CPUE standardization for skipjack tuna (*Katsuwonus pelamis*) of the EU purse-seine fishery on floating objects (FOB) in the Indian Ocean

David M. Kaplan[1], Maitane Grande[2], María Lourdes Ramos Alonso[3], Jose Carlos Báez[3], Jon Uranga[2], Antoine Duparc[1], Taha Imzilen[1], Laurent Floch[1], Josu Santiago[2]

Abstract

Abundance indices for Katsuwonus pelamis (SKJ) in the Indian Ocean were derived from the European purse seiner CPUE series (2010-2021) for fishing operations made on floating objects (FOBs). GAMM and GLMM approach were used to standardize the SKJ catch per floating object set. The GLMM approach has been applied to compare the outputs when using an alternative modelling approach and both approaches have been compared to nominal annual CPUE time series. To account for the effort creep, additional explanatory variables have been included in the models. FOB sets have been classified to non-followed FOBs (i.e., randomly encounter FOBs for which the purse seiner has no previous information) and followed- FOBs (dFADs for which the purse seiner is likely to have previous information and therefore the dFAD was not randomly encounter). Densities of instrumented buoys at the 1° × 1°*-month scale and vessel capacity have also been included as explanatory variables. The time of the set relative to local sunrise has been estimated by comparing logbook catch-effort data with VMS vessel trajectory data and this variable has been integrated in the analysis to account for changes in fish aggregations around the FOBs over the course of the day.*

[1] MARBEC (U. Montpellier, CNRS, Ifremer, IRD), Sète, France

[2] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA). Txatxarramendi ugartea z/g, 48395 Sukarrieta - Bizkaia, Spain

[3] Instituto Español de Oceanografía, Centro Oceanográfico de Málaga, Puerto Pesquero, s/n Apdo., 29640, Fuengirola, Málaga, Spain

Introduction

This paper details the methodology and results for the standardization of Indian Ocean skipjack (SKJ) catch per purse seine floating object (FOB) fishing set (hereafter referred to as "CPUE standardization" for simplicity) using data from the EU (Spanish & French) fishing fleet over the period 2010-2021. Major innovations with respect to previous standardization efforts include:

- Inclusion of new predictor variables related to the local time of day of fishing sets and fishing on followed (a.k.a. owned) versus not-followed dFADs
- Corrections to the methodology used to calculate (EU) dFAD buoy densities on $1^{\circ} \times 1^{\circ}$-month strata to both produce a more representative estimate of density and account for the water area of $1^{\circ} \times 1^{\circ}$ grid cells
- The use of both GLMM and GAMM models to standardize CPUEs
- Presentation of a new theoretical approach to developing a model for SKJ CPUE standardization combining two components, one for total catch and the other for species composition. This new approache is presented as a perspective for future work this year or the following
- Evaluation of two approaches to predicting the standardized CPUEs, one based on the approach that has been used previously of averaging predictions over space and the other based on weighting spatial and temporal averaging by the average amount of fishing activity in each spatio-temporal strata

Details regarding all of these innovations are provided in the methods below. For this standardization, we have not used the VAST methodology developed by Akia et al. (2022) primarily due to time limitations and Sosthène Akia having completed his doctorate and moving onto a new position. This methodology also provides fewer advantages when estimating a simple catch per set model as opposed to the multi-component approach used for estimating YFT CPUE.

Methods

Catch-effort dataset

The catch-effort dataset in this study consisted of French and Spanish FOB sets over the period 2010-2021. The initial dataset consisted of 59,092 FOB sets corresponding to 59,062 logbook entries for which catch has been corrected using the T3 procedure (Pianet et al. 2000). The dataset was filtered in order to remove the following incomplete data entries:

- Null sets (3,577 sets)
- Fishing activities corresponding to multiple fishing sets for which estimating catch per set and whether or not sets were on followed dFADs is problematic (60 sets)

- Sets for which set time could not be determined based on comparison with VMS data (9,302 sets). In particular, Spanish VMS data for the period 2010-2012 was not available at the time analyses were carried out, but we expect to obtain this data in the near future and rerun models with this data

After applying all of these filters, the final dataset used for building CPUE standardization models consisted of 46,660 sets.

Major changes in the predictor variables used

In addition to the standard temporal, spatial, fleet and vessel identifier predictor variables included in previous standardization efforts (e.g., Akia et al. 2022), new predictors have been included in this standardization effort:

- **dFAD densities:** The method for calculating 1° × 1°-month dFAD densities has now been standardized across fleets to be the average of (28-31) daily density estimates over each month. Densities have been divided by the water area of each spatial grid cell (in units of the area of a 1° × 1° grid cell at the equator) so final densities have units of average number of EU dFADs per 1° × 1° grid cell at the equator.
- **Hours since local sunrise:** The set time has been estimated for each set based on identifying the closest VMS position for the corresponding vessel between 1 hour before sunrise and 1 hour after sunrise. No time was reported if the VMS and logbook positions differed by more than 20 km in French sets. As VMS data was not yet available for all components of the EU fleet at the time of developing this paper, some data was excluded as no set time could be calculated. Estimated set times were calculated with respect to local time of sunrise and converted to decimal hours to calculate the number of hours since sunrise for the set.
- **Set on followed dFAD:** A boolean indicator of whether or not a FOB set was on a dFAD being tracked by the fishing vessel. Previous work has shown that sets on tracked dFADs (with echosounders) catch on average 10% more tuna than sets on untracked dFADs (Wain et al. 2021). In the case of the Spanish dataset, the sets were classified following the procedure described in Akia et al. (2022), whereas French data was classified following the procedure described in Wain et al. (2021).

In addition to these variables, we also calculated the number of followed and non-followed dFADs within 20 nm (37.04 km) and 250 km of the set position on the day of the set, however, last minute technical issues prevented us from using these variables in model development. The distance of 20 nm corresponds roughly to the maximum distance at which a dFAD can be seen from the vessel crow's nest with binoculars and/or bird radar, and, therefore, one would expect that the total number of (followed and non-followed) dFADs within this area would impact dFAD catches. The distance of 250 km corresponds roughly to the maximum distance a vessel can travel in one 12 hour time period, and, therefore, one would expect the number of followed dFADs within this area to impact dFAD catch rates. We expect to include these distance-based dFAD density variables in future CPUE standardization exercises.

