A note on different methods for standardization of fishing efforts

Eldho Varghese*, J. Jayasankar, T. V. Sathianandan, Somy Kuriakose, K. G. Mini, Reshma Gills, M. Muktha, V. Sreepriya and A. Gopalakrishnan
ICAR-Central Marine Fisheries Research Institute, Kochi-682 018, Kerala

*E-mail: eldho.varghese@icar.gov.in

Abstract

Surplus production models are widely employed to evaluate the condition of fish stocks, encompassing the entire stock, overall fishing effort, and the total yield derived from the stock. These models operate under the assumption that variations in population biomass result from hikes due to growth and reproduction, as well as drops due to natural and fishing mortality. Utilizing Catch-Per-Unit-Effort (CPUE) as input, these models rely on the presumption that CPUE is directly proportional to the biomass of fish stock in the sea. An inherent challenge in fitting such a production model lies in determining CPUE, whether in terms of units operated or in hours of operation/actual fishing hours (AFH) or in any measure of fishing efforts. Given the heterogeneous nature of fishing fleets in tropical regions, they are often categorized into boat-gear categories, where fishing units within each category share similar characteristics and performance. When assessing the collective impact of the fishing operations of the entire fleet on the exploitation of fish stock, nominal addition of the efforts of different boat-gear categories may lack meaningfulness without prior effort adjustment to enhance comparability. In tropical regions, due to the varying capacities of gears and the potential presence of multiple species in each gear, the effort expended to catch a resource cannot be simply considered as the sum of the duration/units of operation of all gears. This paper aims to underscore various effort standardization methodologies found in the literature for different situations, offering insights into the challenges faced in tropical fisheries and proposing a way forward.

***Keywords**: Surplus production models, stock assessment, sample-based surveys, standardization of fishing effort, abundance, GAM, GLM, GLMM, delta methods*

Introduction

Fish stock assessments are crucial for understanding the health and sustainability of fish populations. These assessments provide valuable information on the size, age, and productivity aspects of fish stocks, which can help managers make informed decisions about fishing regulations which may be furthered to harvest control rules/quotas. Additionally, fish stock assessments can help identify areas where habitat restoration or conservation efforts may be needed and play a critical role in maintaining healthy and sustainable fisheries. These assessments as they are practised worldwide, entail determining the parameters of population dynamics models by fitting them to research and monitoring data.

Surplus production models are commonly used for assessing the state of fish stocks and they deal with the stock, the fishing effort and the total yield obtained from the stock in their entirety. These models assume that variation in population biomass results from addition due to growth and reproduction and loss due to natural and fishing mortality. The results of this fitting process are then used to estimate quantities, such as current abundance, that are important for decisionmakers. When fitting stock assessment models, a range of data types can be utilized. However, these data sets generally need to include details about removals due to harvesting and an indicator of relative abundance. Ideally, this indicator of abundance should be derived from fishery-independent data collection methods like surveys. However, acquiring fishery-independent data is often prohibitively expensive or challenging. In such cases, reliance on fishery-dependent data becomes necessary. Consequently, assessments of numerous stocks rely exclusively on fishery-dependent data. The most common and easily obtainable form of fishery-dependent data is the catch and effort information from commercial or recreational fishers, typically presented as catch-per-unit-of-effort (CPUE) or catch rate (Maunder and Punt, 2004).

As the fishing fleet is heterogeneous in most cases, it is partitioned into boat-gear categories in each of which the fishing units have similar characteristics and performance. When it comes to measuring the combined effect of the fishing operations of the entire fleet to the exploitation of a fish stock, it becomes apparent that adding together effort exerted by different boat-gear categories is not always meaningful without first applying effort adjustment to increase their comparability (Stamatopoulos and Abdallah, 2015). Hence, standardization of commercial catch and effort data is important in fisheries where in standardized abundance indices based on fishery-dependent data are a fundamental input to stock assessments (Bishop, 2006).

The objective of this paper is to highlight diverse methodologies for standardizing fishing efforts as identified in the literature across various situations. It seeks to provide insights into the challenges encountered in tropical fisheries and put forward potential pathways for addressing them.

Methods for standardization of fishing efforts

There is a lot of literature available on the standardization of the fishing effort. These methods deeply depend on the characteristics of the gear being operated and the availability of the information. This choices/listing is more based on the generic nature of the underlying approach and relevance to multigear multispecies scenarios. Following are the few methods available in the literature for the standardization of fishing effort:

Standard vessel/gear based approach
(Beverton and Holt, 1957)

This method consists of selecting a reference gear/vessel and determining the relative fishing power/effort (RFP) of all other vessels/gears by

$$RFP_i = \frac{C_i/E_i}{C_s/E_s}$$

where RFP_i is the relative fishing power/effort for vessel/gear *i*, with C_i representing the total catch by vessel/gear *i* during the specified period when both the standard vessel/gear and vessel/gear *i* were present in the fishery. C_S represents the total catch by the standard vessel during the same period. E_i denotes the total days fished (or another measure of fishing effort) by vessel *i* during the specified period, while E_S represents the total days fished by the standard vessel during the same period.

The standardized catch rate for year *t*, is then defined as

$$I_t = \frac{\sum_i C_{t,i}}{\sum_i RFP_i \, E_{t,i}}$$

where $C_{t,i}$ is the catch by vessel *i* in year *t*, and $E_{t,i}$ the number of days fished by vessel/gear *i* in year *t*. This approach is a simple method for estimating fishery yields, but it may not be suitable for situations with multiple factors and when no long-term fishing vessels are available for comparison.

