

IOTC WORKING PARTY ON METHODS (WPM16)

A Stratified Baseline-Anchored Simulator for Precision and Reliability Metric for Longline Fisheries Observer Programs' Coverage Design and Validation by Integrating the Coefficient of Variation: A Case Study of Kenya.

Kelvin Wachira¹, Stephen Ndegwa¹, Elizabeth Mueni¹, Alex Lukwendah¹, Benedict Kiilu¹, Zachary Ogari¹, Donald Obiero¹, Collins Ndoro¹

Kenya Fisheries Service, State Department for Blue Economy and Fisheries

Abstract

Effective fisheries management relies on accurate, unbiased, and spatially resolved information on catch, effort, and bycatch composition. This study integrates longline observer datasets collected along the Kenyan coastline with advanced analytical frameworks to standardize Catch Per Unit Effort (CPUE), quantify uncertainty through the Coefficient of Variation (CV), and evaluate sampling biases that influence bycatch estimates. The study employed modern statistical, spatial, and computational tools to improve the precision, transparency, and interpretability of observer-based indices of abundance and mortality.

At the core of the methodology, set-level and sample-level data were harmonized and validated using the R ecosystem (tidyverse, sf, lubridate, ggplot2). Each observed set was assigned a $1^\circ \times 1^\circ$ grid cell based on midpoints between setting and hauling coordinates, forming the spatial foundation for analysis. Nominal CPUE was calculated as kilograms per 1,000 hooks and subsequently standardized through generalized linear models (GLMs) and delta-lognormal frameworks, adjusting for spatiotemporal and operational covariates. Variability in CPUE estimates was quantified using both analytical and bootstrap-derived CVs, providing a measure of precision across years, species, and strata.

Species composition and sampling uncertainty were addressed using nested resampling approaches that simultaneously captured variation in catch proportions and sampling effort. This dual-layer resampling allowed realistic propagation of uncertainty, improving the reliability of standardized indices. Data validation modules within the workflow flagged statistical and protocol deviations—such as outlier catches, under-sampling, and inconsistent spatial coverage—ensuring data integrity before downstream analysis.

Spatial analyses were performed using the `sf` and `rnatualearth` packages to construct numbered coastal grids, overlaying observer data on Kenya's Exclusive Economic Zone (EEZ). Grid-level summaries were developed to show (1) numbered $1^\circ \times 1^\circ$ grids along the coastline; (2) species distribution and relative abundance per grid; and (3) total allocated catches by grid, visualized as heatmaps. These maps highlighted species-specific spatial hotspots and potential bycatch concentration zones, providing visual insight into effort distribution and catch dynamics.

In connection with broader fisheries management objectives, the study aligns strongly with contemporary debates on observer coverage, bycatch bias, and sampling precision. Bycatch remains a significant source of unaccounted mortality for commercial, recreational, and ecologically sensitive species. Although observers provide the most reliable data on at-sea mortality, the assumption of random sampling within observer programs is often violated. Opportunistic or voluntary participation by vessels introduces non-random sampling bias, compromising the accuracy of bycatch and CPUE estimates. The present framework incorporates diagnostic modules to assess representativeness and to compare observed versus unobserved vessel behavior—critical steps for bias correction in longline fisheries data.

Simulation and literature evidence suggested that achieving at least 20% observer coverage yields acceptable precision for common species, whereas rare or infrequently encountered species require coverage levels of 50% or more to reduce uncertainty to acceptable levels. In the Kenyan longline fishery context, spatial heterogeneity in effort and species composition implies that coverage requirements may vary substantially across grid cells and seasons (Southeast Monsoon vs. Northeast Monsoon). The integrated CV-based approach provides a quantitative benchmark for designing efficient sampling strategies that balance logistical constraints and statistical rigor.

Ultimately, this study demonstrated a reproducible, transparent, and scalable approach to fisheries observer data analysis. It bridges operational sampling design with modern computational analytics, ensuring that observer-derived indices of CPUE and bycatch are both statistically robust and ecologically meaningful. The combined use of standardization models, spatial stratification, and resampling-based uncertainty quantification enhances the credibility of bycatch and effort estimates in data-limited regions such as the Western Indian Ocean. The resulting methodology not only supports precision-based observer allocation but also strengthens the foundation for sustainable fisheries management, adaptive monitoring, and regional stock assessment frameworks.

Keywords: Observer Validation, Coefficient of variation, Catch per Unit Effort, Spatial stratification, Bycatch estimates, Monte Carlo Resampling, Uncertainty, Fisheries Monitoring

Introduction

The management of longline fishery is complex due to an ever-present challenge of bycatch that are caught in substantial quantities, which involve non-target species, sharks, rays, seabirds, turtles and cetaceans. Therefore, it becomes paramount to strengthen monitoring of the fishery in order to account for these incidental catches and safeguard progress to sustainable management of the fisheries resources as well as fulfill conservation mandates (Kiilu et al., 2025).

Kenya has a functional national observer program that deploys fisheries staff as scientific observers onboard national and foreign flagged vessels in the Kenyan Exclusive Economic Zone (EEZ) and the high seas (Fondo & Omukoto, 2021). The implementation of the observer program is critical function by the Kenya Fisheries Service as stipulated in the Fisheries Act Cap 378 and forms part of the backbone in the monitoring, control and surveillance (MCS) functions (Kiilu et al., 2025). However, as evidenced in literature and national observer reports, the program suffers constraints ranging from finances, logistical operations, knowledge and capacity among others that often contribute to gaps in data coverage (Babcock et al., 2003). The strive for balance by fisheries observer programs between capped observer efforts and the necessity for precise estimates of target, bycatch and protected, endangered and threatened (PET) species for the longline fisheries has become unbearably inevitable (Gray & Kennelly, 2018). The coefficient of variation (CV) is a global technique that has been applied as a scale-free metric of relative precision to set monitoring targets and validation of core indices and estimates relevant to management of longline fisheries (Thompson, 2012).