Modeling approaches

Three different CPUE standardization modeling approaches were considered, two of which were carried out and the third of which is presented as a perspective for future work: (i) a GAMM model, (ii) a two-part model that combines a GAMM for estimating total catch per set and a second model (GAMM or random forest) for estimating the proportion of catch that is SKJ, and (iii) a GLMM model.

1-part GAMM model

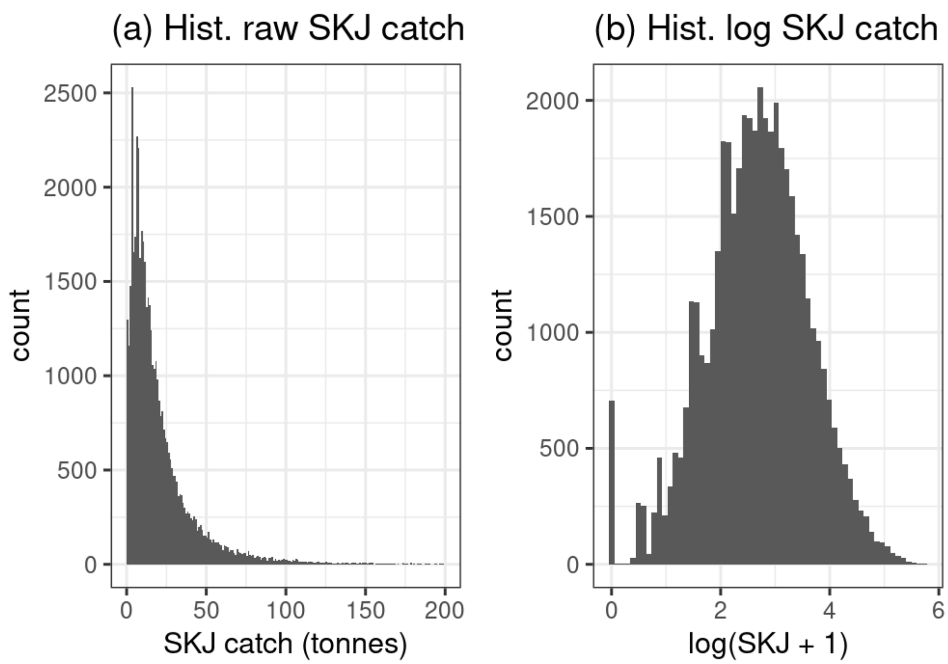

Figure 1: Histograms of SKJ catch data in model training dataset before (a) and afer (b) log transformation.

A single-component general additive mixed-effects model (GAMM) was also run with $\log(SKJ + C)$ as the variable to be predicted, where SKJ is the T3-corrected (Pianet et al. 2000) catch of skipjack per purse seine FOB set. As for a small number of sets (703 sets) zero SKJ catch was reported, a small constant, C, was added to SKJ catch before taking the log. This constant C was chosen to be 1 tonne as this amount is generally used as the limit between null and non-null sets and was observed to produce a response variable that was reasonably close to normally distributed before running the model (Figure 1) and the resulting GAMM had reasonably good model diagnostics (see Results below).

Predictor variables for the GAMM model were longitude and latitude as a tensor product smooth by quarter, year and month as a tensor product smooth cyclic in the month dimension, density (in units of mean daily number of buoys per 1° × 1° water grid cell at the equator), vessel capacity (in units of m^3) and hours since sunrise as individuals smooths, and vessel country and whether or not the set was on a followed dFAD as non-smoothed, categorical predictors. Vessel identifier was included as a categorical random effect. The precise command used to general the GAMM model was:

```
gm = gamm(logskj~te(lon,lat,by=quarter,k=13) + te(year,month,k=10,bs=c("c
r","cc")) +
           s(density,k=10) + s(capacity,k=10) + s(hours_since_sunrise) +
           country + follow,
         data=data,random=list(vessel_id=~1))
```

2-part GAMM model

There are several issues with the above mentioned single-component approach linked to the use of T3-corrected species composition data (Pianet et al. 2000). One is that this methodology has long been identified as not entirely satisfactory in terms of its treatment of spatial and temporal variability of catch composition (Duparc et al. 2018). The second is that there is no simple way to incorporate a measure of the uncertainty in species composition into the CPUE standardization process. As such, we propose, but have not yet implemented, an alternative approach to developing standardization models for FOB catch that has two components:

1) A first component that estimates total tropical tuna catch per FOB set. This component could be a GLMM or GAMM model of the types developed above and below.
2) A second component that estimates the proportion of catch per FOB set that is a given species (in the case of this report SKJ) based on port sampling of species composition. This component could be a beta regression GAMM model or a random forest model similar to that developed in Duparc et al. (2019).

There are several important potential advantages of this approach. One is that the response variables for both components are very close to the raw data found in captain logbooks and obtained by port sampling. The total tropical tuna catch per set is generally only slightly modified by post logbook corrections, with raising factors typically being close to 1. It is considered to be known with high certainty. The second component can be based directly on port sampling data and, therefore, the combined model will naturally incorporate uncertainty in species composition due to limited port sampling. Another advantage of this approach is that it naturally separates out the uncertainty due to species composition into one component, allowing us to develop optimal models for predicting species composition. As models for species composition improve, e.g., using the random forest approach of Duparc et al. (2019) or future Bayesian spatio-temporal auto-regressive models, these can be directly incorporated into the standardization process.

Though we had hoped to develop this two-component model for this year's CPUE standardization, ultimately this was not possible. We therefore propose it as a perspective for future standardization efforts.

GAMM prediction/standardization approaches

CPUE standardization is based on predicting models on a standard spatio-temporal grid, fixing fishing-efficiency- and catchability-related variables at standardized values, and then averaging over space (and potentially other predictors) to obtain a standardized estimation of abundance. We implemented two different approaches to this spatial averaging process for the 1-component GAMM model. The first is the approach that has traditionally been used based on predicting catch in each 1° × 1°-month strata occupied by the fishery and then averaging (or summing) over 1° × 1° grid cells. This spatial averaging is based on the assumption that set size is a true predictor of abundance in each strata. Though spatial thinning is generally used to remove cells with very low fishing effort from the prediction step, this method still has the disadvantage that it combines results from grid cells with potentially highly varying sampling effort (i.e., numbers of fishing sets). Furthermore, catch per set is only partially satisfactory as an estimator of abundance as it implicitly assumes that the number of FOB fish schools is constant over space (so that abundance is entirely reflected in set size), an assumption that is unlikely to be globally valid.