Relative effort based approach
(Robson, 1966)

A more direct approach to standardize fishing effort is proposed by Robson (1966), although it necessitates the availability of additional data. The method operates based on the notion of "relative fishing power". With the fishing power of vessel B relative to vessel A means:

$$PA(B) = \frac{CPUE \text{ of vessel } B}{CPUE \text{ of vessel } A}$$

applied when two boats are fishing under identical conditions (simultaneously and in the same area). Vessel A is commonly referred to as the "Standard vessel." Suppose the boats are participating a certain fishery can be divided into 5 homogenous groups, so that each group consists of boats with similar fishing powers. Suppose also that the CPUE is in units of catch per unit time and further that the following data have been collected:

Boat type	A (Standard)	B	C	D	E
Fishing Power (*PA*)	1.0	*PA (B)*	*PA (C)*	*PA (D)*	*PA (E)*
Number of Boats (*N*)	*NA*	*NB*	*NC*	*ND*	*NE*
Average number of fishing days per boat (*d*)	*dA*	*dB*	*dC*	*dD*	*dE*

The total effort would then be estimated by:

Total effort = 1.0**NA*dA + PA (B)*NB*dB + PA (C)*NC*dC + PA (D)*ND*dD + PA (E)*NE*dE*

In specific instances, one may infer that the fishing effectiveness correlates with certain attributes of the boat or gear, readily accessible, such as GRT (tonnage) or HP (horsepower), or their combination for trawlers, and, for instance, the quantity or length of nets for gill netters. Since the focus typically revolves around relative effort, the fishing power (PA) can be easily substituted with the characteristics of the boat or gear.

Derived effort based approach (Sparre, 1998)

In general, a suitable measure of fishing effort is the one that demonstrates a linear relationship with the catch rate (Sparre, 1998).

The relative effort is

$$\frac{Yield}{CPUE} = Effort \quad or \quad CPUE = \frac{Yield}{Effort}$$

Since the effort of different gears is assessed in terms of units per year and hence to ensure compatibility among various gear types (effort units), each unit needs to be converted into CPUE, which is then further converted into "relative CPUE". The relative catch per unit of effort of gear *i* (*i*=1 to *k*) in year *y* is defined as follows:

$$R_i(y) = \frac{CPUE_i\,(y)}{Mean\ \{CPUE_i(y1,\ y2,\ \ldots,\ yn\ )\}}$$

where $CPUE_i\,(y) = Y_i(y)/f_i(y)$ = catch per unit effort of gear *i* in the year *y*, $Y_i(y)$ = yield of gear *i* in the year *y*; and $f_i(y)$ = effort of gear *i* in the year *y*.

The total yield of the species under examination denoted as $Y_T(y)$, encompasses both the catch covered by the catch/effort sampling scheme and the unaccounted yield. When this total yield is divided by the weighted sum of relative CPUE values, it yields a quantity proportional to the total effort $R(y)$, as $Y_T(y)/R(y)$.

The normalized effort for the year *y* is $E(y) = \frac{Y_T(y)/R(y)}{Mean(Y_T/R)}$,

where $Y_T(y)/R(y)$ is the relative effort of year *y*, $Y_T(y)$= total yield of all gears (including gears for which effort is not known), $R(y) = \sum_{i=1}^{k} [R_i(y) * Y_i(y)/Y_E(y)]$ is the sum of related CPUE weighted by the yields in the year *y* and $Y_E(y) = \sum_{i=1}^{k} Y_i(y)$ is the sum of yields of gears for which effort is known (yield of sampled gears), per year.

Multigear mean standardization (MGMS) (Daniel et al., 2016)

The method named multi-gear mean standardization (MGMS) combines catch per unit effort data that standardizes catch per unit effort data across gear types (Daniel *et al.*, 2016). The calculation of MGMS begins by standardizing the CPUE data for each gear using a form of mean centering. First, the total catch (*TC*) of all *i* species in each observation *j* per unit of effort *e* is calculated as $\frac{TC_i}{e}$. Next, for each gear, the mean total catch per unit effort $\frac{\overline{TC_i}}{e}$ is calculated. To standardize the data for each gear, the CPUE of species *i* in observation *j* (C_{ij}/e) is divided by the mean total catch per unit effort across all observations, yielding:

$$MSC_{ij} = \frac{C_{ij}/e}{\overline{TC}/e}$$

where MSC_{ij} is mean standardized catch of species *i* in observation *j*. Once CPUE data for each gear are converted to MSC_{ij}, they can be combined across gears and the resulting sums provide the basis for further analysis.

Generalized linear models (GLMs) and generalized additive models (GAMs)

Approaches built on GLMs and GAMs represent statistical approaches employed to model the correlation between catch (response variable) and factors such as effort, environmental variables, and other covariates (predictor variables). These methods are adept at accommodating non-linear relationships and variability within catch data, providing flexibility for capturing intricate patterns. The standardization of catch and effort data is most commonly achieved through the application of Generalized Linear Models (GLMs), as introduced by Nelder and Wedderburn in 1972. Gavaris (1980) is recognized as a pioneer in utilizing the GLM approach for this purpose, marking the first instance of its application. He expanded upon the use of multiplicative models (Robson, 1966) for standardization by explicitly incorporating assumptions of log-normal errors. Gavaris (1980) employed an Analysis of Variance (ANOVA) model, exclusively incorporating categorical explanatory variables, on the natural logarithm of CPUE. Hilborn and Walters (1992) gave an excellent exposition on the use of Generalized Linear Models (GLM) for the standardization of fishing efforts.

