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 Wachira[1], Stephen Ndegwa[1], Elizabeth Mueni[1], Alex Lukwendah[1], Benedict Kiilu[1], Zachary Ogari[1], Donald Obiero[1], Collins Ndoro[1]

Kenya Fisheries Service, State Department for Blue Economy and Fisheries

Abstract

Effective fisheries management requires accurate, spatially-resolved data on target, bycatch and protected, endangered and threatened species. This study integrates longline observer datasets (2018-2025) from the Kenyan coastline to standardize Catch Per Unit Effort (CPUE), quantify its uncertainty, and evaluate sampling biases. We employed a comprehensive, reproducible framework using modern statistical and computational tools within the R ecosystem. Data were harmonized and mapped onto a 1° × 1° grid of Kenya's Exclusive Economic Zone. Nominal CPUE (kg/1000 hooks) was standardized using generalized linear and delta-lognormal models, adjusting for spatiotemporal and operational covariates.

Uncertainty in CPUE and species composition was rigorously quantified using analytical and bootstrap-derived Coefficients of Variation (CVs). A nested resampling approach simultaneously captured variation in catch proportions and sampling effort, ensuring realistic propagation of uncertainty. The workflow included diagnostic modules to flag statistical deviations, outliers, and spatial coverage gaps, safeguarding data integrity. Spatial analyses produced gridded summaries of species distribution, relative abundance, and biodiversity, highlighting species-specific hotspots and bycatch concentration zones.

A critical focus was addressing non-random sampling bias inherent in voluntary observer programs. The framework assesses the representativeness of observed vessels using historical observer data from Kenya, a vital step for bias correction. Our analysis indicates that while 20% observer coverage yields acceptable precision for common species, rare species require coverage exceeding 60% to reduce uncertainty to manageable levels, with requirements varying spatially and seasonally. This CV-based approach provides a quantitative benchmark for designing efficient, stratified sampling strategies that balance logistical constraints with statistical rigor.

Ultimately, this study demonstrates a scalable and transparent methodology that bridges operational sampling with modern analytics. By integrating standardization models, spatial stratification, and resampling-based uncertainty quantification, it enhances the credibility of fisheries indices in data-limited regions, directly supporting sustainable management, adaptive monitoring, and robust stock assessments for the Western Indian Ocean.

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

1.0 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.

1.1 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.

1.2 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 (Mclnerny, 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.

2.0 Materials and Methods

2.1 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 $n = total\ sets$ 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).

2.2 Description of Spatial Stratification

The positions for start setting and end hauling were obtained from the logbook dataset, which contained the reported catches for both individual species and aggregated species categories. The line connecting these two points, theoretically set to haul, represents the spatial trajectory of the longline fishing gear and consequently defines the spatial distribution of catches (Francis, 1984).

Each longline set was converted into a LINESTRING geometry using the dateline to ensure accurate spatial continuity, resulting in one geometry per set (Pebesma & Bivand, 2023). This enabled precise spatial allocation of each fishing set across 1° × 1° grid cells. A polygon grid at this finer spatial resolution was constructed and clipped to the boundaries of the Western Indian Ocean FAO Area 51, thereby establishing a spatial stratification framework for area-based management (FAO, 2002).

Subsequently, each LINESTRING was intersected with the FAO-cropped 1° grid, producing individual line segments corresponding to each combination of set and grid cell. The length of each segment within a given cell was calculated following the method described by Jolly & Hampton (1990). These segments were then divided into geodetic length fractions as follows:

For each set line L_i, its intersection with grid cell c was used to determine the fractional distribution of catch and effort:

$wt_{ic} = \mathrm{total}_{\mathrm{wt_i}} x \frac{length(L_i \cap cell_c)}{length(L_i)}$, (Pebesma & Bivand, 2023; Francis, 1984)

This procedure was necessary due to the absence of positional data at the hook or basket level, making it impossible to precisely locate individual hooks along the gear. Instead, basket-level catches were proportionally distributed across intersected grid cells according to line length fractions. This approach ensured that both catch and effort (standardized CPUE) were appropriately allocated, even when basket positions were missing or incomplete.

Finally, reported catches of all species and species categories were distributed across their respective intersected 1° grid cells by these fractional weights and aggregated to the cell level. The resulting logbook dataset provided cell-level catch and effort summaries, forming the basis for area stratification and for generating spatial templates representing the "true" population distribution (Pebesma, 2018).