Due to these limitations, we also implement a second approach to developing a spatially-averaged standardized CPUE. In this approach, the predictions in each 1° × 1°-month strata are weighted by the total number of fishing sets over the entire time series of the data carried out in the given grid cell and the corresponding quarter (i.e., the weightings are stratified by 1° × 1°-quarter). As the number of sets times the average catch per set is the total catch, this approach is akin to using total catch as an indicator of abundance, except that the spatial distribution of fishing effort is standardized over time. This method will place more weight on core fishing areas where most fishing effort occurs relative to the previously described methodology.

Before implementing both standardization approaches, the spatial area to be used for predictions was thinned to remove 1° × 1° grid cells with little fishing effort. Predictions were only made for grid cells that collectively represent the smallest number of grid cells accounting for at least 95% of the FOB fishing sets in each quarter included in the model training data. The resulting modeling domains for each of the four quarters are shown in Figure 2.

Variables related to fishing efficiency and catchability were fixed at their median values from the training data set. Specifically, when calculating standardized CPUEs, dFAD density was fixed at 18.4 dFADs per 1° × 1° ocean grid cell at the equator, vessel capacity was fixed at 2119 m^3 and estimated hours since sunrise of the fishing set was fixed at 2.89 hours. Predictions were made for all levels of categorical predictor variables vessel country and whether or not a set was followed for each space-time strata and then averaged across levels, weighting the resulting predictions by the overall prevalence of each level in the model training data (e.g., fraction of followed versus non-followed sets).

Predictions from the log-normal GAMM model were converted back to absolute catch per set using the standard formula for estimating the expected value of a log-normal distribution (Fletcher 2008):

$$\mu_Y = \exp\left(\mu_X + \frac{\sigma_X^2}{2}\right) \qquad (1)$$

where μ_X is the expected value predicted by the GAMM model on the log-transformed data, σ_X^2 is the residual variance of the GAMM model (i.e., the `scale` parameter of the model outputs) and μ_Y is the final predicted catch per FOB set.

When averaging GAMM model predictions to obtain annual standardized CPUEs, standard errors were combined via simple addition, equivalent to assuming that all uncertainties in model predictions are correlated. Though undoubtedly inexact, this assumption will lead to conservative estimates of uncertainty (i.e., somewhat larger than reality). This issue can be corrected to obtain more exact uncertainty estimates using a bootstrap approach based on the Cholesky trick (Andersen 2022), but there was insufficient time to do so before the WPTT data prep meeting. We hope to implement this correction at a later date.

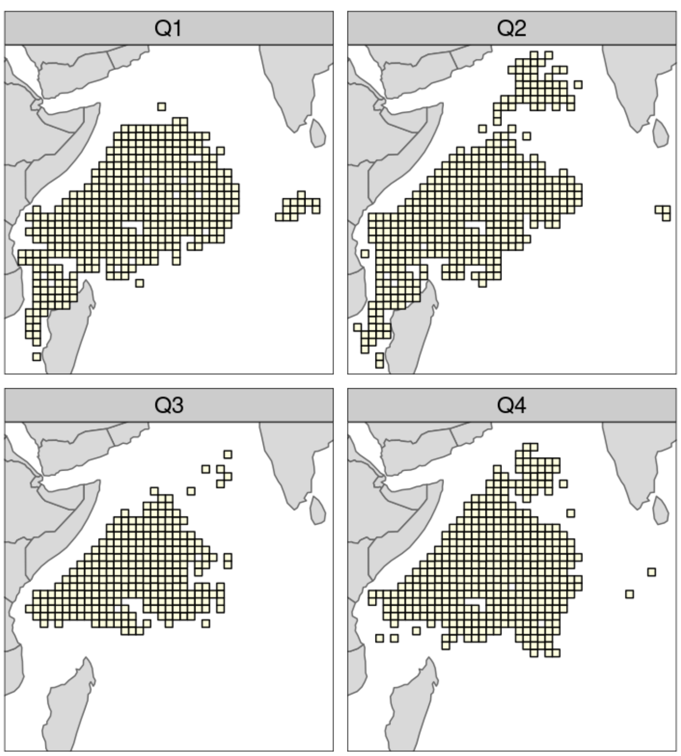

Figure 2: The $1^{\circ} \times 1^{\circ}$ grid cells used for model prediction for each quarter. The quarter number is indicated at the top of each panel.

GLMM model

GLMM model development follows that already presented in Akia et al. (2022). The selection of explanatory variables was conditional to: a) the F significance test, and b) the relative percentage of deviance explained by adding the factor in evaluation (factors explaining more than 1% were selected). Once the variables were selected, the final model included the variable "vessel" and the interaction year:quarter as random effects to obtain the estimated index per year using a GLMM.

Results

GAMM

Model diagnostics and significance of predictor variables

GAMMs are actually implemented in the R package mgcv as the combination of a linear mixed-effects (LME) model for estimating the random effect and a GAM model for estimating the final model with smooths after removing the variance explained by the random effect. Both of these components provide standard diagnostic plots, including a residuals-versus-fitted plot for the LME model (Figure 3) and a QQ-plot for the GAM (Figure 4). Both of these plots indicate an adequate fit of the data to the model assumptions.

All predictors included in the model, including smoothed, direct and random effects, had a significant impact on SKJ catch per FOB set (see model summaries below and Table 1).

