How Observation Coverage Shapes Bycatch Metrics in the Tropical Tuna Purse Seine Fishery

Giancarlo M. Correa[1], Jon Ruiz[1], María Lourdes Ramos[2], Arnaitz Mugerza[1], Miguel Herrera[3], Nekane Alzorriz[4]

[1]AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta, Bizkaia, Spain
[2]Centro Oceanográfico de Canarias, Instituto Español de Oceanografía (IEO-CSIC), C. Farola del Mar 22, 38180 San Andrés, Santa Cruz de Tenerife, Spain
[3]Organización Productores Asociados Grandes Atuneros Congeladores (OPAGAC-AGAC), Cl. de Ayala, 54, 28001 Salamanca, Madrid, Spain
[4]Asociación Nacional de Armadores de Buques Atuneros Congeladores (ANABAC), Txibitxiaga Kalea, 24, 48370 Bermeo, Bizkaia, Spain

Abstract

Keywords

bycatch, pelagic species, spatiotemporal models, purse seine fleet, tunas

1 Introduction

Purse seine fishing is a technique that targets and catch entire fish schools in the surface by encircling them with a fishing net called "seine". In the tropical oceans, purse seine is employed to target tropical tunas such as skipjack (*Katsuwonus pelamis*), yellowfin (*Thunnus albacares*), and bigeye (*Thunnus obesus*). In addition to target species, tropical tuna fisheries catch non-target species collectively referred to as bycatch, which can be discarded at sea, retained to be sold on local markets, or consumed on board (Hall et al., 2017). The ratio of bycatch to target tuna catch in the purse seine fleet is considered relatively low in comparison to other fishing gears, such as longlines, that can result in substantial levels of bycatch (Amandè et al., 2010; Liu et al., 2008; Peatman et al., 2023). However, the impact on pelagic populations and ecosystems may be important, especially for vulnerable long-lived species with low reproductive rates (Dulvy et al., 2014). Therefore, it is essential to conduct studies on bycatch rates and their variability over space and time; however, they are often complicated by a lack of bycatch data recorded in fisheries logbooks, taxonomic identification of bycatch species, among other factors.

One of the most reliable sources of information to quantify the amount of bycatch is the use of on-board trained scientific data collectors (a.k.a. observers). When designing an observer sampling program, the level of coverage required will depend on the objectives of the observer program, which might vary from estimating bycatch of protected species, to improving bycatch and catch data for assessment of fish populations, to collecting biological data. In some cases, it may be necessary to have an exact count of the total incidental mortality of bycatch species, especially threatened or endangered species, so a 100% observer coverage may be needed. However, in most cases, a level of 100% observer coverage is not attainable, then the coverage level chosen must ensure that the total bycatch estimate is sufficiently accurate and precise. Then, assuming these observed units are representative of unobserved activity, design-based (e.g., ratio estimators, Cochran (1977)) or model-based (e.g., generalized linear models, Coelho et al. (2020)) approaches can be used to expand the observed bycatch to the remainder of the fishery. One of the main features of ratio estimators is that they do not incorporate a formal underlying statistical model (i.e. are free of any assumptions regarding data structure), and therefore are broadly used in fisheries worldwide, including tuna purse seine fisheries (Amandè et al., 2012; Amandè et al., 2010).

Despite their widespread use, there are a number of potential issues in applying ratio estimators to estimate bycatch. First, using observed catches of target species or any other measure of effort implicitly makes an assumption about a linear relationship between non-target and target catches (Amandè et al., 2012; Fonteneau and Richard, 2003). This may be unrealistic since the distribution of catches of non-target species is often zero-inflated or has a small number of observations containing extremely high values (Ortiz and Arocha, 2004), and the liner relationship may not hold (Stock et al., 2019). Second, the boundaries of strata used in a ratio estimator can be somewhat arbitrary whenever poststratified boundaries are used. For instance, Amandè et al. (2010) defined strata in the Atlantic Ocean based on ecological features for estimating bycatch of the tuna EU purse seine fishery, which may not be adequate for all bycatch species. Third, for rare-event bycatch species, it is common for zero bycatch events to be observed in a given year (ratio estimator is equal to 0), and when bycatch events are observed, the ratio estimator often delivers implausibly high estimates. Lastly, a final and related point is that within each stratum, bycatch rates are assumed to be uniform, while in reality they may vary by season, depth, or other factors.