At the basket level, the sum of individual species weights within each basket was used to compute the total basket weight. For each set, the total weight of all species across observed baskets was compared to the total reported set weight to validate sampling completeness (Francis, 1984). This comparison revealed that some baskets within sets were unobserved or missing data, confirming partial observation at the basket level.

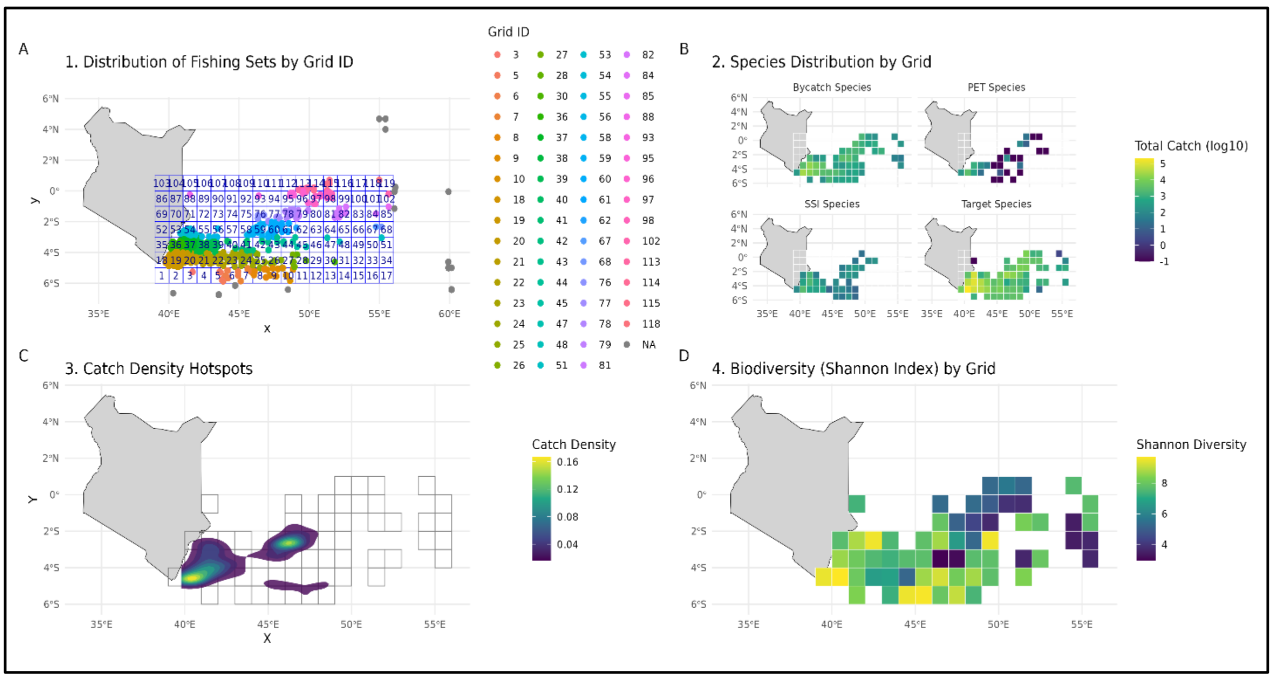

Figure 1. Comprehensive Overview of Catch Composition Per 1°x1° Grid Cell from Kenya's Observer Catch and Effort Data 2018-2015, with illustration as follows: (a) Distribution of fishing sets by grid cell, (b) species distribution (bycatch, PET, SSI & target species), (c) Catch diversity hotspots, (d) Shannon biodiversity index (Developed in R base 4.4.3)

2.3 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:

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

2.4 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 ad caters from multiplicative effects (Hilborn & Walters, 1992).