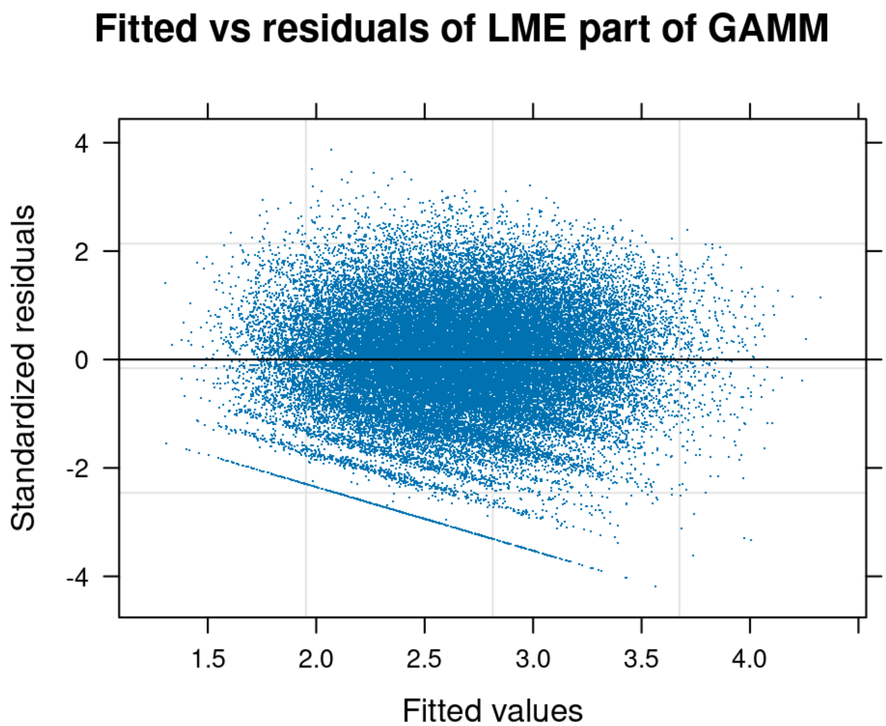

Figure 3: Fitted values versus residuals for LME part (i.e., random part) of GAMM.

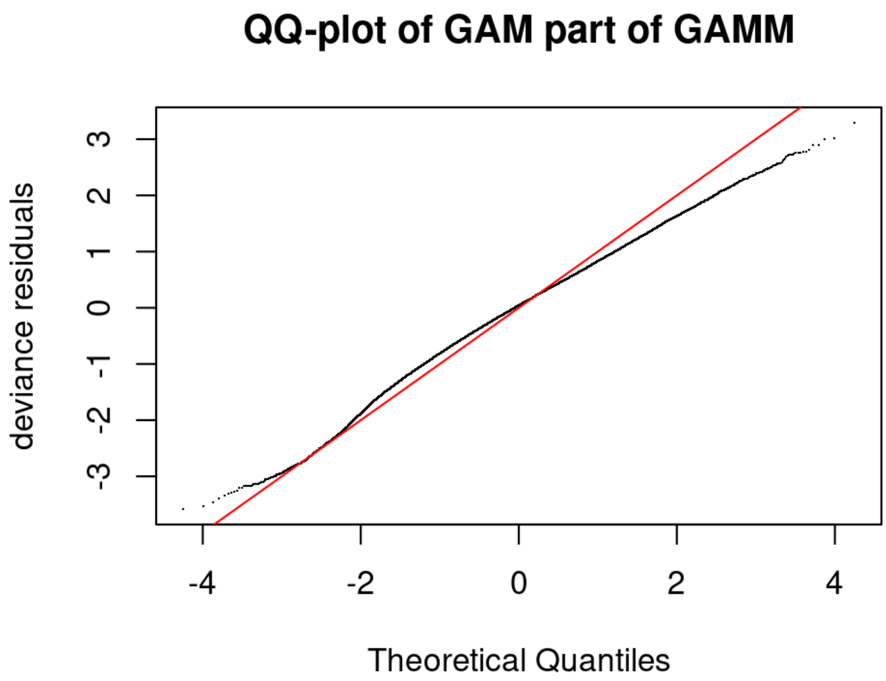

Figure 4: QQ-plot of GAM part (i.e., non-random part) of GAMM.

ANOVA table for LME component of GAMM model (i.e., model for estimating random effect):

```
  numDF denDF  F-value p-value
X    19 46608 1375.267  <.0001
```

Summary output from GAM part of GAMM model (i.e., non-random part of model):

```
Family: gaussian
Link function: identity

Formula:
logskj ~ te(lon, lat, by = quarter, k = 13) + te(year, month,
    k = 10, bs = c("cr", "cc")) + s(density, k = 10) + s(capacity,
    k = 10) + s(hours_since_sunrise) + country + follow

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.57412    0.03007  85.597  < 2e-16 ***
countryspain  0.14555    0.03984   3.653 0.000259 ***
followTRUE    0.03976    0.00858   4.634  3.6e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
                           edf Ref.df       F  p-value    
te(lon,lat):quarter1   42.507 42.507   8.705  < 2e-16 ***
te(lon,lat):quarter2   37.169 37.169   9.648  < 2e-16 ***
te(lon,lat):quarter3   27.324 27.324  15.901  < 2e-16 ***
te(lon,lat):quarter4   31.374 31.374  17.486  < 2e-16 ***
te(year,month)         79.504 79.504  38.824  < 2e-16 ***
s(density)              4.373  4.373  20.515  < 2e-16 ***
s(capacity)             2.220  2.220  12.762 4.86e-06 ***
s(hours_since_sunrise)  7.035  7.035 320.335  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.19   
  Scale est. = 0.72304   n = 46660
```

(a) Parametric terms

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.574	0.030	85.597	0
countryspain	0.146	0.040	3.653	0
followTRUE	0.040	0.009	4.634	0

(b) Smoothed terms

	Edf	Ref.df	F	p-value
te(lon,lat):quarter1	42.507	42.507	8.705	0
te(lon,lat):quarter2	37.169	37.169	9.648	0
te(lon,lat):quarter3	27.324	27.324	15.901	0
te(lon,lat):quarter4	31.374	31.374	17.486	0
te(year,month)	79.504	79.504	38.824	0
s(density)	4.373	4.373	20.515	0
s(capacity)	2.220	2.220	12.762	0
s(hours_since_sunrise)	7.035	7.035	320.335	0

Table 1: Summary statistics and p-values for fixed and smooth terms included in the non-random part of the GAMM model.