GLMs are characterized by the statistical distribution governing the response variable, typically (though not always) the catch rate, and how a linear combination of certain explanatory variables correlates with the anticipated value of the response variable. The fundamental premise of a GLM lies in the assumption that the connection between a function of the expected response variable value and the explanatory variables follows a linear pattern.

$$g(\mu_i)=\boldsymbol{x}_i^T\boldsymbol{\beta} \quad \text{.................................1}$$

where g is the differentiable and monotonic link function, $\mu_i=E(Y_i)$, $\boldsymbol{x}_i$ the vector of size m that specifies the explanatory variables for the i^{th} value of the response variable, $\boldsymbol{\beta}$ is a vector (of size m) of the parameters, and y_i the i^{th} random variable.

To overcome the linearity assumption of GLMs, Generalized Additive Models (GAMs) were developed and are effective models in establishing a relationship between predictor variables and the response. GAMs offer the ability to represent a broader spectrum of response curves compared to GLMs. Many researchers are opting for GAMs over GLMs, particularly in fisheries science, where their predominant application mirrors that of GLMs—specifically, the standardization of abundance data.

Generalized additive models (GAMs; Hastie *et al.*, 2001) are extensions of generalized linear models that involve generalizing Eq. (1) by replacing the linear predictor by an additive predictor:

$$g(\mu_i)=\mu+\sum_{j=1}^{p} f_j(x_i) \quad \text{..........................2}$$

where f_j is a smooth function (such as a spline or a loess smoother). The degree of smoothness achieved is balanced against the deviance by a tuning constant, often chosen by cross-validation, so that estimation is by the method of maximum penalized likelihood rather than of maximum likelihood. This gives GAMs a partially non-parametric aspect.

Methods for zero-inflated data

Databases containing information on catch and effort frequently exhibit a substantial proportion of entries where the catch value is zero, despite a recorded non-zero effort. Instances where effort is marked as zero must be addressed, either as trivial cases if they coincide with zero catch or as errors necessitating resolution (e.g., removal) before conducting any analyses. This pattern is particularly pronounced for less abundant species and those categorized as bycatch. Regrettably, these species often represent crucial sources, if not the sole source, of data for standardized catch rate indices that track changes in abundance (Ortiz and Arocha, 2004). The prevalence of zero values can undermine the assumptions underlying the analysis, posing a risk to the reliability of inferences if not appropriately modeled, as emphasized by Lambert (1992). Moreover, the abundance of zeros can introduce computational challenges. The following approaches are commonly adopted to deal with zero inflated data:

(a) Zero-inflated models

Zero-inflated models are often used when dealing with data that has excess zeros. These models typically assume that the observed data is a mixture of two processes: one that generates zeros and another that generates the remaining values. Common models for zero-inflated data include zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) models. Rochman *et al.* (2017) attempted to standardize CPUE to estimate relative abundance indices based on the Indonesian longline dataset time series using GLM with Tweedie distribution. Setyadji *et al.* (2018) used GLM to standardize CPUE and to estimate relative abundance indices based on the Indonesian longline dataset. Six GLM models were considered viz., negative binomial, zero inflated Poisson, zero-inflated negative binomial, Poisson hurdle, and negative binomial hurdle models. AIC and BIC were used to select the best models among all those evaluated.

(b) Delta methods

Traditional Generalized Linear Model (GLM) analyses, relying on log-transformed data, assume that no CPUE observation equals zero. To address these challenges within a GLM framework, the delta-lognormal method (Pennington 1983, 1996; Lo *et al*., 1992) has been employed. This method handles zero catches separately, modelling them independently, and then employs a GLM for positive catches. The models for zeros and the GLM are then integrated to generate an abundance index. Delta-GLM and Delta-GAM models are extensions of GLMs and GAMs, respectively, used for standardizing catch and effort data. They focus on modelling the differences (delta) between observed and expected catch rates, allowing for better handling of count data and overdispersion.

Generalized linear mixed models (GLMMs)

Generalized Linear Mixed Models (GLMMs), as introduced by Pinheiro and Bates in 2000, expand upon the Generalized Linear Model (GLM) approach by allowing certain parameters in the linear predictor to be considered as random variables. This extension enables more flexibility in modelling and accommodates the inclusion of random effects. In recent analyses of catch and effort data, various studies (Chang, 2003; Miyabe and Takeuchi, 2003; Rodríguez-Marín *et al*., 2003; Brandão *et al*., 2004; Ortiz and Arocha, 2004) have employed GLMMs, treating some of the model parameters as random effects. This utilization of random effects is particularly valuable in addressing interactions between variables, such as year and other categorical factors like area. The incorporation of random effects allows for a more meaningful representation of the underlying complexities in the data, contributing to a more comprehensive and accurate modelling approach. By considering random effects, GLMMs can provide more accurate estimates of catch rates, especially when dealing with hierarchical data structures.

Spatial models

Spatial models are analytical tools used in various fields to represent and analyze the spatial relationships and patterns of phenomena across geographic space and can help to understand, simulate, and predict the behaviour of processes that exhibit spatial dependencies. Spatial Autoregressive Models (SAR) and Spatial Regression Models are two popular models which are useful when analyzing data from different geographic locations and can account for spatial dependence in the standardization process.