The observer program in Kenya implements training techniques from the IOTC and the SWIOFC that has structured data collection forms and reporting templates for catch and effort, biological sampling, and compliance information. This study focused on catch and effort data in order to review and assess precision estimates at the primary sampling unit, which in this case is the basket (hooks between floats/radio buoys). In order to comprehensively assess precision, it was inevitable to compile all data and simulate sampling techniques and finally review the statistics involved.

Therefore, study aimed to assess the efficiency of alternative stratification schemes unique to longline fishery in addition to estimates of CV for a given defined stratum, and

thus, demonstrates proof to select the most statistically efficient strata for fisheries managers need to understand the marginal gains in precision for every new stratification scheme.

Objectives

1. To Build per-stratum baseline standardised CPUE templates from vessel logbook and observer data.
2. Simulate observer sampling under alternative set-coverage and basket-sampling schemes
3. Compute per-stratum CV of CPUE (catch rate estimates) and 95% confidence bounds
4. Identify minimal coverage and within-set sampling fraction that meets management targets. This identifies which species/strata estimates are statistically reliable and which suffer from insufficient sampling effort due to high variance or low coverage.
5. To validate that the current operational stratification scheme is statistically appropriate for the longline fishery by verifying that the strata successfully partition the fishing effort into relatively homogenous units, whereby catch rates within a stratum are less variable than across the entire fishery.

Justification and Rationale

Therefore, aggregating catch totals of vessel logbooks with basket-level fishing behaviour from observer data has potential to generate an authentic baseline population that conserves both reported catch totals and within-set heterogeneity (Kesavan Nair & Alagaraja, 1988). In addition, the simulation of sampling techniques for observer candidate set and basket across different strata allows for proof-based design of coverage levels for sets and basket sampling efforts. In essence, this procedure is critical because it manages the phenomenon of potential uncertainty intrinsic in proportionate sampling of increasingly unstable natural patterns like quantifying the variation in sampling (Thompson, 2012). The strive to attain 100% coverage over the whole fishery for observer program steadily continues and with that reality, estimated total catches for target, bycatch and PET species becomes a random estimate, which is conditional to inaccuracies (Babcock et al., 2003). For example, reported observer data particularly for elusive catches of PET species, usually demonstrate highly overdispersed distribution whereby, a majority of fishing sets tend to report zero bycatch or PET species while only a few sets record extremely high levels of these species (Curtis & Carretta, 2020). It is due to this high uncertainty in the fishery that causes conventional simple sampling techniques unreliable (Little & Rubin, 2019).

Stratified random sampling schemes based on factors that drive catches such as fishing area, target species, depth or seasons is more efficient than simple random sampling (SRS), which fails to validate whether sampled vessel effort is truly representative of the unobserved effort by rigorously assessing the statistical efficiency of various stratification and allocation strategies (Babcock et al., 2003). This study aimed to correct that by comparing both logbook data and observer data for species category totals and imputed by mean for unobserved sets and baskets per stratum. Then acknowledged Newton and Geyer's (1994) technique in a simulation through nested bootstrap and Monte Carlo (MC) propagation (Manly, 2018) as well as deployment of the Horvitz-Thompson (HT) estimator (Gokpinar & Arzu Ozdemir, 2012) to apply the CV as the definitive metric for quantifying the uncertainty property of estimated total catches of target, bycatch and PET species categories (Wakefield et al., 2018; Hulliger, 1995). Therefore, the techniques were then invested to demonstrate a probabilistic evaluation of the precision of estimated catches by confirming that a low CV denotes a statistically stable estimate whereas a high CV would then signal that the monitoring program is statistically deficient for that particular species category or stratum (McInerny, 2014; Babcock et al., 2003).

Furthermore, the simulation permits for the testing of different stratification and allocation schemes such as placing more observers in areas known for high bycatch (Babcock et al., 2003). Also, the technique creates opportunities for advanced model-based and inclusion of the finite population correction (FPC) (Thompson, 2012) the current simulation abstracts away from complex design-based and FPC.

Materials and Methods

The positions for start setting and end hauling were obtained from the logbook dataset, which contained the catches for both reported and species categories. The line, theoretically set to haul, denotes the location of the longline fishing gear and consequently, establishes a spatial distribution of the catches (Francis, 1984). Thereafter, we made a LINESTRING for every set accurately with dateline for one geometry per set (Pebesma & Bivand, 2023) in order to allow distribution of every set that is positioned in each 5° cell. Then a polygon of 5° x 5° grid, which was cropped into the Western Indian Ocean FAO Area 51 in order to create a stratification of grid cells required for area of management (FAO, 2002).

Consequently, each LINESTRING was intersected with the FAO cropped grid to generate a line segment of per combination of set and cell and computed the length of every set segment within individual cells (Jolly & Hampton, 1990). The set segments were split into geodetic length fractions as follows:

Construction of the line segment from setting to hauling: L_i . Intersect L_i with the 5° grid, then the fraction of line length inside each cell was used to supply set catches to those cells:

$$wt_{ic} = \text{total}_{wt_i} \chi \frac{\text{length}(L_i \cap \text{cell}_c)}{\text{length}(L_i)}, \text{ (Pebesma \& Bivand, 2023; Francis, 1984)}$$

This process was possible because of the absence of hooks at the basket-level (discussed above) and thus, hook locations along individual cells was not possible hence, use of catches for baskets so as to preserve supply of catch and effort (adjusted CPUE) by length fraction in limited or missing positions of baskets. Finally, reported catches of species and species categories were distributed along each intersected cell by fraction and combined for every grid cell. The output was a logbook table eventually had cell-level data that was used for the area stratification as well as generate templates for the 'true' population (Pebesma, 2018).