Marginal effects of predictor variables

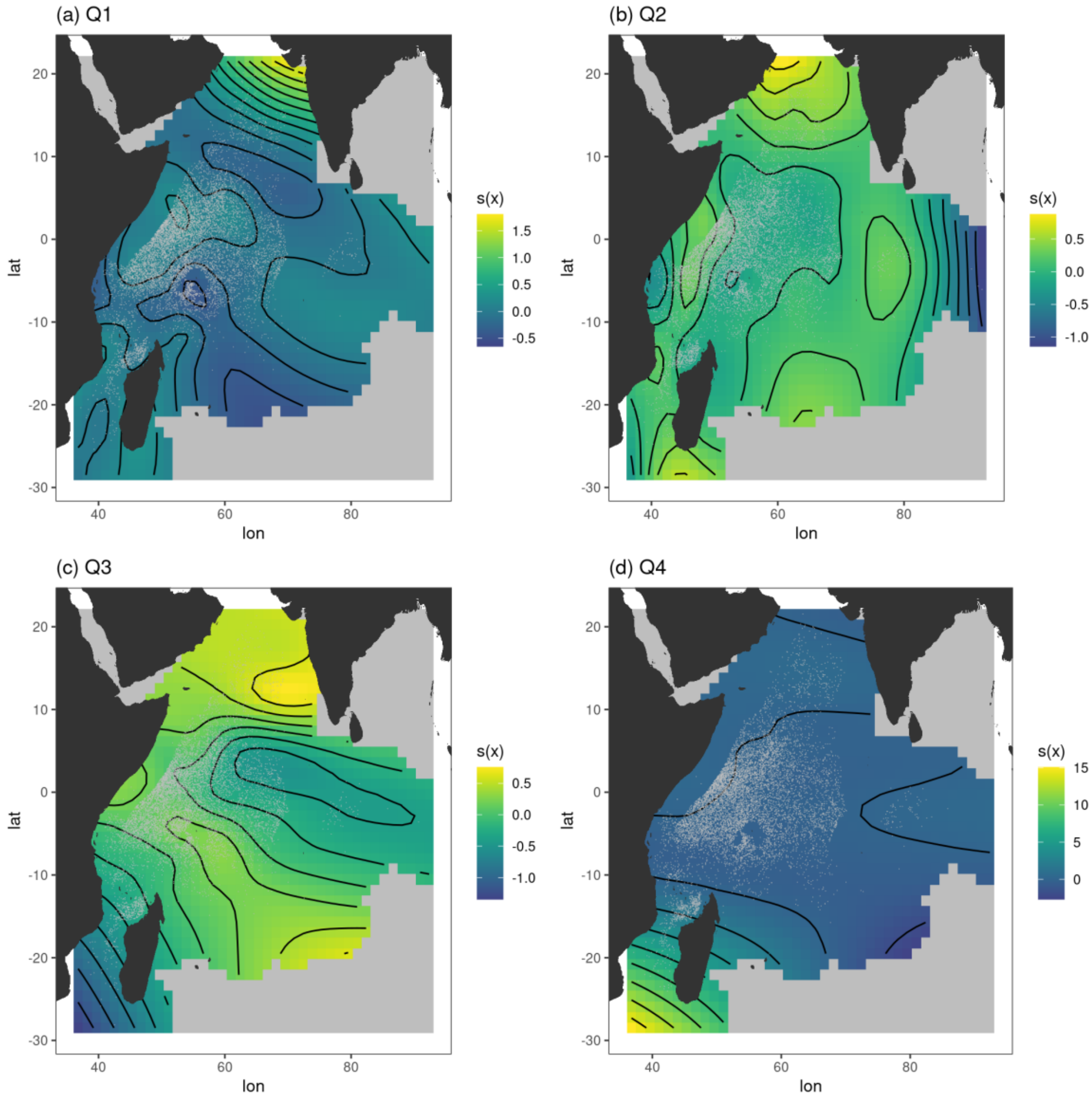

Figure 5: Marginal effect of lon,lat on log SKJ catch per FOB set for each of the four quarters.

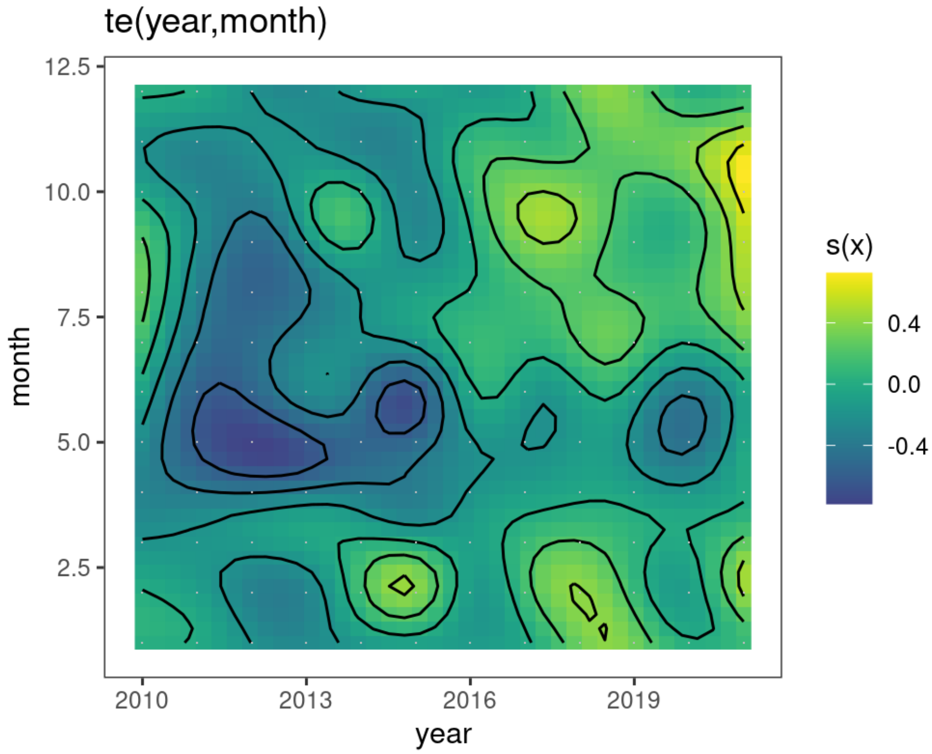

Figure 6: Marginal effect of year,month on log SKJ catch per FOB set.