Spatiotemporal models are increasingly adopted in multiple fisheries applications (Ducharme-Barth et al., 2022; Grüss et al., 2023; Thorson, 2019), including undertaking bycatch analyses (e.g., Yan et al. (2022)). These models can provide detailed predictions for any location based on spatial autocorrelation in the observations. However, they are also complicated and require more data to generate robust predictions, which make it unsuitable for data-poor fisheries. In the majority of cases, spatially explicit model-based estimators have increased precision relative to simpler estimators that assign observations to strata (Thorson et al., 2015; Thorson and Ward, 2013). There are a number of additional advantages of spatial models, including the ability to better quantify shifts in distribution (Thorson et al., 2016) and improved ability to identify fine-scale hotspots of high bycatch (Cosandey-Godin et al., 2015). For tuna fisheries, spatiotemporal models like generalized additive models (GAMs) have recently been used to obtain annual estimates of the most important bycatch species (Dumont et al., 2024; Peatman et al., 2023).

In this study, we implemented a simulation experiment to evaluate the impacts of different levels of sampling coverage on the annual bycatch estimates derived from design-based and model-based estimators. Our hypothesis is that model-based estimators may provide more accurate bycatch estimates under low sampling coverage scenarios. We used data from the Spanish purse seine tuna fishery operating in the Atlantic Ocean as study case. We performed our analyses differentiating by set type: operating on free schools (FSC) or floating objects (FOB), since they have different bycatch dynamics (Peatman et al., 2023). Our simulation experiment may be extended to other fisheries with fine-scale bycatch information and support the implementation of sampling programmes across tuna RFMOs.

2 Methods

In the following sections, we describe the data used in our analyses, the design and model-based estimators, and the simulation framework. The analyses described below were performed by set type independently.

2.1 Data

Our analyses use data collected by scientific observers aboard tropical purse seine vessels operating in the Atlantic Ocean between 2015 and 2023 (*observers data*). The dataset includes records from both the Spanish scientific monitoring program (EU Data Collection Framework) and the industry-funded Best Practices program, covering Spanish vessels and those under other flags affiliated with ANABAC and OPAGAC.

Regardless of the monitoring program, observers followed a standardized protocol. They recorded detailed information for each fishing set, including estimates of target tuna catch and bycatch. For larger bycatch species such as elasmobranchs and billfish, all individuals were counted. For smaller, more abundant species, estimates were often based on visual assessments. In addition to counts, observers conducted length sampling to convert numbers into biomass using species-specific length-weight relationships. Taxa were identified to the species level whenever possible, although in rare cases, only higher taxonomic groups (e.g., family) were recorded.

The *effort data* used in our analyses pertain exclusively to the Spanish fleet for the same period and fishing ground. This information is sourced from the logbooks completed by the captains, who are required to record details of each fishing trip, including the location and date of every fishing operation.

This study does not present results for all taxa found in the purse seine fleet's bycatch. Instead, a targeted selection of species or groups of species was made to represent three categories: the most abundant taxa, the rare or less prevalent ones, and those considered vulnerable or of special interest. The list of these taxa are presented in Table XX.