Therefore, the model was performed as:

$E[Y_i] = H_i^{obs} * \mu_i$, so that $CPUE_{s,i}^{nominal} \sim Gamma\ (mean = \mu_{s,i}, link = \log)$, 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.

$$E[Y_{is}] = \mu_{is} = \exp{(\beta_0 + \beta_1 \log(H_i^{obs}) + \beta_2 Area_i + \beta_3 Season + \beta_4 Species_s(wt^{raised}) + \epsilon_i)}$$

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 categorial that impacts spatial stratum developed for 5° grid cell, $Season_i$ as a categorical outcome, $Species_s(wt^{raised})$ 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: $CPU\bar{E}_{h,s}^{std} = \exp{(\bar{\beta}_o + \bar{\beta}_{2,h} + \bar{\beta}_{3,s} + \bar{\beta}_{4,s} + \cdots)}$, which eliminates effort and sampling biases (Hiborn & Walters, 1992).

2.5 Generate Baseline Templates for True Population

This stage was important to introduce the standardised CPUE $(CPU\bar{E}_{h,s}^{std})$ 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 $(CPU\bar{E}_{h,s}^{std})$ (already defined above). Therefore,

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

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): $w_{ibs}^{pop} = P_{ib} * Y_{is}$, $b = 1, \ldots, B_i$.

Thus, the totals for each basket assigned became:

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

3.0 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 $CPU\bar{E}_i^{std}$. 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:

$p_i^{(prop)} = \frac{size_i}{\sum_{j=1}^{N} size_j}$, whereby N 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 (Thompson, 2012). 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: $\left(1 - p_i^{(prop)}\right)^{m_h}$(Hulliger, 1995). On the hand, the probability of that particular set i being picked at least once was denoted as:

$\pi_i^{(PPSWR)} = 1 - \left(1 - p_i^{(prop)}\right)^{m_h}$ (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 quicky 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).

3.1 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:

3.1.1 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).

3.1.2 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:

$$\bar{P}_{is}^{(r)} = \frac{\sum_{b \epsilon \boldsymbol{S}_i^{(r)}} w_{ibs}}{\sum_{b \epsilon \boldsymbol{S}_i^{(r)}} \sum_s w_{ibs}}$$

- 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.

4.0 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): $SE(\bar{Y}) = \sqrt{\frac{1}{M-1} \sum_{i=1}^{M} (\bar{Y}_i - \bar{Y})^2}$, 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.

5.0 Assumptions

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

II. In the absence of sampled CPUE at basket level as discussed above, a true CPUE ratio raise $(\frac{\text{sampled}_{\text{catch}}}{\text{sampled}_{\text{hooks}}}) * H_i$ 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 an unbiased for species total since direct observed weights were used as basis for proportioning (Cochran, 1977).

III. 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.

IV. 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.

6.0 Tools and software

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

7.0 Results and Discussions

7.1 Spatial Stratification in Context of Kenya's Observer Program

The spatial analysis based on the 1°×1° grid cells provided a detailed overview of catch composition and species distribution patterns from Kenya's longline observer data (2018–2025). The results are summarized in four key visual outputs: (a) distribution of fishing sets by grid cell, (b) species distribution across bycatch, PET (Protected, Endangered, and Threatened), SSI (Species of Special Interest), and target groups, (c) catch diversity hotspots, and (d) Shannon biodiversity index values (figure 1).

The distribution of fishing sets by grid cell revealed uneven observer coverage across the fishing grounds, with several grid cells showing dense activity concentrated in limited areas. This pattern highlights the spatial variability in fishing effort and underscores the importance of increasing observer coverage at the basket or set level. In areas where observer presence was sparse, the representativeness of catch and effort data was reduced, limiting the precision of spatially explicit CPUE and biodiversity estimates.

The species distribution maps demonstrated that bycatch, PET, SSI, and target species are spatially distinct, often occupying different ecological zones within the Western Indian Ocean. Such segregation indicates that pooling data across wide areas without sufficient sampling resolution may mask important spatial patterns in species occurrence and abundance. These findings strongly reinforce the need for 100% basket-level sampling or, at minimum, observer coverage exceeding 60% to capture the true variability in catch composition and species interactions.

The catch diversity hotspots further illustrated localized areas of high species richness, particularly along productive oceanographic zones near the Kenyan EEZ boundary. Similarly, the Shannon biodiversity index revealed moderate to high diversity in several nearshore and offshore grid cells, consistent with known multi-species assemblages in the region. Together, these indicators demonstrate that catch diversity is not uniformly distributed and that targeted sampling strategies are necessary to account for this spatial heterogeneity.