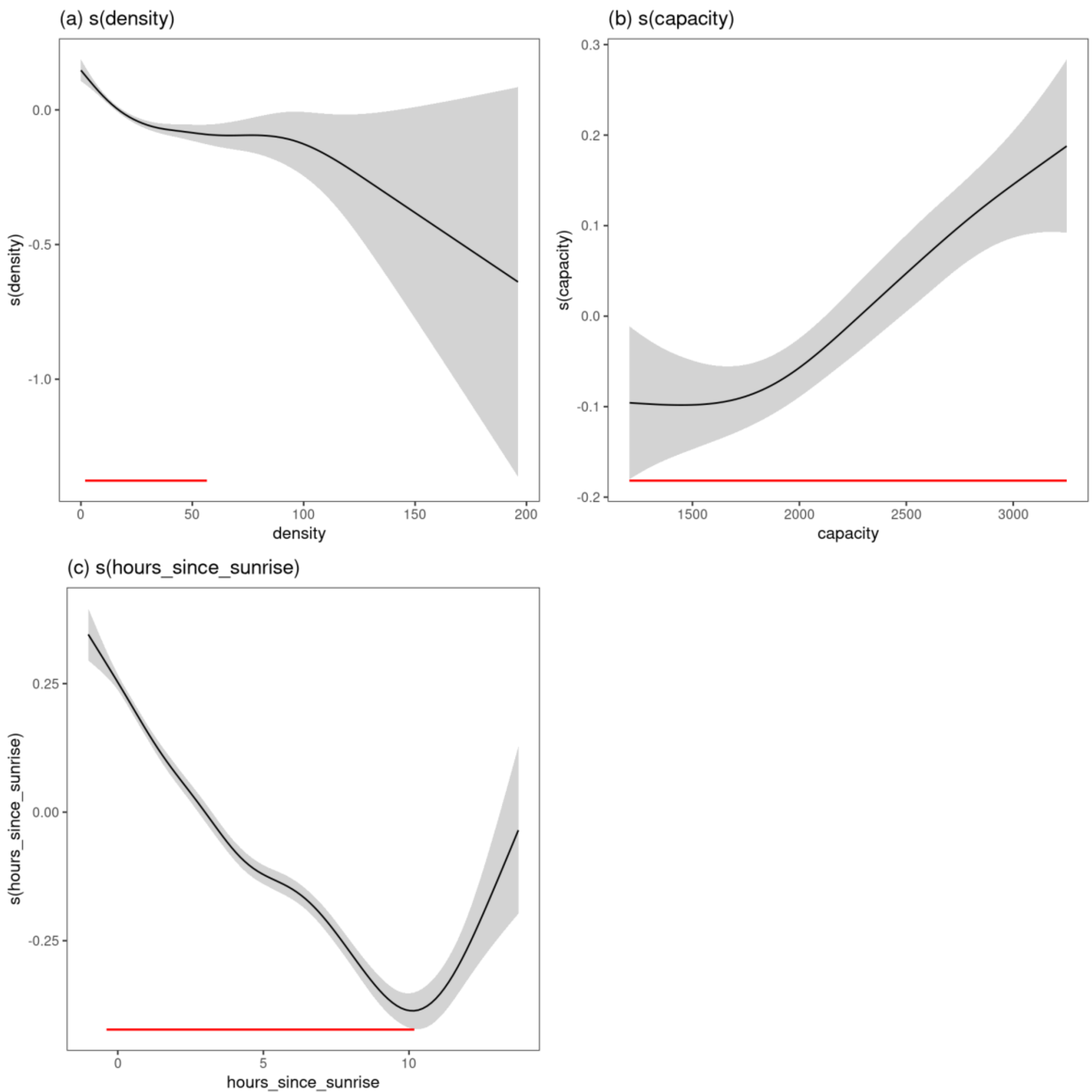

Figure 7: Marginal effects of individual smooths on log SKJ catch per FOB set. The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.

Vessel carrying capacity had an increasing, approximately linear impact on log SKJ catch per set, whereas dFAD density and set time hours since sunrise both had decreasing effects (Figure 7; note the different scales on the panels). The impacts of spatial (Figure5) and temporal (Figure6) predictors on log SKJ catch are more difficult to interpret.

Standardized CPUEs

Table 2: Annual spatially weighted and unweighted standardized CPUEs and nominal CPUEs for SKJ catch per FOB set in the Indian Ocean European purse seine fleet. Values are in

Year	Unweighted, Mean	Unweighted, 2.5%	Unweighted, 97.5%	Weighted, Mean	Weighted, 2.5%	Weighted, 97.5%
2010	18.03	14.71	22.08	18.97	15.72	22.88
2011	14.17	11.69	17.15	14.44	12.09	17.21
2012	12.68	10.81	14.86	12.76	11.07	14.69
2013	14.14	12.17	16.42	14.35	12.58	16.37
2014	15.81	13.60	18.37	16.27	14.24	18.57
2015	15.80	13.54	18.42	16.21	14.13	18.59
2016	17.66	15.26	20.42	18.12	15.94	20.58
2017	20.15	17.48	23.23	20.95	18.49	23.73
2018	23.03	20.09	26.41	23.78	21.10	26.80
2019	21.08	18.32	24.26	21.72	19.20	24.57
2020	18.54	15.70	21.88	19.36	16.62	22.55
2021	25.22	21.80	29.17	26.42	23.21	30.08