2.2 Model fitting

We use the *sdmTMB* geostatistical spatiotemporal model (Anderson et al., 2024) to fit taxon-specific bycatch per set (in weight) using the observers data. Geostatistical spatiotemporal models have become widely used in fisheries over the last decade (Anderson et al., 2024; Thorson, 2019), and are used when georeferenced data (e.g. each has a corresponding latitude and longitude) with an underlying spatial process is available. *sdmTMB* is written in Template Model Builder (TMB, Kristensen et al. (2016)) and R (Team, 2025) for a friendly user interface, and can be viewed as an extension of generalized linear mixed models (GLMMs), but with additional spatial and spatiotemporal components, which are approximated as random effects.

Mathematically, the model structure can be expressed as:

$$u_{s=1:S,t} = f^{-1}(Xb + \omega + \epsilon_t)$$

where$\breve{}$ represents the bycatch predictions at locations $s = 1 : S$ for a given year t, $f^{-1}()$ is the inverse link function, X represents the design matrix of fixed effects, b is the vector of estimated parameters, ω are the estimated latent spatial effects, and ϵ_t represents year-to-year latent spatiotemporal effects. ω represents a spatial intercept that is constant with time while ϵ_t represents spatial deviations over time, and both are modelled as Gaussian random fields (GRFs):

$$\omega \sim MVN(0, \Sigma_\omega)$$
$$\epsilon \sim MVN(0, \Sigma_\epsilon)$$

where MVN is the multivariate normal distribution, and the covariance matrix Σ is modelled with Matérn covariance (Lindgren et al., 2011; Matérn, 1986), which defines the rate at which spatial covariance decays with distance. *sdmTMB* approximates the GRF by relying on the Stochastic Partial Differential Equation (SPDE) approach using the Integrated Nested Laplace Approximation in *R-INLA* to reduce computational costs (Rue et al., 2009). The first step when using the SPDE approach is to construct the mesh, which, in our case, was composed of triangles covering the studied area with a minimum allowed triangle edge length (*cutoff*) of 1.5 degrees.

For some taxa, especially the less recurrent ones, the inclusion of both the spatial and spatiotemporal terms caused the model failed to converge. For those cases, we reran the model only including the spatial term in order to simplify the model structure.

We used the Tweedie distribution $Tweedie(\mu, \phi^2, p)$, where $1 < p < 2$, and a log link function (Tweedie, 1984). The Tweedie model is an extension of compound Poisson model derived from the stochastic process where the weight of the response variable (e.g., catch data) has a gamma distribution and has an advantage of handling the zero-catch data in a unified way (Shono, 2008). For fixed effects, we incorporated the year and quarter effects as factors, and the target tuna catch (the sum of skipjack, yellowfin, and bigeye tunas) as a continuous covariate.

For all fitted models, we checked that the maximum gradient was smaller than 1e-03, the Hessian was invertible, and standard errors were estimated for all fixed effects and did not look unreasonably large ("safety checks"). We then used the *DHARMa* R package (Hartig, 2022) to evaluate the model residuals. Standard raw residuals are not always appropriate when using generalized linear models, and other types of residuals are commonly used instead. *DHARMa* uses a simulation-based approach to create readily interpretable scaled (quantile) residuals for generalized linear mixed models. We analyzed two plots produced by *DHARMa*: 1) the QQ plot residuals, which detects overall deviations from the expected distribution, and 2) the residual vs. predicted plot, which detects trends in residuals along model predictions and simulation outliers.

2.3 Simulation

One of the advantages of using models like *sdmTMB* is that we can simulate new observations using a new dataset ("prediction dataset") containing the same covariates used when fitting the model. In our case, we used the effort data as the prediction dataset and simulated new observations (i.e., bycatch in weight for every fishing set in the effort data) using the fitted models for each taxa in Section 2.2. These simulated observations keep the statistical properties of the original bycatch data. We refer to the effort data with simulated bycatch observations as the *simulated data*.

We then took a subset of the simulated data with different sampling coverage scenarios: 5%, 10%, 20%, 30%, 40%, 50%, 70%, and 90%. To approximate real-case situations, we performed this subsetting stratified by year, and then selected the fishing trips observed under that sampling coverage scenario. The obtained *sampled data* represent the observers data that would have been obtained from an observers program with the specified sampling coverage.