The computation of the sum of species type in a basket derived the basket-weight-total and also, per set calculation involved summation of total weight of species composition per observed basket and therefore, sampled total weights of species composition per basket was validated to be less than or equal to the observed total weight per set (Francis, 1984). This outcome confirmed that there were missing basket-level data particularly from unobserved baskets per set.

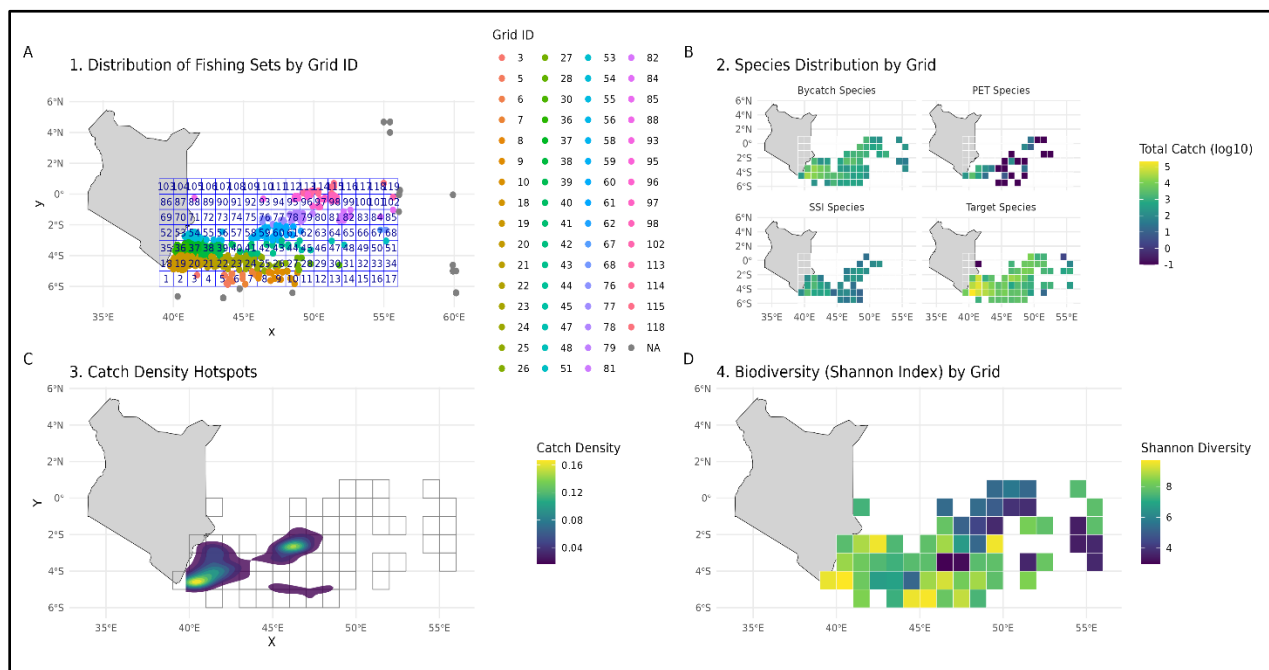


Figure 1. Comprehensive Overview of Catch Composition Per 1°x1° Grid Cell from Kenya's Observer Catch and Effort Data 2018-2015 (Developed in R base 4.4.3)

Data sourcing, Incorporation and Validation

The fisheries observer dataset involves sampled data of catch and effort data $n = 6928$ from 2018 to 2025, which represents the historical dataset of fisheries observer program onboard longline fishing vessels. Similarly, vessel logbook data $n = \text{total sets}$ that reflects fishing sets of the observer data was used to provide total catches and total number of hooks for every set for validation of with those exact observed total of sets captured in the observer data. Therefore, we combined the two datasets to generate a schema of the following fields per row: date, start-setting positions, end-hauling positions, season (NEM and SEM), set unique identification (id), basket index, species type, species category (target, bycatch, PET), total weight per species composition (kg), observed total weight per set (calculated catch) and observed total number of hooks per set. Generally, data cleaning was performed in R (R Core Team, 2024) with packages tidyverse and dplyr (Wickham, 2014; Wickham et al., 2023) and vectorisation of spatial coordinates (Pebesma, 2018).

Proportion of Weight Raise and Imputation for unobserved Baskets

Here the focus was on observed basket species weights in order to compute species proportions and raise to the total set weight. At the same time, impute (mean) for unobserved baskets to generate complete basket templates for a 'true' population (Quinn & Deriso, 1999). Thus, due to the limitation of absence of number of observed hooks per basket, sampled basket weights proportions within set were used to allocate total weight of sets proportionally to observed baskets as well as species categories whereby, the approach becomes: for set i let sampled basket species weights be w_{ibs} for sampled baskets $b \in S_i$ (Harley et al., 2001): sampled weight of species composition is calculated as $w_{is}^{sample} = \sum_{b \in S_i} w_{ibs}$ and total sampled weight of baskets per set i (including species composition in baskets) as $W_i^{sample} = \sum_{b \in S_i} \sum_s w_{ibs}$ (Quinn & Deriso, 1999). In addition, the proportion of sampled weight (per basket) for species was calculated as $s: \bar{P}_{ibs} = \frac{w_{ibs}^{sample}}{W_i^{sample}}$, and showed the ratio of how much of sampled catch each basket contributed. Then, estimates were done for the total number of baskets in a set, whereby, B_i refers to total number of baskets hauled per set.

