IOTC-2023-WPTT25-08

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], Giancarlo Morón Correa[2], María Lourdes Ramos Alonso[3], José Carlos Báez[3], Jon Uranga[2], Antoine Duparc[1], Taha Imzilen[1], Laurent Floch[1] and Josu Santiago[2]

1. MARBEC (Univ. Montpellier, CNRS, Ifremer, IRD), av. Jean Monnet, CS30171, 34203 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

* *Correspondence:* David M. Kaplan david.kaplan@ird.fr

ABSTRACT

Abundance indices for Katsuwonus pelamis (SKJ) in the Indian Ocean were derived from the European purse seine 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 potential 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 with three distinct classes of tracking buoys: without an echosounder, with a one-frequency echosounder and with a two-frequency echosounder. Densities of instrumented buoys at the $1^\circ \times 1^\circ$-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. Results of both GLMMs and GAMMs indicate an initial decrease in standardized skipjack catch per set over the period 2010-2012 followed by an increase in standardized skipjack catch per over the period 2012-2021. This paper represents an update of that presented at the data preparatory meeting with the following major changes: division of followed buoys into categories based on echosounder technology, inclusion of Spanish data from 2010-2012 and inclusion in GAMM models of an interaction between year and follow/echosounder tracking buoy categories.

1 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, fishing on not-followed vs. followed (a.k.a. owned) dFADs and the number of frequencies of echosounder buoys for 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 GLMMs and GAMMs to standardize CPUEs
- Presentation of two theoretical approaches to developing models for SKJ CPUE standardization, one involving a single model for SKJ catch per set and the other combining two models, one for total catch and the other for species composition. Results are presented for the first of these, whereas the second 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, and in particular the last two involving different methodological approaches, 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 moved onto a new position. This methodology also provides fewer advantages when estimating a more simple catch per set model.

This paper represents an update of that presented at the WPTT data preparatory meeting in April 2023 with the following major changes:

- Division of followed buoys into categories based on echosounder technology
- Inclusion of Spanish data from 2010-2012 (previously excluded due to inability to estimate set time)
- Inclusion in GAMM models of an interaction between year and follow/echosounder tracking buoy categories

2 Methods

2.1 Catch-effort dataset

The catch-effort dataset in this study consisted of French and Spanish FOB sets over the period 2010-2021. Due to issues with port sampling data for the Spanish fleet from 2020, only catch-effort data for the French fleet were used for this year. The initial dataset consisted of 59,092 FOB sets corresponding to 59,062 logbook entries. The dataset was filtered 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, set time of day 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, either due to incomplete VMS data or closest VMS data being more than 20 km from the set position (2,209 sets)

After applying all of these filters, the final dataset used for building CPUE standardization models consisted of 53,285 sets.

2.2 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 $1^\circ \times 1^\circ$-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 $1^\circ \times 1^\circ$ grid cell at the equator) so final densities have units of average number of EU dFADs per $1^\circ \times 1^\circ$ 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. Estimated set times were compared 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).
- **Number of echosounder frequencies used by the tracking buoy for sets on followed dFADs:** For followed sets, the number of echosounder frequencies was determined based on tracking buoy model and converted into 3 classes: (i) buoys without echosounders (i.e., zero frequencies), (ii) single frequency echosounder buoys, and (iii) two or more echosounder frequencies. We decided to use a plus class for two or more frequencies as buoys with more than two frequencies are extremely rare.
- **Follow+echo:** The followed boolean variable and the number of echosounder frequencies categorical variable were combined into a single categorical variable with four levels, the three levels of the echosounder frequency variable plus a level corresponding to dFADs that are not followed. This was the final categorical variable used in models, the breakdown of which by level is shown in Table 1.

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,

Table 1: The number of sets in each of the follow+echo categories. Note that class `echo_2freq` is a plus class used for all tracking buoys with two or more frequencies.

Set category	No. sets	% sets
no follow	27435	51.5
no echo	902	1.7
echo_1freq	20197	37.9
echo_2freq	4751	8.9

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.

2.3 Modeling approaches

Three different CPUE standardization modeling approaches were considered, two of which were carried out (the one-part GAMM and GLMM) and the third of which (the two-part model) 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. We also present two different methodologies for generating standardized predictions from GAMM models.