Then, using the sampled data, we estimated the annual bycatch using two approaches: ratio and model-based estimator, which are described below.

2.3.1 Ratio estimator

We used the spatially-stratified bycatch–over-target catch ratio. For a given taxon, the ratio ($R_{y,a}$) was calculated for every defined $5 \times 5^{\circ}$ grid a in the study area and year y as follows:

$$R_{y,a} = \frac{B_{y,a}}{T_{y,a}}$$

where $BY_{y,a}$ is the total bycatch and $TG_{y,a}$ the total tropical tuna catch obtained from the sampled data. In a few cases, there could happen that $T_{y,a} = 0$ (e.g., sets with target catch equal to zero or "null sets"), so $R_{y,a}$ could not be calculated. Therefore, exclusively for those cases, we assumed that $R_{y,a} = 0$.

Then, assuming a linear relationship between bycatch and target catch, we calculated the total bycatch:

$$\hat{B}_{y,a} = R_{y,a} T^*_{y,a}$$

where $T^*_{y,a}$ is the total target catch in grid a and year y obtained from the effort data. Especially for low sampling coverage scenarios, it is expected to have missing $R_{y,a}$ values for some grids due to the sampled data do not cover all the grids in the study area. Therefore, exclusively for those grids, $R_{y,a}$ was calculated using B_y/T_y, where B_y and T_y are the total bycatch and target catch in the whole area, respectively, derived from the information in the sampled data.

Finally, the annual bycatch estimate is calculated: $\hat{B}_y = \sum_a \hat{B}_{y,a}$.

2.3.2 Model-based estimator

We followed the same modelling framework described in Section 2.2. Once the fitted model is obtained, we then made predictions using the effort data, which generated predicted bycatch observations for every fishing set in that dataset. Then, we summed the predicted bycatch values per year to estimate the annual bycatch $\hat{B}_y$.

Especially when the sampling coverage is low, we could find cases when a given taxon is not detected in the sampled data (i.e., bycatch equal to zero for all fishing sets). In those cases, the model-based estimator was not run and we assumed $\hat{B}_y = 0$ for all years. Another special case is when the model did not pass the safety checks (see Section 2.2). In those cases, we were not able to produce bycatch estimates $\hat{B}_y$ and reported the rate of model failure.

2.4 Performance

The procedure explained in Section 2.3 was repeated 100 times ("replicates") with different seeds to produce the simulated data, therefore we obtained 100 annual bycatch estimates by the ratio and model-based estimators for each taxa q. We calculated the relative error for every replicate i: $RE_{i,q,y} = (\hat{B}_{i,q,y} - B_{i,q,y})/B_{i,q,y}$, where $B_{i,q,y}$ represents the true annual bycatch obtained from the simulated data. The width of the 95% quantile of RE over replicates was used as a measure of precision and the median as a proxy of bias, and these metrics were used to compare the performance of the ratio and model-based estimators.

3 References

Amandè, M.J., Ariz, J., Chassot, E., De Molina, A.D., Gaertner, D., Murua, H., Pianet, R., Ruiz, J., Chavance, P., 2010. Bycatch of the European purse seine tuna fishery in the Atlantic Ocean for the 2003–2007 period. Aquatic Living Resources 23, 353–362. https://doi.org/10.1051/alr/2011003

Amandè, M.J., Chassot, E., Chavance, P., Murua, H., De Molina, A.D., Bez, N., 2012. Precision in bycatch estimates: The case of tuna purse-seine fisheries in the Indian Ocean. ICES Journal of Marine Science 69, 1501–1510. https://doi.org/10.1093/icesjms/fss106

Anderson, S.C., Ward, E.J., English, P.A., Barnett, L.A.K., Thorson, J.T., 2024. sdmTMB: An R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv : the preprint server for biology. https://doi.org/10.1101/2022.03.24.485545

Cochran, W.G., 1977. Sampling Techniques, 3rd ed. John Wiley & Sons, New York.