Thereafter, we estimated the total number of baskets per set by conducting a mean imputation to generate weights for unobserved baskets and was achieved by inferring reasonable number of baskets from observed total weights per set or logbook set totals (kgs) $W_i^{(set)}$ as $B_i = \frac{W_i^{(set)}}{\bar{w}_i^{(sample)}}$, whereby, $\bar{w}_i^{(sample)}$ becomes the mean weight of sampled basket (Kimura, 1981).

Furthermore, we normalised proportions and expanded to total weights per set using, $\sum_{b=1}^{B_i} \bar{P}_{ib}^{(all)} = 1$ in order to ensure that summation of proportions becomes 1 (Quinn & Deriso, 1999). Then, calculated the total weights for a set by scaling the basket weights as: $\bar{w}_{ib}^{all} = \bar{P}_{ib}^{all} * W_i^{set}$ this confirmed that both observed baskets and imputed baskets were combined to represent the expanded total catch for that set.

Similarly, the same computation was repeated for estimating species composition per basket since the observed baskets have total weights of species types and was imputed for basket-level species composition by, calculating the proportion of species type per observed basket as

$p_{is} = \sum_{b \in obs} w_{ibs} / W_i^{(sample)}$, then for imputed baskets utilise the same ratio:

$\bar{w}_{ibs}^{(imputed)} = p_{is} * \bar{w}_{ib}^{(all)}$ and consequently, the summation over all baskets will generate the expanded total weights of species composition per set (Thompson, 2012). According to previous literature, the imputation fortifies variance when coverage is less than 100% whereas, the ratio of weight expansion reduces bias because it safeguards proportional heterogeneity within sets (Thompson, 2012; Pennington, 2001; Cotter, 1998). The ratio estimation of the total catch per set $\bar{Y}_i$ under an unbiased proportional sampling then became:

$$\bar{Y}_i = W_i^{set} * \frac{\sum_{b \in sample} w_{ibs}}{\sum_{b \in sample} w_{ib}} \text{ (Cochran, 1977).}$$

CPUE standardisation at Set-Level

The computation for expanded total catch per set and records of total number of hooks per set was done to generate nominal CPUE (kg/1000 hooks) and then applied a GLM model to standardised nominal CPUE into indices of abundance that was comparable across stratum (Lo et al., 1992). Then, summed totals into population sets and population baskets for stratum templates (Thompson, 2012). Standardisation of CPUE satisfies its comparison across sets and thus, nominal CPUE at set-level for species s in set i as:

$CPUE_{is}^{nom} = \frac{\bar{Y}_{is}}{H_i^{obs}} * 1000$, where, $\bar{Y}_{is}$ indicates the raised total catch (kg) for species s in set i whereas H_i^{obs} , represents the total number of hooks observed per set. Thereafter, nominal CPUE was standardised using a gaussian GLM (Lo et al., 1992) to enhance comparability of CPUE across spatial and temporal strata and thus, ensured minimal bias due to non-uniform effort or sampling techniques (Hilborn & Walters, 1992). Consequently, GLM was computed as: let Y_{is} be total catch of species type s in set i and adopted a Gamma-log, which ensures that responses are positive and caters from multiplicative effects (Hilborn & Walters, 1992).

Therefore, the model was performed as:

$E[Y_i] = H_i^{obs} * \mu_i$, so that $CPUE_{s,i}^{nominal} \sim \text{Gamma}(\text{mean} = \mu_{s,i}, \text{link} = \log)$, and model was fitted with an offset $\log(H_i^{obs})$ (Lo et al., 1992). Then standardisation was done by strata of area, season and species type.

$$E[Y_{is}] = \mu_{is} = \exp(\beta_0 + \beta_1 \log(H_i^{obs}) + \beta_2 \text{Area}_i + \beta_3 \text{Season} + \beta_4 \text{Species}_s(\text{type}) + \epsilon_i)$$

The model parameters were defined as, β_0 to indicate the intercept inferring to the abundance index of baseline and $\log(H_i^{obs})$ performs the offset for hooks to achieve adjusted effort. The model involved effects at stratum level such as Area_i as categorical that impacts spatial stratum developed for 5° grid cell, Season_i as a categorical outcome, $\text{Species}_s(\text{type})$ to acknowledge basket-level species composition effects and ϵ_i to represent residual error (Hiborn & Walters, 1992). Thus, indices of CPUE standardization per stratum was generated as: $CPUE_{h,s}^{std} = \exp(\bar{\beta}_0 + \bar{\beta}_{2,h} + \bar{\beta}_{3,s} + \bar{\beta}_{4,s} + \dots)$, which eliminates effort and sampling biases (Hiborn & Walters, 1992).

Generate Baseline Templates for True Population

This stage was important to introduce the standardised CPUE ($CPUE_{h,s}^{std}$) to generate a hypothetical but natural population, which the sampling simulation procedure MC will be undertaken (Thompson, 2012; Manly, 2018). The stage had two phases: (i) template for population sets, since each set represented a feature in the population with its parameters. (ii) template population baskets whereby, every set is expanded into total number of baskets B_i to create B_i^{full} and thereafter, assign individual weights of species composition for every basket by disaggregating raised total weights of species across the baskets using B_i^{full} template (Cochran, 1977).