2.3.1 One-part GAMM model

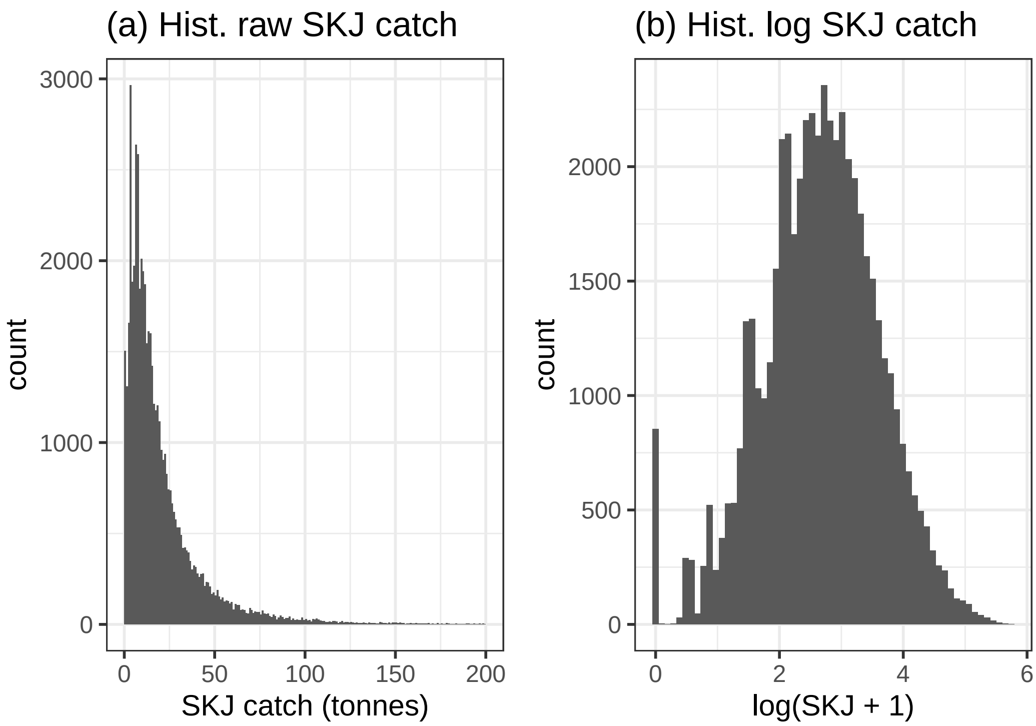


Figure 1: Histograms of SKJ catch data in model training dataset before (a) and afer (b) log transformation.

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 (852 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 acceptable 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, including an interaction between year and follow+echo, density (in units of mean daily number of buoys per $1^{\circ} \times 1^{\circ}$ water grid cell at the equator), vessel capacity (in units of m^3) and hours since sunrise as individuals smooths, and vessel country and follow+echo as non-smoothed, categorical predictors. Vessel identifier was included as a categorical random effect. The precise command used to generate the GAMM model was:

```
gm = gamm(logskj~te(lon,lat,by=quarter,k=13) +
            s(year,by=follow_echo,k=10,bs="cr") +
            s(month,k=10,bs="cc") +
            ti(year,month,k=10,bs=c("cr","cc")) +
            s(density,k=12) + s(capacity,k=10) + s(hours_since_sunrise) +
            country + follow_echo,
          data=data,random=list(vessel_id=~1))
```

2.3.2 Two-part model

There are several issues with the above mentioned approaches 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 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 type developed above.
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). 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.

2.3.3 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. The first is the approach that has traditionally been used based on predicting catch in each $1^{\circ} \times 1^{\circ}$-month strata occupied by the fishery and then averaging (or summing) over $1^{\circ} \times 1^{\circ}$ 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 $1^{\circ} \times 1^{\circ}$-month strata are weighted by the total number of fishing sets carried out in that grid cell and the corresponding quarter (i.e., the weightings are stratified by grid cell-quarter) over the entire time series of the data. 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 $1^{\circ} \times 1^{\circ}$ 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 16.3, vessel capacity was fixed at 2119 and estimated hours since sunrise of the fishing set was fixed at 5.12. Predictions were made for all levels of categorical predictor variables vessel country and follow+echo 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 Spanish versus French sets).

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

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

where μ_X is the expected value predicted by the GAMM model, σ_X^2 is the residual variance of the GAMM model (i.e., the `scale` parameter of the GAM summary outputs plus the variance explained of the vessel identifier random effect LME model, which was very small compared to the unexplained variance of the GAM model) and μ_Y is the final predicted catch.

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., 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 it was decided that it was best to use a more conservative approach.