Overall, the 1°×1° grid approach provided enhanced spatial precision, allowing for a finer-scale understanding of fishing activity, species distribution, and biodiversity. The results support the conclusion that improved observer coverage and basket-level sampling are essential for reducing uncertainty in bycatch and biodiversity assessments and for strengthening ecosystem-based management of Kenya's longline fishery.

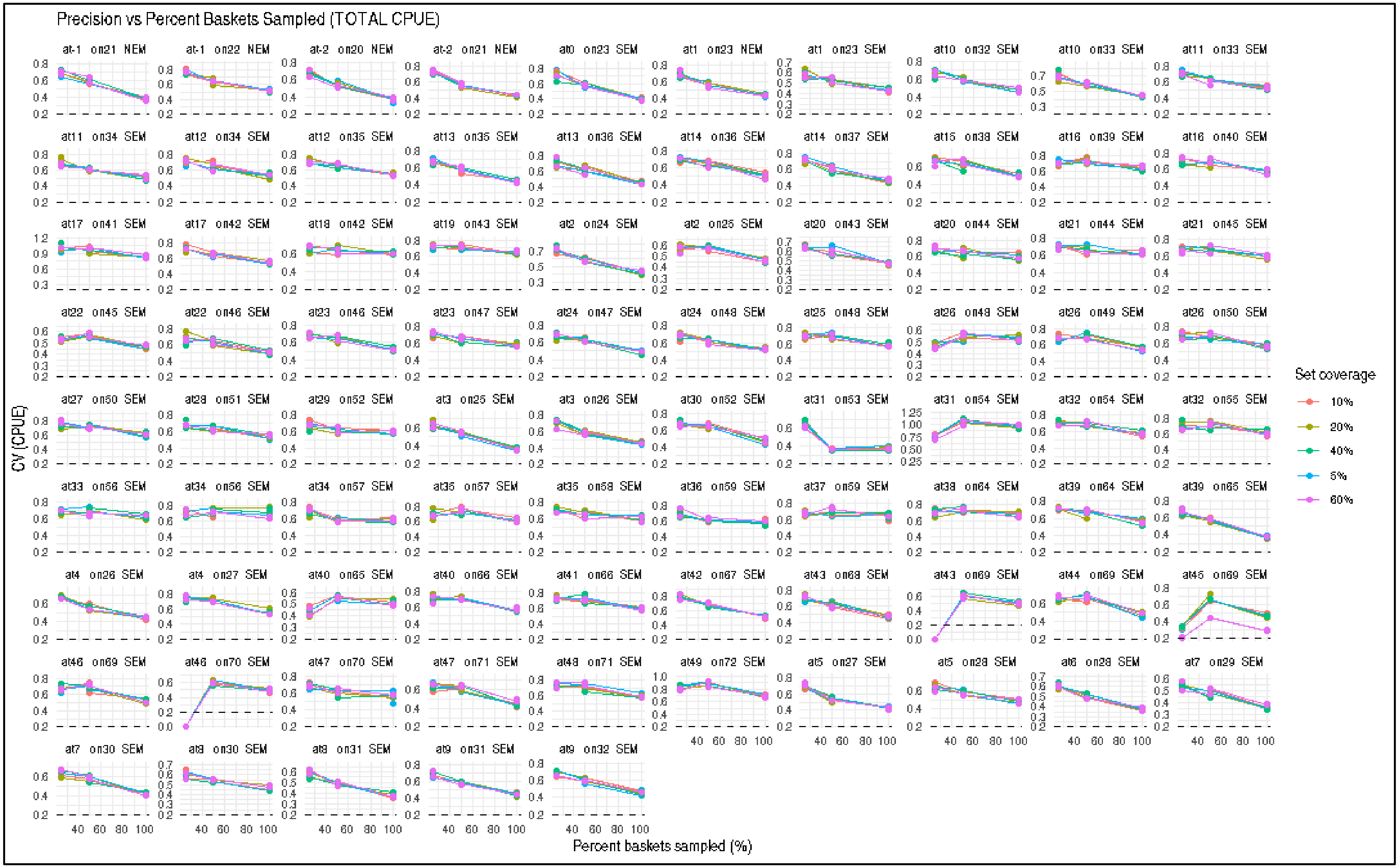

Figure 2. Precision vs Percentage of Basket Sampled (Stand-CPUE) per stratum

In figure (2) the Monte Carlo (MC) simulation results illustrate how precision (CV of CPUE) changes with increasing percent of baskets sampled and varying set coverage levels across spatial strata (1°×1° grids). Each panel represents a stratum for example *at01–at80* whether both SEM and NEM seasons, with curves showing how precision improves as more baskets are sampled within a set.