Phase 1: Construction of template for population sets for every observed set was performed so that the output generated a set with rows bearing parameters as follows: unique identification (id) of set i , stratum (area * season), H_i^{obs} , $\bar{Y}_i$, w_{is}^{sample} and p_{is} , Y_{is} and ($CPUE_{h,s}^{std}$) (already defined above). Therefore,

$$\text{Population}_{sets[i]} := (i, \text{stratum}, H_i^{obs}, \bar{Y}_i, \{\bar{P}_{is}\}_s, \{Y_{is}\}_s, CPUE_i^{std})$$

Phase 2: Construction of template for population baskets were created by amplifying individual set i into B_i baskets and then assigned the weight of species composition into those baskets was done for each set i : $\sum_{b=1}^{B_i} \sum_s w_{ibs}^{pop} = \bar{Y}_i$ (refer above) and likewise, for each species s in set i : $\sum_{b=1}^{B_i} w_{ibs}^{pop} = Y_{is}$ (refer above). In addition, the P_{ib} ratio output already derived for baskets in each set was considered to demonstrate observed

pattern at basket-level. Thereafter, we distributed species s across baskets through a deterministic disaggregation (Cochran, 1977): $w_{ibs}^{pop} = P_{ib} * Y_{is}$, $b = 1, \dots, B_i$.

Thus, the totals for each basket assigned became:

$$w_{ib,total}^{pop} = \sum_s w_{ibs}^{pop} = P_{ib} \sum_s Y_{is} = P_{ib} * \bar{Y}_i \text{ (Cochran, 1977).}$$

In order to confirm that the allocation was conservative, we computed $\sum_{b=1}^{B_i} w_{ib,total}^{pop}$ to confirm it was the same as $\sum_s P_{ib} * Y_i = Y_i$ (Thompson, 2012).

Eventually, the population basket template developed had features represented in every row as: (i, b, s) and weight (kgs) w_{ibs}^{pop} .

Preparation for Stratified Statistics and Inclusion of Species Composition Uncertainty into Sampling Scheme

This step was increasingly important to setup the required ingredients for a stratum-level configuration that comprised of inclusion probability that warrants the simulation of sampling schemes discussed above (stratified, PPS) and then evaluate Horvitz-Thompson values and approximate variance (Pebesma & Bivand, 2023). The parameters that were involved in the computation were derived from previous outputs as follows: stratum h , sets $i \in h$, species s and basket b . Accordingly, N_h represented the number of sets in stratum h and for set i it contained hooks H_i^{obs} , total raised catch $\bar{Y}_i$, total raised species catch $\bar{Y}_{is}$ and $CPUE_i^{std}$. In addition, total estimated number of hooks per stratum was calculated as: $H_h = \sum_{i \in h} H_i^{obs}$ (Hiborn & Walters, 1992).

As explained above, the technique adopted for PPS aimed to conduct deliberate oversampling of sets as opposed to SRS, so as to capture most of the total catch from those sets. The technique as argued by (Babcock et al., 2003), elevates the precision of calculated total catches since the uncertainty in population is attributed to a small number of sets that are have flourishing catches or increased CPUE (Kimura, 1981). Thus, the focus was the size measure $size_i$ in set i was derived as the i^{th} set, which was adopted as $\bar{Y}_i$. The probability ratio of selecting i^{th} set in one (m_h) draw with replacement was then computed as the $size_i$ divided by the summation of all other size measures $size_j$ and denoted as:

$$p_i^{(prop)} = \frac{size_i}{\sum_{j=1}^N size_j}, \text{ whereby } N \text{ represents the total number of sets in the population (Jolly \& Hampton, 1990).}$$

The subsequent step was the sampling itself with-replacement boosted probability π_i (). In essence, m_h units were sampled with replacement from the derived true population

and thus, on the one hand, probability of not selecting a particular set i for every m_h unique draw was denoted as: $(1 - p_i^{(prop)})^{m_h}$ (Hulliger, 1995). On the hand, the probability of that particular set i being picked at least once was denoted as:

$\pi_i^{(PPSWR)} = 1 - (1 - p_i^{(prop)})^{m_h}$ (Hulliger, 1995). Therefore, the features for this approximation were adopted such that if both $p_i^{(prop)}$ and m_h became small, the output of $m_h * p_i^{(prop)}$ was an acceptable approximation for $\pi_i^{(PPSWR)}$, which substantiated the assumption for PPSWR. However, in cases that $size_i$ prevailed with increased $p_i^{(prop)}$, it meant that huge sets influenced $\pi_i^{(PPSWR)}$ to quickly arrive at 1, which inferred that there was a higher chance that the sets would be chosen in the sample (Thompson, 2012). According to (cite), the computational function for $\pi_i^{(PPSWR)}$ is inherently placed on caveat of 0 and 1 (Cochran, 1977).

The HT estimator discussed above was deployed in the sampling process as an unbiased estimator for the overall population due to uncertainty supplied in the PPSWR (Pennington, 1996). Hence, HT estimator for species s was derived as: $\bar{T}_{HT}^{(s)} = \sum_{i \in S} \frac{\bar{Y}_{is}}{\pi_i^{(PPSWR)}}$, and $\bar{Y}_{is}$ as defined above, represented the set-level catch for species s in set i while the summation refers to sets S for oversampled sets (FAO, 2011). In retrospect, the unbiasedness of HT is usually tied to π_i as the authentic first-order inclusion probability derived from the true without-replacement sampling scheme (Pennington, 1996) while, its precise variance measure of the HT estimator relies upon second-order joint probabilities π_{ij} (Newton & Geyer, 1994; Hulliger, 1995). Thus, since $\pi_i^{(PPSWR)}$ used in this step was a derivative approximation from the authentic probability, the precision of the estimates is dependant on the properties of the approximation (Pennington, 1996). In that regard, the need to perform sampling process through bootstrap was paramount so that samples are resampled M times in consideration of the sampling regime and evaluated the empirical variance of HT (Hulliger, 1995). The process of bootstrap or Monte Carlo resampling ensured that both sampling scheme complexities i.e. for both PPS and stratification as well as species composition uncertainty are accommodated inherently in the sampling procedure of sets and baskets, and generates an authentic variance associated with those schemes (Thompson, 2012).