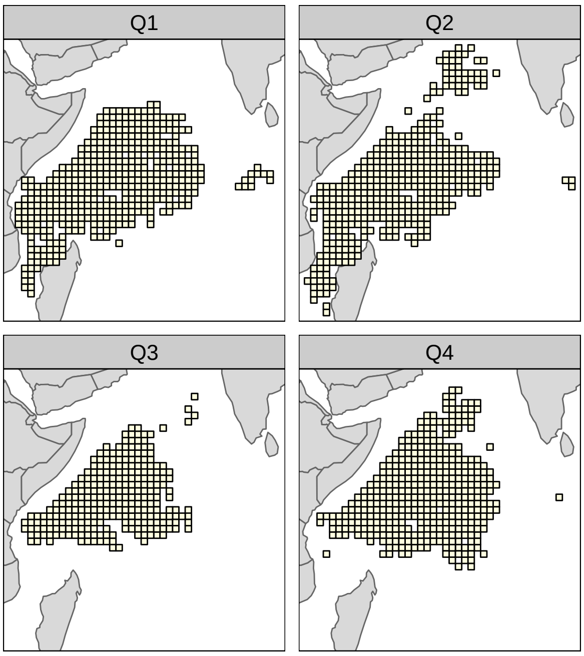


Figure 2: The $1^\circ \times 1^\circ$ grid cells used for model prediction for each quarter. The quarter number is indicated at the top of each panel.

2.3.4 GLMM model

GLMM model development follows that already presented in Akia et al. (2022) with the exception that the new predictor variables described above have been included in the model. 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.

3 Results

3.1 GAMM

3.1.1 Model diagnostics and significance of predictor variables

GAMM models are actually implemented 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 2).

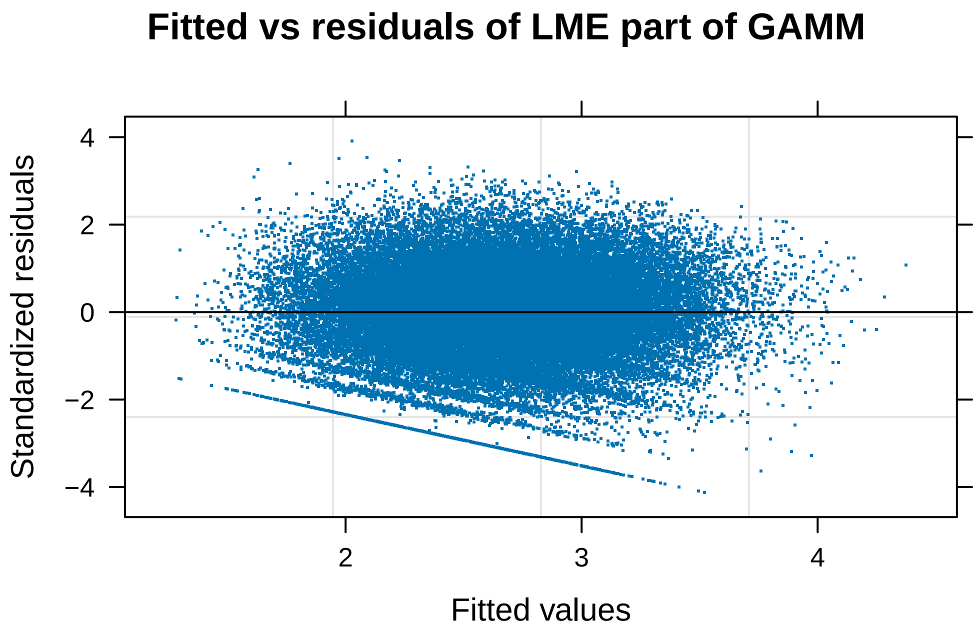


Figure 3: Fitted values versus residuals for LME part (i.e., random part) of GAMM.