Machine learning approaches

Machine learning algorithms, such as Random Forest, Gradient Boosting, and Support Vector Machines (SVM), can be applied to standardize catch and effort data. In a study conducted by Yang *et al.* (2020), SVM was applied to standardize longline catch per unit fishing effort for Bigeye tuna (*Thunnus obesus*) in the tropical fishing area of the Atlantic Ocean. The researchers evaluated three parameter optimization methods: a Grid Search method, and two enhanced hybrid algorithms, namely SVM in combination with particle swarm optimization (PSO-SVM) and genetic algorithms (GA-SVM). These optimization methods were employed to strengthen the performance of SVM, providing a more robust and accurate tool for CPUE standardization in fisheries data.

Like GAMs, neural networks offer increased flexibility in representing relationships between CPUE and explanatory variables. Maunder and Hinton (2006) pioneered a neural network approach for estimating relative abundance based on CPUE data. Their key innovation involved incorporating the year effect as a categorical variable within the neural network framework. Unlike GLMs , which are constrained to linear relationships (with the option of higher-order and interaction terms), neural networks enable the data to determine these relationships, allowing for more nuanced, non-linear modelling. Warner and Misra (1996) provide a comprehensive introduction to the connection between neural networks and regression, elucidating the terminology used in both. However, a drawback of neural networks is the potential existence of multiple solutions arising from common estimation techniques. These diverse solutions stem from different initial weights. Preliminary investigations indicate that these varied solutions yield comparable estimates of the year effect (Maunder and Hinton, 2006).

Environmental data integration

Integrating environmental variables (such as sea temperature, chlorophyll concentration, and ocean currents) into standardization models helps account for environmental influences on fish behaviour and distribution. Hinton and Nakano (1996) introduced a comprehensive habitat-based standardization (HBS) method that establishes an analytical framework, and consequently a statistical framework, for integrating an understanding of the distributions of environmental factors, fishing gear, and species into the standardization of CPUE. The fundamental concept is that if a hook is deployed in an environment preferred by a species, say bigeye tuna, it has an elevated probability of capturing that species. This becomes particularly crucial, for instance, in standardizing the effort of longline gear targeting tuna, given that the depth of the gear has increased over time as fishermen pursued bigeye tuna, which are generally located at greater depths in the water column.

Methods for multispecies multigear fishery

In tropical region, the marine fishery is of complex multi-species nature where in different species are caught by several fishing gears and each gear harvests several species making it difficult to obtain the fishing effort corresponding to each fish species. Since the capacity of the gears vary and each gear may harvest multiple species, the effort made to catch a resource cannot be considered as the sum of duration/units of operation of all the gears, making the nominal figures less relevant or rather intriguing.

As standardising efforts or CPUE stem from the kind of nominal measures of quantification available as basic data, a clear picture of the methodology adopted for landings/catch and effort collection is a mandatory requirement. Hence, as a typical point under focus towards an understanding of the marine fish landings data collection system followed in the Indian scenario is presented in brief.

India has a well-established data collection and estimation system for generating information on species-wise and fishing gear-wise marine fishery resources landings and fishing effort for different maritime states every month using skilled observers in fish landing ports. The method was developed by ICAR-Central Marine Fisheries Research Institute jointly with ICAR-Indian Agricultural Statistics Research Institute following a scientific sampling scheme named "Stratified Multistage Random Sampling Design (SMRSD)" (Sukhatme *et al*., 1958; Srinath *et al*., 2005), where stratification is done over space and time as well as sub regional/zonal levels. This system of data collection and estimation has been in use since 1960. The sampling frame was created by gathering information on marine fishing villages, landing centres, crafts, and gears, among other things, and it is updated on a regular basis to reflect changes in the sector through all India frame surveys. Species-wise catch, fishing effort, details of fishing crafts and gears and other related information are collected through this sampling scheme.

The population that is being attempted to be assessed through the samples is two-dimensional with zone-month as the parametric index. The zones are sub-civic spatially contiguous divisions that may be equated to districts within the administrative provinces, states, in India. The parameters like total catch, effort and catch rates pertaining to these zone-month populations are estimated through a two-stage sampling procedure, with the first one having strata and a pseudo-strata of time intervals within a month. The sampling units are accordingly the fishing vessel or unit selected at the second level after the selection of a landing centre/ fishing harbour on a particular day (lcd) of the zone-month.

In spatial stratification, based on the fishing intensity, geographical boundaries and number of landing centres, each maritime state is divided into suitable non-overlapping regions called fishing zones. These zones have been further stratified into substrata, depending on the intensity of fishing. The number of centres may vary from zone to zone (Fig. 1).