Coelho, R., Infante, P., Santos, M.N., 2020. Comparing GLM, GLMM, and GEE modeling approaches for catch rates of bycatch species: A case study of blue shark fisheries in the South Atlantic. Fisheries Oceanography 29, 169–184. https://doi.org/10.1111/fog.12462

Cosandey-Godin, A., Krainski, E.T., Worm, B., Flemming, J.M., 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Canadian Journal of Fisheries and Aquatic Sciences 72, 186–197. https://doi.org/10.1139/cjfas-2014-0159

Ducharme-Barth, N.D., Grüss, A., Vincent, M.T., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., Thorson, J.T., 2022. Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application. Fisheries Research 246, 106169. https://doi.org/10.1016/j.fishres.2021.106169

Dulvy, N.K., Fowler, S.L., Musick, J.A., Cavanagh, R.D., Kyne, P.M., Harrison, L.R., Carlson, J.K., Davidson, L.N., Fordham, S.V., Francis, M.P., Pollock, C.M., Simpfendorfer, C.A., Burgess, G.H., Carpenter, K.E., Compagno, L.J., Ebert, D.A., Gibson, C., Heupel, M.R., Livingstone, S.R., Sanciangco, J.C., Stevens, J.D., Valenti, S., White, W.T., 2014. Extinction risk and conservation of the world's sharks and rays. eLife 3, e00590. https://doi.org/10.7554/eLife.00590

Dumont, A., Duparc, A., Sabarros, P.S., Kaplan, D.M., 2024. Modeling bycatch abundance in tropical tuna purse seine fisheries on floating objects using the Δ method. ICES Journal of Marine Science 81, 887–908. https://doi.org/10.1093/icesjms/fsae043

Fonteneau, A., Richard, N., 2003. Relationship between catch, effort, CPUE and local abundance for non-target species, such as billfishes, caught by Indian Ocean longline fisheries. Marine and Freshwater Research 54, 383–392. https://doi.org/10.1071/MF01268

Grüss, A., McKenzie, J.R., Lindegren, M., Bian, R., Hoyle, S.D., Devine, J.A., 2023. Supporting a stock assessment with spatio-temporal models fitted to fisheries-dependent data. Fisheries Research 262, 106649. https://doi.org/10.1016/j.fishres.2023.106649

Hall, M., Gilman, E., Minami, H., Mituhasi, T., Carruthers, E., 2017. Mitigating bycatch in tuna fisheries. Reviews in Fish Biology and Fisheries 27, 881–908. https://doi.org/10.1007/s11160-017-9478-x

Hartig, F., 2022. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software 70. https://doi.org/10.18637/jss.v070.i05

Lindgren, F., Rue, H., Lindström, J., 2011. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. Journal of the Royal Statistical Society Series B: Statistical Methodology 73, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Liu, K.-M., Tsai, W.-P., Joung, S.-J., 2008. Preliminary estimates of blue and mako sharks bycatch and cpue of Taiwanese longline fishery in the Atlantic Ocean (No. SCRS/2008/153). ICCAT (International Commission for the Conservation of Atlantic Tunas), Madrid, Spain.

Matérn, B., 1986. Spatial Variation, 2nd ed. Springer-Verlag, New York, NY.

Ortiz, M., Arocha, F., 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: Billfish species in the Venezuelan tuna longline fishery. Fisheries Research 70, 275–297. https://doi.org/10.1016/j.fishres.2004.08.028

Peatman, T., Allain, V., Bell, L., Muller, B., Panizza, A., Phillip, N.B., Pilling, G., Nicol, S., 2023. Estimating trends and magnitudes of bycatch in the tuna fisheries of the Western and Central Pacific Ocean. Fish and Fisheries 24, 812–828. https://doi.org/10.1111/faf.12771