Species Composition and Scheme Variance Propagation Across Strata with Monte Carlo and HT Simulator

The purpose of this approach was to determine how observed coverage and basket sampling could be synthesized into precision for CPUE of species composition, as well as translate into design-based planning of effort and targets. The simulation handled set observations and within-set basket sampling across various sampling scheme and then calculated HT values of totals for every stratum and finally aggregate resampled estimates for composition variations. Thus, to be able achieve the desired outcomes, the procedure was structured to perform computations for every replicate as follows:

True Population per Stratum

In the above analysis, already generated population templates for individual strata was:

$$Population_{stratum,h}: \{CPUE_{is}^{std}, w_i, Area_i, Season_i, Species_s\}$$

whereby $CPUE_{is}^{std}$, is the standardised CPUE for species s in set i derived above and $w_{i,h}$ is the overall weight of a stratum obtained from total effort or total number of sets in population (Efron & Tibshirani, 1993).

Therefore, mean CPUE of the true population for every stratum was calculated as:

$$\mu_{true}^{CPUE} = \sum_{i=1}^N w_{i,h} * CPUE_{is}^{std} / \sum_{i=1}^N w_{i,h}$$

and consequently, the overall catch estimates of the true population of stratum h is derived from: $Y_{true,h} = \sum_{i=1}^N CPUE_{is}^{std} * H_h$.

Therefore, let N_h become number of sets per stratum, which is obtained from population sets computed. Thus, by selecting $m_h = round(coverage_{level} * N_h)$ (Cochran, 1977), the simulator withdraws m_h sets according defined sampling scheme discussed above. The sampling scheme computed were random sampling conserved to be equivalent of equal-probability sampling (Gokpinar & Arzu Ozdemir, 2012), stratified sampling was considered as similar to random although it was executed per individual stratum. Then, performed sampling with probability proportional to size with replacement and approximation for inclusion probability (Efron & Tibshirani, 1993) to generate expansion across sets with the estimator per species as $\bar{T}_{HT}^{(s)}$ (Hulliger, 1995).

Monte Carlo Propagation

Application of the Monte Carlo (MC) simulator (Manly, 2006) accounted for uncertainty in deriving species composition since proportions $\bar{P}_{is}$ had been calculated from a subset of baskets. The technique for nested bootstrap (Manly, 2018), was adopted for individual MC replicates and for every sampled set as follows:

- Observed baskets S_i were resampled (with replacement) to generate a bootstrap sample $S_i^{(r)}$ at the basket level and this process of resampling whole baskets was

vital to shield covariance of multi-species within a basket (Kunz et al., 2007). Thereby, we reworked proportions $\bar{P}_{is}^{(r)}$ from baskets that were resampled as:

$$\bar{P}_{is}^{(r)} = \frac{\sum_{b \in S_i^{(r)}} w_{ibs}}{\sum_{b \in S_i^{(r)}} \sum_s w_{ibs}}$$

- Similarly, we recalculated the raised total catches for within-set expansion, which was accomplished by inflating basket-weights by proportion $\bar{Y}_{is}^r = H_i^{obs} * \bar{P}_{is}^{(r)}$.

The nested resampling was undertaken within individual replicates of MC so that uncertainty in species composition structure was supplied to the variance of across-replicates (Hulliger, 1995; Hilborn & Walters, 1992). In addition, propagating uncertainty, baskets were resampled entirely (the basket-level vectors across species) when building population of species $\bar{P}_{is}$ inside the MC replicates in order to capture composition sampling variance. This technique satisfied our objective of integrating both composition and design variability from the application of a nested bootstrap from basket sampling and MC set selection respectively.

Compute CPUE and empirical CV over Monte Carlo replicates

The MC draws samples of size $n = p * N_{total}$ from the 'true' population according to specified sampling scenarios and the stratified ratio estimator calculates estimated totals as per species categories using HT. Then, the simulation calculates the distribution of CV after M iterations by calculating the mean of the estimates ($\bar{Y}$) and the SE of the estimated total (the SD of the M estimates): $SE(\bar{Y}) = \left(\frac{1}{M-1} \sum_{i=1}^M (\bar{Y}_i - \bar{Y})^2 \right)^{1/2}$, and the overall

CV simulation for particular sampling scheme was evaluated as: $CV_{sim} = \frac{SE(\bar{Y})}{\bar{Y}} * 100$. Conclusively, bootstrapped CV was computed to generate 95% CI.

Assumptions

Logbook reported set totals are treated as reliable baselines and within-set expansion, the ratio estimation assumed sampled baskets are representative.

In the absence of sampled CPUE at basket level as discussed above, a true CPUE ratio raise $\left(\frac{\text{sampler}_{catch}}{\text{sampler}_{hooks}} \right) * H_i$ and inference of total catch of baskets B_{full} that would have been applied for within-set expansion patterns was not possible and therefore, the study assumed that the observed baskets are representative of the within-set species composition by weight. Consequently, those sampled baskets that represented the set composition, were then disaggregated from the set totals by sampled-weight proportions,

which is unbiased for species total since direct observed weights were used as basis for proportioning (Cochran, 1977).

The proportion-based raising fails to acknowledge uncertainty between the relationship of sampled baskets and unobserved baskets, which may have been possible using hooks or basket positions. Therefore, due to this mismatch, the variance was included through bootstrap of basket-level resampling conducted for both basket and set-level in order to reflect uncertainty in true population.