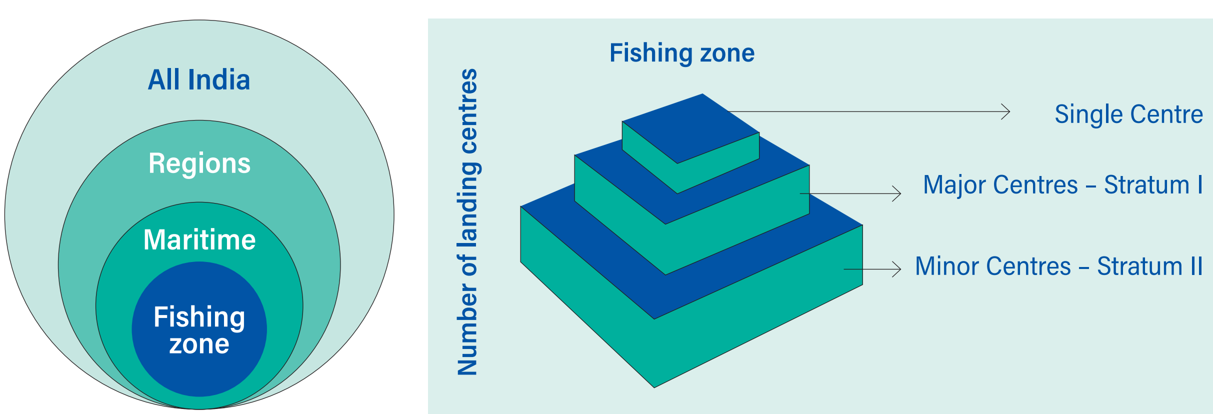

Fig. 1. Spatial stratification

Time strata	Days in a month									
1	1	2	3	4	5	6	7	8	9	10
2	11	12	13	14	15	16	17	18	19	20
3	21	22	23	24	25	26	27	28	29	30

Fig. 2. Temporal stratification

The landing centres are classified into High-Intensity Landing Centres (number of vessels in operation 300 or more), Major Landings Centres (number of vessels in operation between 100-299) and Minor Landing Centres (number of vessels in operation less than 100). The sampling coverage is more for High-Intensity Landing Centres than that for Major Landings Centres and it is still less for Minor Landing Centres. Among the fish landing centres, the major fisheries harbours/centres are classified as single-centre zones for which there is exclusive and extensive coverage.

The temporal stratification (Fig. 2) is more conventional than statistical, wherein the landing centre days to represent the population are ensured to spread evenly throughout the month, which is a major component defining the population. This gives enough support to take into account all the periodic oscillations noticed in resource availability within a month.

Suppose there are 10 landing centres in a zone, there will be 300 landing centre days (10 centres x 30 days) in a month. A month is divided into three groups, each with ten days. A day is selected at random from the first five days of a month, and the next five consecutive days are chosen automatically and form cluster groups of two consecutive days. In the remaining ten-day groups, the clusters are systematically selected with an interval of ten days. Normally, in a month, there will be nine clusters of two days each. Among the total number of landing centres in the given zone, nine centres are selected with replacement and allotted to the nine cluster days described earlier. Thus, nine landing centre days are observed in a month. The observations are made as per Table 1.

Table 1. Data collection during a landing centre day

24 hrs landings (One landing centre day)	Data collection method
1200 hrs to 1800 hrs of 1st day	By observation on the first day
0600 hrs to 1200 hrs of 2nd day	By observation on the second day
1800 hrs of the 1st day to 0600 hrs of the 2nd day (night landing)	By enquiry on the second day

During an observation period, when the number of boats/craft landings is high, it may not be practically possible to record the catches of all boats landed. Hence, the following procedure given in Table 2 is adopted (Alagaraja, 1984):

Table 2. Number of boats/crafts to be observed

Number of boats/crafts landed	Fraction to be observed
≤ 15	100 %
Between 16 and 19	First 10 and 50 % from the remaining
Between 20 and 29	1 in 2
Between 30 and 39	1 in 3 etc.

In the case of single centre zones, sixteen to eighteen days are selected randomly in a month and the units (fleets) landed on a selected day (either as a cluster of 2 days or a single day itself) is enumerated.

In the data collection system, dedicated technicians (harbour-based observers) with species identification skills visit the landing centres according to work schedules generated under SMRSD and record different aspects of the fishery from sampled boats.

With the introduction of computers and information technologies, the access and dissemination of information have become easier. ICAR-Central Marine Fisheries Research Institute took the lead in developing an online system for the collection and retrieval of data on marine fish landings and other related parameters named Fish Catch Survey and Analysis (FCSA) and the system has been operational since 2018 and was proven to be an excellent system for the data collection and estimation of marine fishery resources (Mini *et al.*, 2023).

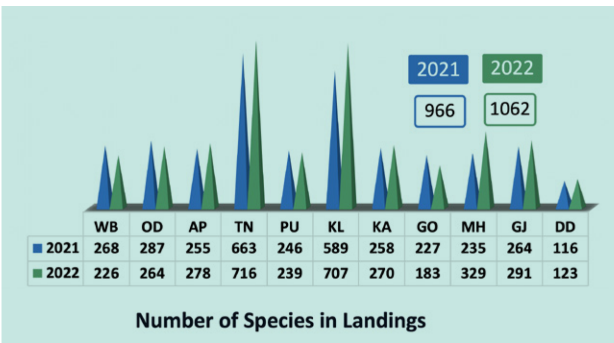

Fig. 3. Species distribution in landings along the coastal states

Based on observed landings and fishing efforts, an estimate of fish landings and fishing efforts for all fleets for a landing centre in a day is made. Monthly zonal landings are estimated using these data. Furthermore, estimates at the District, State, and National levels are obtained on a Monthly, Quarterly, and Yearly time scale.