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. Journal of the Royal Statistical Society Series B: Statistical Methodology 71, 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Shono, H., 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research 93, 154–162. https://doi.org/10.1016/j.fishres.2008.03.006

Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., Semmens, B.X., 2019. The utility of spatial model-based estimators of unobserved bycatch. ICES Journal of Marine Science 76, 255–267. https://doi.org/10.1093/icesjms/fsy153

Team, R.C., 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Thorson, J.T., Pinsky, M.L., Ward, E.J., 2016. Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. Methods in Ecology and Evolution 7, 990–1002. https://doi.org/10.1111/2041-210X.12567

Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science 72, 1297–1310. https://doi.org/10.1093/icesjms/fsu243

Thorson, J.T., Ward, E.J., 2013. Accounting for space–time interactions in index standardization models. Fisheries Research 147, 426–433. https://doi.org/10.1016/j.fishres.2013.03.012

Tweedie, M.C.K., 1984. An Index Which Distinguishes between Some Important Exponential Families, in: Statistics: Applications and New Directions: Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Indian Statistical Institute, Calcutta, Calcutta, pp. 579–604.

Yan, Y., Cantoni, E., Field, C., Treble, M., Flemming, J.M., 2022. Spatiotemporal modeling of bycatch data: Methods and a practical guide through a case study in a Canadian Arctic fishery. Canadian Journal of Fisheries and Aquatic Sciences 79, 148–158. https://doi.org/10.1139/cjfas-2020-0267

4 Tables

Table 1: List of bycatch taxa by set type (FOB=floating object, FSC=free school), and their classification ('Group' column).

Set type	Taxon	Short name	Description	Group
FOB	*Elagatis bipinnulata*	E. bipinnulata	-	Common
FOB	Balistidae	Balistidae	Mostly *Canthidermis maculata*	Common
FOB	Coryphaenidae	Coryphaenidae	Mostly *Coryphaena hippurus*	Common
FOB	*Acanthocybium solandri*	A. solandri	-	Common
FOB	Carangidae	Carangidae	Mostly *Caranx crysos*	Common
FOB	Carcharhinidae	Carcharhinidae	Mostly *Carcharhinus falciformis*	Special interest
FOB	*Makaira nigricans*	M. nigricans	-	Special interest
FOB	Sphyrnidae	Sphyrnidae	Mostly *Sphyrna mokarran*, *Sphyrna lewini*, and *Sphyrna zygaena*	Special interest
FOB	Cheloniidae	Cheloniidae	Mostly *Eretmochelys imbricata*, *Chelonia mydas*, *Lepidochelys olivacea*, *Lepidochelys kempii*, and *Dermochelys coriacea*	Special interest
FOB	Mobulidae	Mobulidae	Mostly *Mobula birostris* and *Mobula mobular*	Special interest
FOB	Alopiidae	Alopiidae	Mostly *Alopias vulpinus*	Rare
FOB	*Xiphias gladius*	X. gladius	-	Rare
FOB	Lamnidae	Lamnidae	Mostly *Lamna nasus* and *Isurus oxyrinchus*	Rare
FOB	*Prionace glauca*	P. glauca	-	Rare

Set type	Taxon	Short name	Description	Group
FSC	Carcharhinidae	Carcharhinidae	See above	Common
FSC	Mobulidae	Mobulidae	See above	Common
FSC	*Istiophorus albicans*	I. albicans	-	Common
FSC	*Makaira nigricans*	M. nigricans	-	Special interest
FSC	Sphyrnidae	Sphyrnidae	See above	Special interest
FSC	Cheloniidae	Cheloniidae	See above	Special interest
FSC	Molidae	Molidae	Mostly *Mola mola*	Special interest
FSC	Lamnidae	Lamnidae	See above	Rare
FSC	*Prionace glauca*	P. glauca	-	Rare
FSC	Istiophoridae	Istiophoridae	Other marlin species than *Istiophorus albicans* and *Makaira nigricans*	Rare

5 Figures