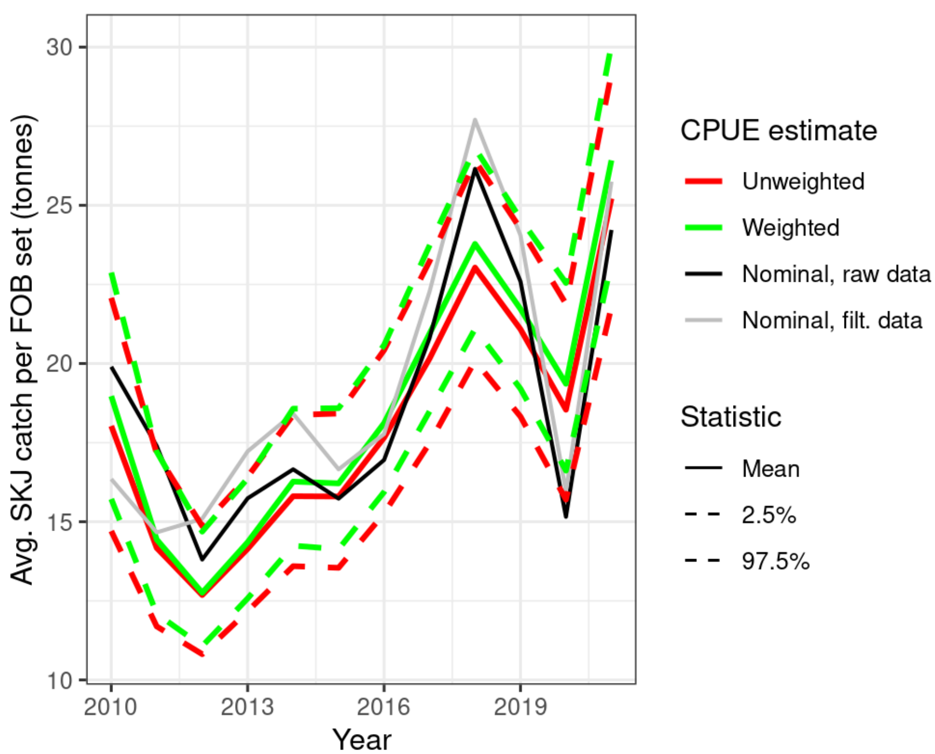

Figure 8: Standardized CPUE predictions from the single-component GAMM model. CPUEs are in units of tonnes of SKJ catch per PS FOB set in the Indian Ocean. Solid curves indicate mean tendencies, whereas dashed curves indicate the upper and lower limits of the 95% confidence interval. Red curves correspond to the spatially unweighted approach to averaging predictions over space, whereas green curves correspond to the spatially weighted approach to spatial averaging. Black and gray curves indicate the nominal CPUE derived from the original, unfiltered data and the filtered data used for training the GAMM model, respectively.

Nominal and standardized CPUE curves are shown in Figure 8 and Table 2. The weighted and unweighted standardized CPUE curves are generally similar to each other. The most notable differences between nominal and standardized CPUEs are reduced variability in standardized CPUEs over the period 2018-2021, approximately the period covered by the YFT quota and COVID. Nominal CPUEs based on the original, unfiltered data and the filtered data used for GAMM model training differ most notably in the early part of the time series, likely due to the exclusion of Spanish data due to partial VMS coverage in this time period needed for estimating set time.

GLMM

Analysis of Deviance Table [GLMM Lognormal]

Table 3: Analysis of deviance for the GLMM Lognormal

	Variable	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)	Dev..Exp
NULL	NA	NA	49611	61466	NA	NA	NA
year	11	1891	49600	59575	163	0	3.08 %
quarter	3	576	49597	58999	183	0	0.94 %
area	61	1184	49536	57815	18	0	1.93 %
pays	1	1123	49535	56692	1068	0	1.83 %
den_water	1	150	49534	56542	142	0	0.24 %
h_sunrise	1	2535	49533	54007	2411	0	4.12 %
numbat	32	925	49501	53082	27	0	1.5 %
capacity	0	0	49501	53082	NA	NA	0 %
follow	1	19	49500	53063	18	0	0.03 %
year:quarter	33	645	49467	52418	19	0	1.05 %
year:area	370	789	49097	51629	2	0	1.28 %

The most significant explanatory factors were incorporated in the following lognormal model:

log(catch) ~ year+quarter+area+pays+h_sunrise+(1|numbat)+(1|year:quarter)+(1|year:area).

The proportion of deviance explained by the model was 16%.

Diagnostic PLOTS [GLMM Lognormal]

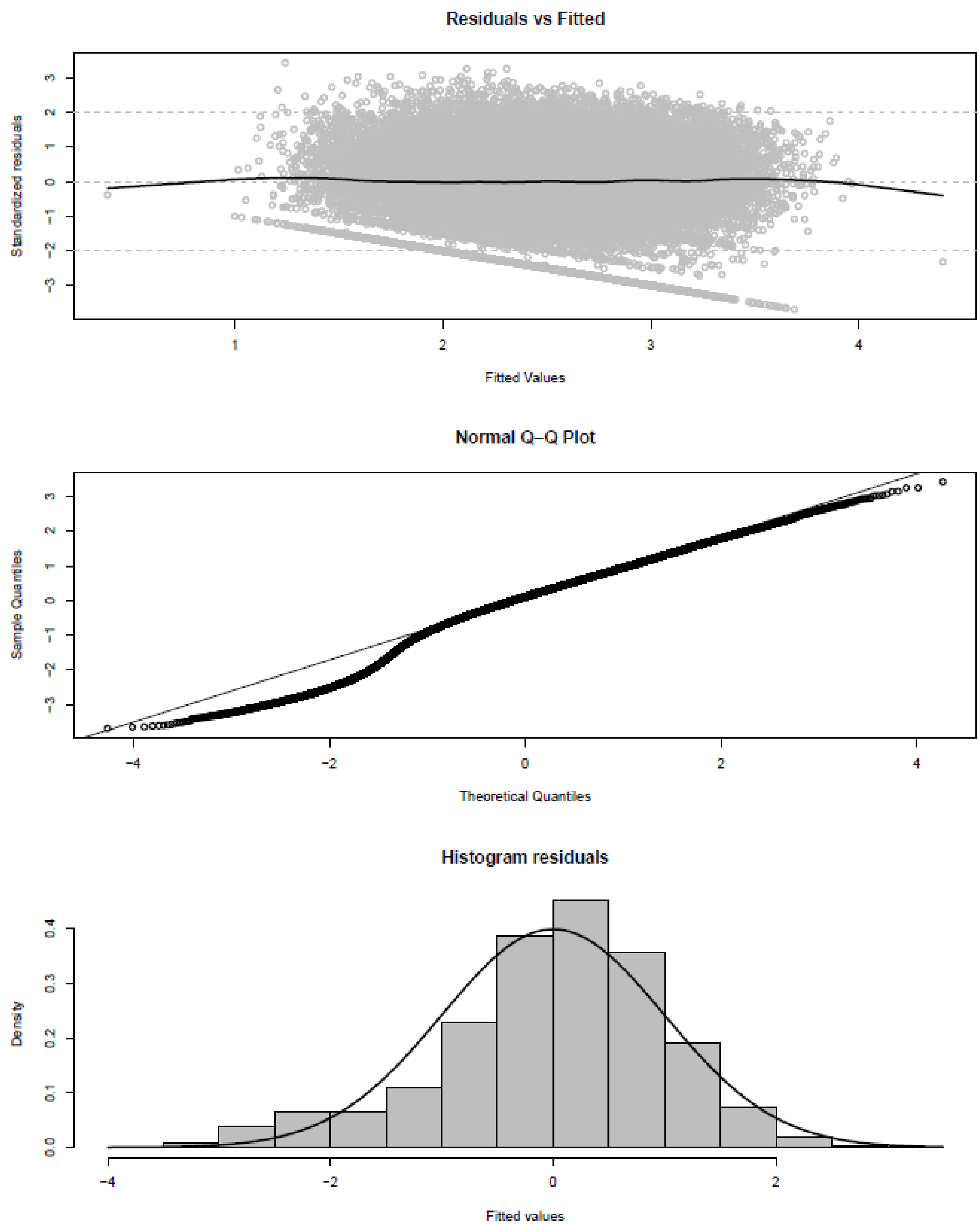

Figure 10. Diagnostic plots for GLMM

CPUE index [GLMM Lognormal]

Table 4: Standardized CPUEs and nominal CPUEs for SKJ catch per FOB set in the Indian Ocean European purse seine fleet from the GLMM.