The diversity of the fishery along the Indian coast is most probably reflected in the number of species documented in the fished taxa in recent years (FRAD-CMFRI, 2022; FRAEED-CMFRI, 2023), as illustrated in Fig. 3. Despite the source being commercial fishery, due to the fact that the record of landings was done by qualified and neutral enumerators with species level exhaustive identification as mandate, this can be viewed as measure of diversity of taxa. Additionally, the variety of gears in operation, as observed even on the southwest coast of India (Varghese *et al.*, 2021), serves as an indicator of the intricate nature of the fishery. The combination of high species diversity and the utilization of multiple types of fishing gears contribute to the complexity of fisheries in tropical countries like India.

As indicated earlier, due to varying capacity of gears and also the incidence of multiple species in some of the gears, a customized method is needed for the standardization of the fishing efforts. The following two approaches can probably handle the situation mentioned above:

a) A simple analytical framework

(Varghese *et al.*, 2020)

This method of standardization requires the species catch, total catch and total fishing effort. Let Y_{ijk} represents the catch of k^{th} species ($k = 1, 2, ..., s$) from i^{th} ($i = 1, 2, ..., g$) gear at the j^{th} ($j = 1, 2, ..., t$) time point (say year) and the corresponding effort is expressed as X_{ij}.

To calculate the component of standardized fishing effort for the species corresponding to each gear, the proportion of catch in the total catch by each gear for each year and a weighting factor for each gear is required. Following is the step-wise procedure of effort standardization:

Step1: Calculate $P_{ijk} = \frac{Y_{ijk}}{Y_{ij.}}$, where $Y_{ij.} = \sum_{k=1}^{s} Y_{ijk}$

Step 2: Obtain the mean and variance of P_{ijk} for each gear and for each species

$$\overline{P_{i.k}} = \frac{1}{t}\sum_{j=1}^{t} P_{ijk} \text{ and } \sigma^2_{i.k} = \frac{1}{t}\sum_{k=1}^{s} (P_{ijk} - \overline{P_{i.k}})^2$$

Step 3: Calculate the weighting factor as

$$W_{i.k} = \frac{\overline{P_{i.k}}}{(\sigma^2_{i.k}+1)} \text{ and } W'_{i.k} = \frac{W_{i.k}}{\sum_{i=1}^{g} W_{i.k}}$$

The weighting factor is then adjusted for unit sum. The decomposition of fishing effort for the species is then obtained by multiplying the corresponding total fishing effort for the gear in the year with the proportion of the species for the year corresponding to the same gear and the weighting factor.

Step 4: Obtain the standardized gear-wise fishing effort as

$$E_{ijk} = W'_{i.k} \times P_{ijk} \times X_{ij}$$

Here, the sum of all the gear efforts would give a total effort. But, the efficiency of gears varies so also the capability to catch in an hour which demands scaling the fishing efforts into a single scale. Hence, it is better to express all gears in terms of a single gear (which may be the least efficient or the most efficient) by deriving a suitable multiplication factor for each fishing gear.

Step 5: Calculate the catch per unit effort (gear-wise) as

$$CP_{ij} = \frac{Y_{ij.}}{X_{ij}} \text{ and } \overline{CP}_{i.} = \sum_{j=1}^{t} \frac{CP_{ij}}{t}$$

The multiplication factor is $\overline{CP}_{i.}'' = \frac{\overline{CP}_{i.}}{\overline{CP}'_{i.}}$ where $\overline{CP}'_{i.}$ is the least efficient or the most efficient gear

Step 6: Obtain the standardized fishing effort for k^{th} species at j^{th} time point as

$$\sum_{i=1}^{g} E_{ijk} \times \overline{CP}_{i''.}$$

For ease of computation, an R package named **F**ishing **E**ffort **Sta**ndardization (**FESta**) (Available at https://CRAN.R-project.org/package=FESta) for standardizing the fishing effort was developed. This package provides a function named "StdEffort" for the standardization of fishing effort expended by various fishing gears to obtain the Catch Per Unit Effort (CPUE) for a particular fish species using the time series of the total catch (landings) by each fishing gear, catch (landings) of a particular species (for which the CPUE is required) by each gear and total effort expended by each gear.

To install the FESta package in R, use the following code below:
install.packages("FESta")

And the usage of the "StdEffort" is:
StdEffort(sp_catch, tot_catch, effort, meg)

Where,

- sp_catch = Time series of catch/landings of a particular species (for which the CPUE is required) by each gear
- tot_catch = Time series of total catch/landings by each fishing gear
- effort = Time series of total effort expended by each gear
- meg = Most efficient gear (it takes value either FALSE (for least efficient gear) or TRUE (for most efficient gear))

An example of fishing effort standardization is given below:

A list named "Example" has been taken for illustration. It contains three data frames named sp_catch (Quantity of the fish species, in tonnes), tot_catch (Quantity of total catch, in tonnes), and effort (Fishing duration, in hours) with the same dimension.

To standardize the fishing efforts expended by various gears, the following codes can be used:

```
library(FESta)
data("Example")
StdEffort(sp_catch=Example$sp_catc,tot_catch
=Example$tot_catch, effort=Example$effort, meg=FALSE)
```

Remark: It is to be mentioned here that, as the method revolves around an identified gear, be it most efficient or otherwise, as reference amongst the gears with some significant contribution to the species landings may be selected for standardization and the gears with very negligible amount of species catch may be computationally insignificant. This always makes it mandatory to select the candidate gears by a strict rigour of pre-processing before reaching the standardization stage.