Across most strata, the CV (Coefficient of Variation) values decline steadily as the proportion of baskets sampled increases. This inverse relationship indicates that precision of CPUE estimates improves with sampling intensity, confirming that limited basket-level coverage introduces significant uncertainty in estimating catch rates. In several strata, CV values drop sharply between 20% and 60% sampling, after which gains in precision taper off — implying a diminishing return beyond 60–80% basket coverage.

Comparisons among the coloured curves (representing 10%, 20%, 40%, 50%, and 60% set coverage) reveal that higher set coverage consistently produces lower CVs across strata. However, some variability remains, particularly in strata with fewer observed sets or greater heterogeneity in catch composition such that even high set coverage did not fully stabilize precision. This variation underscores that spatial heterogeneity in fishing effort and species composition can amplify sampling uncertainty unless both basket and set coverage are adequate. Therefore, in fisheries management perspective that involves planning of observer deployment and validation of the datasets, the study revealed that:

- Attention to adequate sampling effort: The precision verses coverage relationship suggests that 60% basket-level sampling combined with ≥50% set coverage achieves acceptable precision (CV ≤ 0.3) for most strata. This balance optimizes observer effort without incurring unnecessary costs.
- Planning for coverage for specific strata: Precision gains vary across strata, indicating that sampling designs should be adaptive. Strata with greater variability in CPUE (e.g., high bycatch or mixed species zones) may require intensified sampling or stratified subsampling protocols.
- Data reporting quality and validation process: The MC-based validation provides a quantitative basis for evaluating observer data quality. By comparing simulated (true population) and sampled CPUE distributions, managers can identify under-sampled strata and potential biases in observer deployment.
- Attention to uncertainties and biasness: Since observer presence and sampling are rarely random, ensuring adequate basket and set coverage helps minimize bias in catch and bycatch estimates, especially critical when estimating bycatch of PET and SSI species.

Conclusively, the precision curves demonstrated that robust observer coverage is vital to data reliability. The use of simulated "true populations" allows for realistic testing of sampling strategies, offering a strong empirical basis for setting minimum coverage targets and validating the representativeness of observer data across Kenya's longline fishery strata.

7.2 Fisheries Management Perspective

The simulation results shown above for both plots and tables demonstrated how effectively the analysis met the three core objectives related to precision, sampling sufficiency, and stratification design for the longline fishery observer data.

To estimated CV for strata with ±95% confidence bounds, each panel in the figure 2 represented a distinct 1°×1° spatial stratum, where the coefficient of variation (CV) was calculated from the Monte Carlo replicates of standardized CPUE. The downward trend in the curves shows that as more baskets are sampled within each set, the variability in CPUE decreases, confirming the computation of precision per stratum. Although the 95% confidence bounds are not displayed directly, the narrowing spread of CVs across simulations indicates increasing confidence in CPUE estimates at higher sampling fractions.

Also, to determine the minimum coverage and basket sampling proportions that fulfills targets in management, the precision curves showed clearly that precision improves rapidly between 20% and 60% basket coverage, after which gains begin to flatten. This pattern identifies an optimal range of sampling coverage (50–60%) that achieves acceptable CV values (typically below 0.3), aligning with management standards for reliable catch rate estimates. Strata that maintain high CVs even at higher coverage levels signal statistically unreliable areas, likely due to low observer effort or high catch variability, and thus need targeted increases in sampling intensity or revised deployment strategies.

Furthermore, to validate operational stratification scheme, by evaluating precision within each 1°×1° grid cell, the analysis confirmed that the current stratification successfully divides the fishery into relatively homogenous units. Within each stratum, CPUE variability was lower than across the entire fishery, showing that fishing effort and catch composition are internally consistent within these grid-based areas. This supports the statistical appropriateness of the stratification scheme for monitoring trends and managing observer effort efficiently.

8.0 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., & Waletrs, 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., Musembei, 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

Mclnerny, 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, 101(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). Tudy 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