Year	Nominal	Standardized	se	cv
2010	9.747	13.080	1.009	0.077
2011	8.601	11.277	0.888	0.079
2012	9.175	8.546	0.554	0.065
2013	10.427	10.060	0.624	0.062
2014	10.256	10.439	0.647	0.062
2015	9.980	10.400	0.640	0.062
2016	11.536	11.836	0.692	0.058
2017	12.996	13.160	0.785	0.060
2018	15.984	16.010	0.905	0.056
2019	14.216	13.597	0.780	0.057
2020	10.389	13.938	0.981	0.070
2021	15.483	16.571	0.978	0.059

log(catch) ~
year+quarter+area+pays+h_sunrise+(1|numbat)+(1|year:quarter)+(1|year:area)

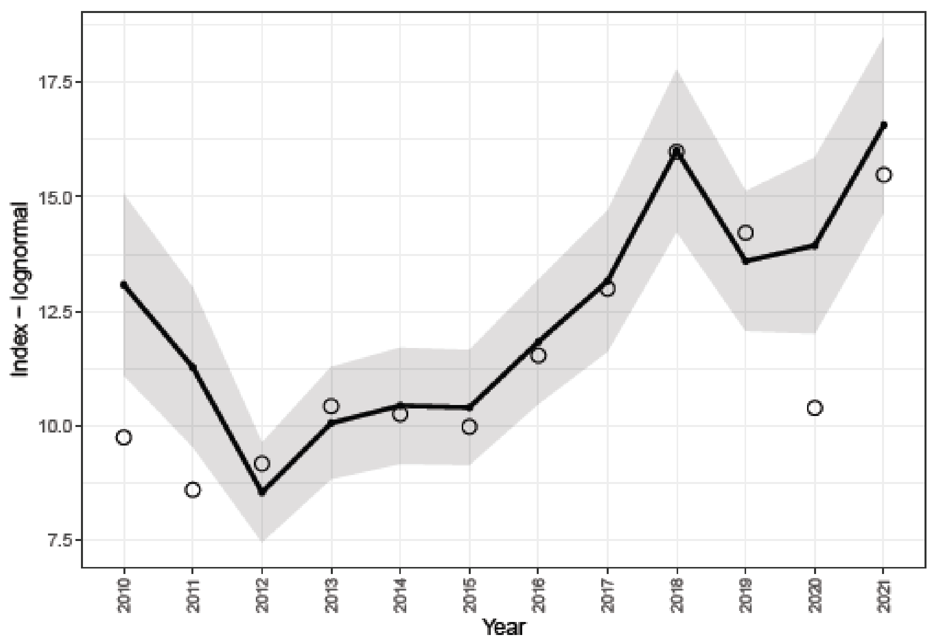

Figure 11: Standardized CPUE predictions from the GLMM model.

Discussion

There are a number of future improvements that could be made to the analyses carried out:

- Implementation of the 2-component model to better assess uncertainty in abundance estimates due to uncertainty in species composition.
- Inclusion of additional Spanish data in model training after obtaining complete VMS data so as to estimate set times.
- Improved estimation of covariance (or lack thereof) in model prediction uncertainties when predicting standardized CPUEs using a bootstrap approach based on the Cholesky trick.
- Inclusion of other predictors that might impact fishing efficiency and catchability, such as vessel age and mixed-layer depth.
- Inclusion of dFAD densities more directly related to fishing decision making and potential catch, such as the total number of dFADs within detection distance of the vessel and the number of followed dFADs within a reasonable travel distance from the set position.

Acknowledgements

We thank the IRD-Ob7 pelagic observatory of the MARBEC laboratory for French IO tropical tuna logbook and observer data management and preparation. We thank Orthongel, ANABAC and OPAGAC, the professional organizations representing the European tropical tuna purse-seine fishery, for facilitating access to European dFAD trajectory data.

References

Akia S, Guéry L, Grande M, Kaplan DM, Baéz JC, Ramos ML, Uranga J, Abascal FJ, Santiago J, Merino G, Gaertner D (2022) European purse seiners CPUE standardization of Big Eye tuna caught under dFADs. IOTC-2022-WPTT24-12. IOTC 24th Working Party on Tropical Tunas (WPTT) Data Preparatory Meeting, Online/Virtual

Andersen MM (2022) Prediction intervals for Generalized Additive Models (GAMs). *Mikkel Meyer Andersen*

Duparc A, Cauquil P, Depestris M, Dewals P, Gaertner D, Hervé A, Lebranchu J, Marsac F, Bach P (2018) Assessment of accuracy in processing purse seine tropical tuna catches with the T3 methodology using French fleet data. IOTC-2018-WPTT20-16_Rev1. IOTC Working Party on Tropical Tunas (WPTT), Victoria, Seychelles

Duparc A, Aragno V, Depetris M, Floch L, Cauquil P, Lebranchu J, Gaertner D, Marsac F, Bach P (2019) Assessment of the species composition of major tropical tunas in purse seine catches: A new modelling approach for the Tropical Tuna Treatment processing.

Case of the French fleet in Indian Ocean. IOTC-2019-WPTT21-10. IOTC Working Party on Tropical Tunas (WPTT21)

Fletcher D (2008) Confidence intervals for the mean of the delta-lognormal distribution. *Environmental and Ecological Statistics* **15**:175–189. doi:10.1007/s10651-007-0046-8

Pianet R, Pallares P, Petit C (2000) New Sampling and Data Processing Strategy for Estimating the Composition of Catches by Species and Sizes in the European Purse Seine Tropical Tuna Fisheries. WPDCS00-10. IOTC Proceedings no. 3

Wain G, Guéry L, Kaplan DM, Gaertner D (2021) Quantifying the increase in fishing efficiency due to the use of drifting FADs equipped with echosounders in tropical tuna purse seine fisheries. *ICES Journal of Marine Science* **78**:235–245. doi:10.1093/icesjms/fsaa216