(b) Biodynamics model-based framework

(Sathianandan *et al.*, 2021)

The basic surplus production model takes the following expressions, one for the calculation of biomass of a species for successive periods termed as the process equation (Eq. 3) and the other relating biomass to catch and fishing effort known as the observation equation (Eq. 4).

$$B_{t+1}=B_t+r\ B_t\left(1-\frac{B_t}{K}\right)-C_t \qquad \ldots\ldots 3$$

$$C_t=q\ B_t\ f_t \qquad \ldots\ldots 4$$

In the multispecies and multigear fishery situation prevailing in tropical regions, a species is usually caught by multiple fishing gears (fishing fleets) and similarly, a fishing gear catches many species. Here, the fishing effort expended by a specific fishing gear results in the catching of many fish species and attributing the total fishing effort expended by the fishing gear to individual species-level effort is a challenging task. This issue is addressed here by incorporating an additional set of gear standardization parameters (λ's with its values summing to unity) in the catch equation in addition to the proportion of catch of the species in the total catch by the gear (Sathianandan *et al.*, 2021). Thus, for each species the expression for standardized fishing effort f_t was derived considering the fishing effort of all the g fishing gears in which the species is caught (Eq. 5). By replacing f_t in equation 4 we get the modified catch equation suitable for the multigear situation (Eq. 6).

$$f_t=\sum_{i=1}^{g} \lambda_i\ P_{i,t}\ f_{i,t} \qquad \ldots\ldots 5$$

$$C_t=\sum_{i=1}^{g} \left(\lambda_i\ P_{i,t}\ f_{i,t}\right)\ q\ B_t \qquad \ldots\ldots 6$$

The symbols used for the above models are described below.

B_t	biomass of the stock corresponding to year *t*
C_t	quantity harvested in year *t*
$f_{i,t}$	fishing effort in hours spend by fleet type *i* in year *t*
$p_{i,t}$	observed proportion of the species/resource in the catch by gear type *i* in year *t*
r	the intrinsic annual growth rate in biomass of the species/ resource
q	overall catchability coefficient in catching the species/resource
K	carrying capacity for the species/resource
λ_i	gear standardization parameter introduced for gear type *i*

The model parameters can be estimated after incorporating the observation error term ϵ_t in the catch equation (Eq. 7). The error terms ϵ_t were assumed to be distributed identically and independently as N(0, σ^2) leading to the expression for the negative log-likelihood (excluding constants) given as equation 8, which was minimised for estimating all the model parameters with $\sum_{i=1}^{g} \lambda_i = 1$ as an additional constraint for λ during minimization.

$$C_t = \sum_{i=1}^{g} \left(\lambda_i P_{i,t} f_{i,t}\right) q B_t e^{\epsilon_t} \quad \ldots\ldots 7$$

$$ln(L) = \frac{n}{2} ln(\sigma^2) + \frac{\sum_{t=1}^{n} \left(ln(C_t) - ln\left(\sum_{i=1}^{g} \left(\lambda_i P_{i,t} f_{i,t}\right) q B_t \right)\right)^2}{2\sigma^2} \quad \ldots\ldots 8$$

The model fitness can be assessed using appropriate statistical measures of goodness of fit or by verifying the closeness of the observed landings time series and its model-predicted values.

Conclusions

The standardization of commercial catch and effort data holds significance in fisheries, especially in cases where standardized abundance indices, derived from fishery-dependent information, play a crucial role in stock assessments. The primary objective of standardization is to minimize bias resulting from the intertwining of apparent abundance patterns with fishing power. Fisheries, particularly those where the fleet has undergone changes in fishing technology over time, face a heightened risk of confounding between fishing power and abundance. In tropical marine fisheries, due to varying gear capacities and the potential presence of multiple species in each gear, considering the effort exerted to catch a resource as the simple sum of the duration of fishing operation or units of operation of all gears is not feasible. An attempt has been made in the paper to highlight various effort standardization methodologies found in the literature for different situations, providing insights into the challenges faced in tropical fisheries. It is important to note that the specific methods and models used for standardization can vary based on the fishery, available data, and research/management objectives. Fisheries scientists and managers need to collaborate to determine the most appropriate standardization techniques for a particular study or assessment.

The way forward

In the past few years, there has been an increasing inclination towards the integration of diverse approaches and the amalgamation of various modelling techniques to enhance the precision and dependability of standardized catch and effort data. The selection of a particular method frequently hinges on the distinct characteristics of the data, the research/ management goals, and the computational resources at hand.

Fishing effort inherently also includes the fishing behaviour of fishers, be it in terms of which fish to target based on demand (indirectly fishing ground selection), scouting for fish (part of actual fishing hours) or selection of what fish catch should be retained (which eventually translates into landings). A Bayesian approach to standardizing fishing efforts could be an option to address these uncertainties in fishing effort brought about by fishing behaviour and variability in the standardization process. As the standardization of fishing efforts aims to account for factors such as changes in fishing practices, gear efficiency, and other variables that may affect the observed catch rates, the Bayesian methods provide a flexible framework for modelling these uncertainties and incorporating prior knowledge into the analysis.

Exploring innovative methods to standardize CPUE in anticipation of changing management requirements and fishermen's responses must be carried out. These elements together constitute a holistic strategy to propel advancements in research within this domain, with the overarching goal of refining the precision and applicability of CPUE standardization techniques. Simultaneously, the strategy needs to be designed to ensure flexibility and adaptability to accommodate shifting management needs and the dynamic nature of fishing practices. Assessing the general applicability of currently available CPUE standardization methods in tropical fisheries and identifying the conditions under which they outperform other methods is also important.

