 p.

FAO, 1981. Methods of collecting and analyzing size and age data for fish stock assessment. *FAO Fish. Circ*. 736, 100 pp.

Ford, E., 1933. An account of the herring investigations conducted at Plymouth during the years from 1924-1933*. J. mar. biol. Assoc. U.K.* 19, 305-384.

Fox, W.W. Jr. (1970). An exponential surplus-yield model for optimizing exploited fish populations. *Trans. Am. Fish. Soc*. 99, 80-88.

Graham, M., 1935. Modern theory of exploiting a fishery, an application to North Sea trawling. *J. Coast. Expl. Mer*. 10, 264-267.

Gulland, J.A., 1955. On the estimation of population parameters for marked members. *Biometrika* 42, 269-270.

Gulland, J.A., 1965. Estimation of mortality rates. *Cons. Expl. Mer. C.M*. 1965, 9 p.

Gulland, J.A., 1983. Fish Stock Assessment: A Manual of Basic Methods. FAO/Wiley, New York, 223 p.

Gulland, J.A. and S.J. Holt, 1959. Estimation of growth parameters for data at unequal time intervals. *J. Cons. CIEM* 25, 47-49.

Holt, S.J., 1963. A method for determining gear selectivity and its application. *ICNAF Spl. Publ*. 5, 106-115.

Jones, R., 1977 . Tagging: theoretical methods and practical difficulties. In "Fish Population Dynamics", J.A.Gulland (Ed.), Wiley Interscience, New York, 46-66.

Jones, R., 1984. Assessing the effects of changes in exploitation pattern using length composition data (with notes on VPA and cohort analysis). *FAO Fish. Tech. Pap*. 256, 118 p.

Laevastu, T. and F. Favorite, 1988. Fishing and Stock Fluctuations. Fishing News Book Ltd. Farnhan, 239 p.

Laevastu, T. and R. Marasco, 1982. Fluctuation of fish stocks and the consequences of the fluctuations to fishery and its management. Fishing Book Ltd, Farnhan, 162 p.

Morales–Nin, H., 1991. Determination of growth of bony fishes based on micro-structures in the otoliths. FAO. DOC. TEC. PeSca, 322, 58 p.

Munro, J.L., 1982. Estimation of the parameters of the von Bertalanffy growth equation from recapture data at variable time intervals. *J. Cons. CIEM* 40, 199-200.

Pauly, D., 1979. Theory and management of tropical multispecies stocks: a review with emphasis on Southeast Asian demersal fisheries. *ICLARM Stud. Rev*. 1, 35 p.

Pauly, D., 1980. On the interrelationships between natural mortality, growth parameters and mean environmental temperature in 175 fish stocks. *J. Cons. Int. Explor. Mer.* 39, 175-192.

Pauly, D., 1983. Some simple methods for the assessment of tropical fish stocks. *FAO Fish. Tech. Pap.* 234, 52 p.

Pope, J.A., A.R. Margetts, J.M. Hamley, J.M. and E.F. Akyûz, 1975. Manual of methods for fish stock assessment. III. Selectivity of fishing gear. *FAO Fish. Tech. Pap*. 41, 65 p.

Powell, D.G. and J.A. Weatherall, 1979. Estimation of mortality and growth parameters from the length frequency of a catch. *Rapp. P.-V.Reun. CIEM*. 175, 167-169.

Prasad, R.R., S.K. Banerji, and P.V.R. Nair, 1970. A quantitative assessment of the potential fishery resources of the Indian Ocean and adjoining seas. *Indian J. Anim. Sci*. 40, 73-98.

Ricker, W.E., 1954. Stock and recruitment. *J. Fish. Res. Board Can*. 11, 559-623.

Rikhter, V.A and V.N. Efanov, 1976. On one of the approaches to estimation of natural mortality of fish populations. *ICNAF Res. Doc. 76/VI/8*, 12 p.

Schaefer, M.B., 1954. Some aspects of the dynamics of population important to the management of commercial marine fisheries. *Bull. IATTC* 1, 26-56.

Schaefer, M.B., 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. *Bull. I – ATTC / BOI. CIAT* 2 , 247-268.

Shepherd, J.G., 1988. Fish stock assessments and their data requirements. In "Fish Population Dynamics" (J. A. Gulland, Ed.), John Wiley & Sons, Chichester, 35-62.

Sparre, P. and C. Venema, 1992. Introduction to Tropical Fish Stock Assessment. *FAO Fish Tech. Pap*. 306, 376 p.

Srinath, M., 1998. Empirical relationships to estimate the instantaneous rate of natural mortality. *Indian J. Fish*. 45, 7-11.

Thompson, W.F. and H. Bell, 1934. Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield, and yield per unit gear. *Rep. Internat. Fish. Comm.* 8, 48 p.

Vivekanandan, E., 2002. Multispecies assessment of the demersal fish stocks along the southeast coast of India. *FAO RAP Publ*. 27, 24-37.

Vivekanandan, E., 2005. Stock Assessment of Tropical Marine Fishes. Indian Council of Agricultural Research, New Delhi, 115 p.

von Bertalanffy, 1934. Untersuchungen iiber die Gesetzlichteiten des Wachtums. I. Augemine Grundlagender Theorie. *Roux'Arch.Entwicklungs-mech. Org.,* 131, 613-653.

Walford, L.A., 1946. A new graphic method for describing the growth of animals. *Biol. Bull.* 90, 141-147.

***

Notes

भारतीय कृषि अनुसंधान परिषद
Indian Council of Agricultural Research
केन्द्रीय समुद्री मात्स्यिकी अनुसंधान संस्थान
Central Marine Fisheries Research Institute
(कृषि अनुसंधान एवं शिक्षा विभाग, भारत सरकार)
(Department of Agricultural Research and Education, Govt. of India)
पोस्ट बोक्स सं. 1603, एरणाकुलम नोर्थ पी.ओ., कोच्ची - 682 018, केरल, भारत
Post Box No. 1603, Ernakulam North P.O., Kochi - 682 018, Kerala, India
Telephone:+91-484-2394867, Fax: +91-484-2394909, 2396685
Email: director@cmfri.org.in, director.cmfri@icar.gov.in
Web: www.cmfri.org.in

BAY OF BENGAL PROGRAMME
INTER-GOVERNMENTAL ORGANISATION
91, Saint Mary's Road, Abhiramapuram, Chennai - 600 018, Tamil Nadu, India.
Tel: +91-44-24936294, 24936188; Fax: +91-44-24936102
Email: info@bobpigo.org; URL: http//www.bobpigo.org; www.oppbob.org