The loss of effort simple standardisation (per 1000 hook) rate at the basket-level was compensated at the set-level by adjusting the CPUE with effects originating at the basket-level for species composition. Thus, since proportional raising at basket-level will only give species composition totals. After raising at basket-level totals, adjusted CPUE at set-level was calculated for raising to total for sets.

Tools and software

As mentioned above, R base software was employed in the analysis and packages used were: sf, ggplot2, lubridate, boot, lme4, naturalearth among others (R Core Team, 2024)

Results and Discussions

Per-stratum precision vs coverage curves

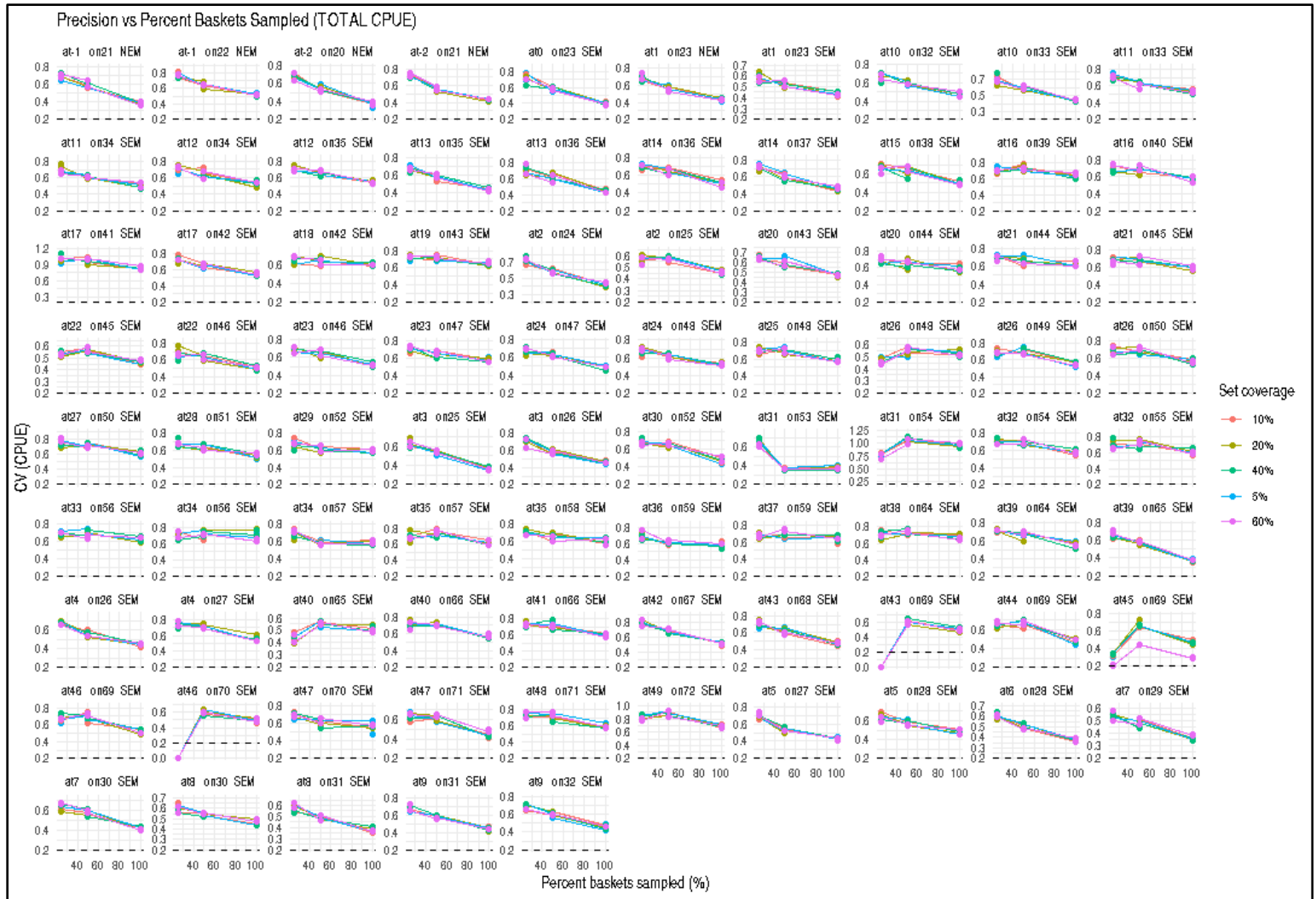


Figure 2. Precision vs Percentage of Basket Sampled (Stand-CPUE) per stratum

The observer data in Kenya showed that basket-level sampling was conducted systematically without exhaustive sampling, which caused the CV to be high showing high variability of mean but as coverage increased at the basket-level, CV consequently increased towards 0.2 or 20% $\pm$ 95% CI.

Fisheries Management Perspective

Rare species require high coverage: species with low encounter rates (like sea turtles or rare sharks) have a high inherent variance and require much higher coverage in order to achieve a desired CV target compared to common target species.

Cost-Benefit Improvement: observer coverage is expensive and management agencies require to justify observer funding based on the data quality generated. The outputs from the simulator provides the answer to the critical question, of how much observer coverage is enough? Thus, by setting acceptable CV for the species categories, managers can use the simulation output to determine the minimum observer coverage percentage required to meet that target. Similarly, managers will then be able to assign limited resources effectively. For instance, the simulation might show that increasing coverage from 10% to 20% cuts the CV in half, but increasing it from 20% to 30% only cuts the CV by a small fraction. This guides the decision to stop at 20% coverage, achieving maximal statistical benefits for the marginal cost.