Integration of advanced technologies, such as satellite imagery, artificial intelligence, and machine learning may be handy in improving the accuracy and efficiency of measuring fishing effort. This could enhance data collection and provide real-time information for more dynamic standardization models. Incorporation of spatial and temporal dynamics into standardization models to understand how fishing effort varies across different locations and seasons is crucial for accurate stock assessments and sustainable management.

Integrating environmental factors into the standardization of fishing efforts can help to account for influence of environmental variables such as temperature, ocean currents, and habitat characteristics etc. on fish behaviour and distribution thereby fisheries management can better understand and respond to the dynamic nature of marine ecosystems. This strategy extends beyond merely focusing on the target species and considers the broader ecological context.

Refining the standardization models/methods that account for the impact of fishing effort on multiple species simultaneously is necessary especially in mixed-species fisheries where the catch of one species may affect the abundance of others. Future research efforts in the standardization of fishing effort should aim to enhance the robustness, accuracy, and adaptability of methodologies to contribute to the sustainable management of tropical fisheries resources.

Standardization of fishing effort is only a part of the entire assessment process which includes a number of steps starting from fish catch data collection, its collation, correction, analysis, inferences and use of outputs. While researchers attempt to fine-tune each step of the process, efforts also should be made to adhere to a unified, comprehensive and locally attuned process of assessment genuinely reflective of the targeted fishery scenario. For this to happen effectively, a seamless collaboration between fishers, department officials, trade organizations, fisheries researchers, fishery managers and policy makers is needed.

Acknowledgement

This work forms a part of the research project *Geo-referenced online information system for marine fisheries on GIS platform to formulate management strategies for sustainable harvest of resources (FRA/GIS/01)*. The authors would like to acknowledge the support and guidance provided by the Director, ICAR-Central Marine Fisheries Research Institute, Kochi.

References

Alagaraja. 1984. *Indian Journal of Fisheries*, 31 (2): 177-208.

Bishop *et al.*, 2006. *Rev. Fish. Biol. Fisheries,* 16: 21–38.

Beverton *et al.*, 1957. *On the Dynamics of Exploited Fish Populations.*

Brandão *et al.*, 2004. *Fisheries Research*, 70(2-3): 339-349.

Chang. 2003. *ICCAT Col. Vol. Sci. Pap*, *55*(2): 453-466.

Daniel *et al.*, 2016, *Canadian Journal of Fisheries and Aquatic Science*, 74(1): 8-14.

FRAD-CMFRI. 2022. *CMFRI Booklet Series No.* 26/2022.

FRAEED-CMFRI. 2023. *CMFRI Booklet Series No.* 31/2023.

Gavaris. 1980. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(12): 2272-2275.

Hastie *et al.*, 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.

Hilborn and Walters. 1992. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*, 587 p.

Hinton and Nakano. 1996. *Bull. Inter-Amer. Trop. Tuna Comm.* 21(4): 169-200.

Lambert. 1992. *Technometrics*, 34(1): 1-14.

Lo *et al.*, 1992. *Canadian Journal of Fisheries and Aquatic Sciences*, 49(12): 2515-2526.

Maunder and Punt. 2004. *Fisheries Research*, 70: 141–159

Maunder and Hinton. 2006. *Inter-American Tropical Tuna Commission special report, Article No. 15*, 15p.

Mini *et al.*, 2023. *Fisheries Research*, 267 (2023): 106821.

Miyabe and Takeuchi. 2003. *ICCAT Col. Vol. Sci. Pap.*, 55: 1190-1207.

Nelder and Wedderburn. 1972. *J. R. Statist. Soc. Ai.*, 137: 370–384.

Ortiz and Arocha. 2004. *Fisheries Research*, 70(2-3): 275-297.

Pennington. 1996. *Fishery Bull.*, 94: 498-505.

Pennington. 1983. *Biometrics*, 281-286.

Pinheiro and Bates. 2000. *Mixed-effects models in S and S-PLUS.*

Robson. 1966. *Res. Bull. ICNAF*, (3): 5-14.

Rochman *et al.*, 2017. *Indonesian Fisheries Research Journal*, 23(1): 29-38.

Rodriguez-Marin *et al.*, 2003. *ICES Journal of Marine Science*, 60(6): 1216-1231.

Sathianandan *et al.*, 2021. *ICES Journal of Marine Science*, 78(5): 1744–1757.

Setyadji *et al.*, 2018. *Turish. Journal of Fisheries & Aquatic Science*, 19(2): 119-129

Sparre. 1998. *FAO Fish. Tech. Paper.*, 306, 1-407.

Srinath *et al.*, 2005. CMFRI Special Publication No. 86, 57p.

Stamatopoulos and Abdallah. 2015. *Journal of Marine Science: Research and Development,* 5: 170p. doi:10.4172/2155-9910.1000170

Sukhatme *et al.*, 1958. *Biometrics*, 14: 78–96.

Varghese *et al.*, 2020. *Journal of the Indian Society of Agricultural Statistics*, 74(1): 33-40.

Varghese *et al.*, 2021. *Mar. Fish. Infor. Serv., T & E Ser.,* 250: 7-17.

Warner and Misra. 1996. *The American Statistician*, 50(4): 284-293.

Yang *et al.*, 2020. *Ocean and Coastal Research*, *68*.