Accomplish regulatory and conservation requirements: observer data captures meta data for PET species category, which is a mandate for regional through RFMOs and national regulations that for instance may require quantifiable and reliable estimates of their fate (mortality, survival rates or injury). Therefore, the simulation is vital for species that are rarely encountered but protected like sea turtles, some shark species or cetaceans. As the results indicate (fig) and previous studies (Curtis, 2020), rare events are associated with elevated variance, which the simulator reliably demonstrates that achieving an acceptable CV for these species will require potentially increased coverage levels, usually >50% more than target species (Babcock et al., 2003). The study also reinforces accountability of stock assessment or conservation measures adopted by scientists and managers since CV demonstrates measure of confidence in the data used to inform the decision. Therefore, by applying the simulator in design development of the observer program satisfies that the reports and datasets generate are defensible and scientifically authentic.

References

- Cochran, W. G. (1977). Sampling techniques, 3rd ed. John Wiley & Sons.
- Cotter, A. J. R. (1998). Methodology of estimating catch weight and species composition from samples. ICES Cooperative Research Report 228
- Francis, R. I. C. C. (1981). Analytical results for generalised random stratified sampling. N.Z, Journal of Marine Freshwater Resources, 15, pp. 451-457
- Babcock, E. A., Pikitch, E. K., & Hudson, C. G. (2003). How much observer coverage is enough to adequately estimate bycatch? *PEW*,
https://www.pew.org//media/legacy/uploadedfiles/wwwpewtrustsorg/reports/protecting_ocean_life/oceanabycatch110403pdf.pdf
- Curtis, K. A., Carretta, J. V. (2020). ObsCovgTools: Assessing observer coverage needed to document and estimate rare event bycatch. *Fisheries Research* 225: 105493.
<https://doi.org/10.1016/j.fishres.2020.105493>
- FAO. (2002). A manual for fisheries observer programmes. FAO Fisheries Technical Paper, No. 414, Rome
- Francis, R. I. C. C. (1984). An adaptive strategy for stratified random trawl surveys. New Zealand Journal of Marine and Freshwater Research, 18(1), 59-71
- Gokpinar, F., & Arzu Ozdemir, Y. (2012). A Horvitz-Thomson estimator of the population mean using inclusion probabilities of ranked set sampling. *Communications in Statistics-Theory and Methods*, 41(6), pp. 1029-1039
- Grey, C. A. & Kennelly, S. J. (2018). Bycatches of endangered, threatened and protected species in marine fisheries, *Reviews in Fish Biology and Fisheries*, 28(3), 521-541
- Harley, S. J., Myers, R. A., & Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences*, 58(9), 1760-1772
- Hilborn, R., & Walters, C. J. (1992). Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty, Chapman & Hall
- Hulliger, B. (1995). Outlier robust Horvitz-Thompson estimators. *Survey Methodology*, 21(1), 79-87
- Jolly, G. M. & Hampton, I. (1990). A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, 47(7), 1282-1291
- Kesavan Nair, A. K., & Alagaraja, K. (1988). Statistical methods in gear selectivity and gear efficiency studies (Doctoral dissertation, Cochin University of Science and Technology)
- Kiilu, B., Kaunda-Arara, B., Oddenyo, R., Okemwa, G., Mueni, E., Musembi, P., Fulanda, B., Menya-Otieno, L., Okeri, M., Nduku, G., Musembi, J. (2025). Vulnerability assessment of elasmobranch species to fisheries in coastal Kenya: Implications for conservation and management policies. *Marine Policy*, 171, 106459

- Kimura, D. K. (1981). Standardised measure of relative abundance based on modelling log(CPUE) and their application to Pacific Ocean perch (*Sebastes alutus*). *Journal du Conseil*, 39(3), 211-218
- Kunz, M., Bassett, B. A., & Hlozek, R. A (2007). Bayesian estimation applied to multiple species. *Physical Review D- Particles, Fields, Gravitation and Cosmology*, 75(10), 103508
- Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. *John Wiley & Sons*.
- Manly, B. F. (2018). Randomisation, bootstrap and Monet Carlo methods in biology. Capman and Hall/CRC, October 13, 2018.
- Maunder, M. N., & Punt, A. E. (2004). Standardising catch and effort data: A review of recent approaches, *Fisheries Research*, 70(2-3), 141-159
- McInerney, M C. (2014). Evaluation of standard fishery assessment techniques for use in longterm monitoring. Minnesota Department of Natural Resources, Investigational Report, 561
- Newton, M. A. & Geyer, C. J. (1994). Bootstrap recycling: A Monte Carlo alternative to the nested bootstrap. *Journal of the American Statistical Association*, 89(427), 905-912
- Pebesma, E., & Bivand, R. (2023). Spatial data science: With applications in R. Chapman and Hall/CRC
- Pebesma, E. (2018). Simple features for R: Standardised support for spatial vector data. *The R Journal*, 10(1), 439-446
- Pennington, M. (2001). Improved estimation of fish abundance from trawl surveys using log-transformed data *ICES Journal of Marine Science.*, 58(5), pp. 1086-
- Thompson, S. K. (2012). Sampling, 3rd Edition. Wiley.
- R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Thompson, S. K. (2012). Sampling Vol. 755. John Wiley & Sons.
- Wakefield, C. B., Hesp, S. A., Blight, S., Molony, B. W., Newman, S. J., & Hall, N. G. (2018). Uncertainty associated with total bycatch estimates for rarely-encountered species varies substantially with observers' coverage levels: informing minimum requirements for statutory logbook validation. *Marine Policy*, 95, 273-282
- Wickham, H. (2014). Tidy data. *Journal of Statistical software*, 59(10), 1-2
- Wickham, H., Francois, R., Henry, L., & Muller, K. (2023). dplyr: A grammar of data manipulation. R package version 1.1.4
- Quinn, T J., & Deriso, R. B. (1999). Quantitative Fish Dynamics. Oxford University Press