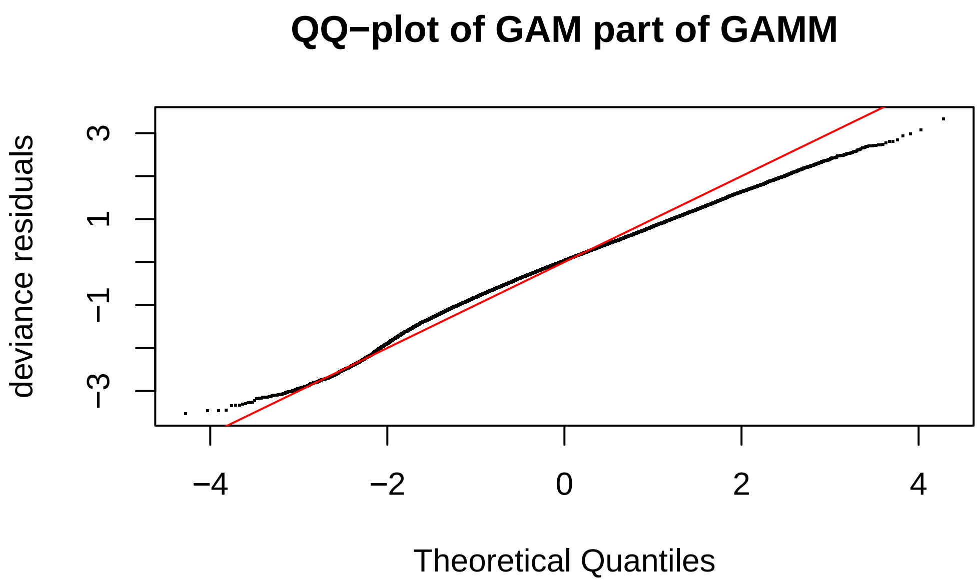


Figure 4: QQ-plot of GAM part (i.e., non-random part) of GAMM.

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

```
  numDF denDF  F-value p-value
X    24 53227 1450.684  <.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) + s(year, by = follow_echo,
    k = 10, bs = "cr") + s(month, k = 10, bs = "cc") + ti(year,
    month, k = 10, bs = c("cr", "cc")) + s(density, k = 12) +
    s(capacity, k = 10) + s(hours_since_sunrise) + country +
    follow_echo

Parametric coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.60522    0.02552 102.098  < 2e-16 ***
countryspain           0.13489    0.03135   4.302 1.70e-05 ***
follow_echono echo    -0.51069    0.11505  -4.439 9.06e-06 ***
follow_echoecho_1freq  0.05005    0.01123   4.455 8.41e-06 ***
follow_echoecho_2freq -0.09884    0.12449  -0.794    0.427
---
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        45.866 45.866  11.50  <2e-16 ***
te(lon,lat):quarter2        36.056 36.056  10.58  <2e-16 ***
te(lon,lat):quarter3        29.641 29.641  16.59  <2e-16 ***
te(lon,lat):quarter4        39.487 39.487  20.18  <2e-16 ***
s(year):follow_echono follow  8.663  8.663 111.69  <2e-16 ***
s(year):follow_echono echo     1.000  1.000   6.53  0.0106 *
s(year):follow_echoecho_1freq  8.380  8.380  97.63  <2e-16 ***
s(year):follow_echoecho_2freq  5.447  5.447  22.98  <2e-16 ***
s(month)                     7.661  8.000  90.34  <2e-16 ***
ti(year,month)              65.020 72.000  14.21  <2e-16 ***
s(density)                   5.836  5.836  19.06  <2e-16 ***
```

Table 2: Summary statistics and p-values for fixed and smooth terms included in the non-random part of the GAMM model.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.605	0.026	102.098	0.000	***
countryspain	0.135	0.031	4.302	0.000	***
follow_echono echo	-0.511	0.115	-4.439	0.000	***
follow_echoecho_-1freq	0.050	0.011	4.455	0.000	***
follow_echoecho_-2freq	-0.099	0.124	-0.794	0.427	

	edf	Ref.df	F	p-value	
te(lon,lat):quarter1	45.866	45.866	11.502	0.000	***
te(lon,lat):quarter2	36.056	36.056	10.576	0.000	***
te(lon,lat):quarter3	29.641	29.641	16.591	0.000	***
te(lon,lat):quarter4	39.487	39.487	20.181	0.000	***
s(year):follow_echono follow	8.663	8.663	111.692	0.000	***
s(year):follow_echono echo	1.000	1.000	6.530	0.011	*
s(year):follow_echoecho_1freq	8.380	8.380	97.634	0.000	***
s(year):follow_echoecho_2freq	5.447	5.447	22.980	0.000	***
s(month)	7.661	8.000	90.338	0.000	***
ti(year,month)	65.020	72.000	14.206	0.000	***
s(density)	5.836	5.836	19.063	0.000	***
s(capacity)	3.111	3.111	14.748	0.000	***
s(hours_since_sunrise)	4.549	4.549	592.017	0.000	***

```
s(capacity)                   3.111  3.111  14.75  <2e-16 ***
s(hours_since_sunrise)        4.549  4.549 592.02  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.188
  Scale est. = 0.72883   n = 53285
```

Checking to see if the basis dimensions chosen for the smooth effects are sufficient with the `gam.check` function of the `mgcv` package indicated that this is the case for all smooths (though, it should be noted that using the `gam.check` function with `gamm` models is generally discouraged; in this case, the variance explained by the random vessel effect is relatively small compared to the unexplained variance and therefore the results of `gam.check` should be reasonably reliable for our model).

3.1.2 Marginal effects of predictor variables

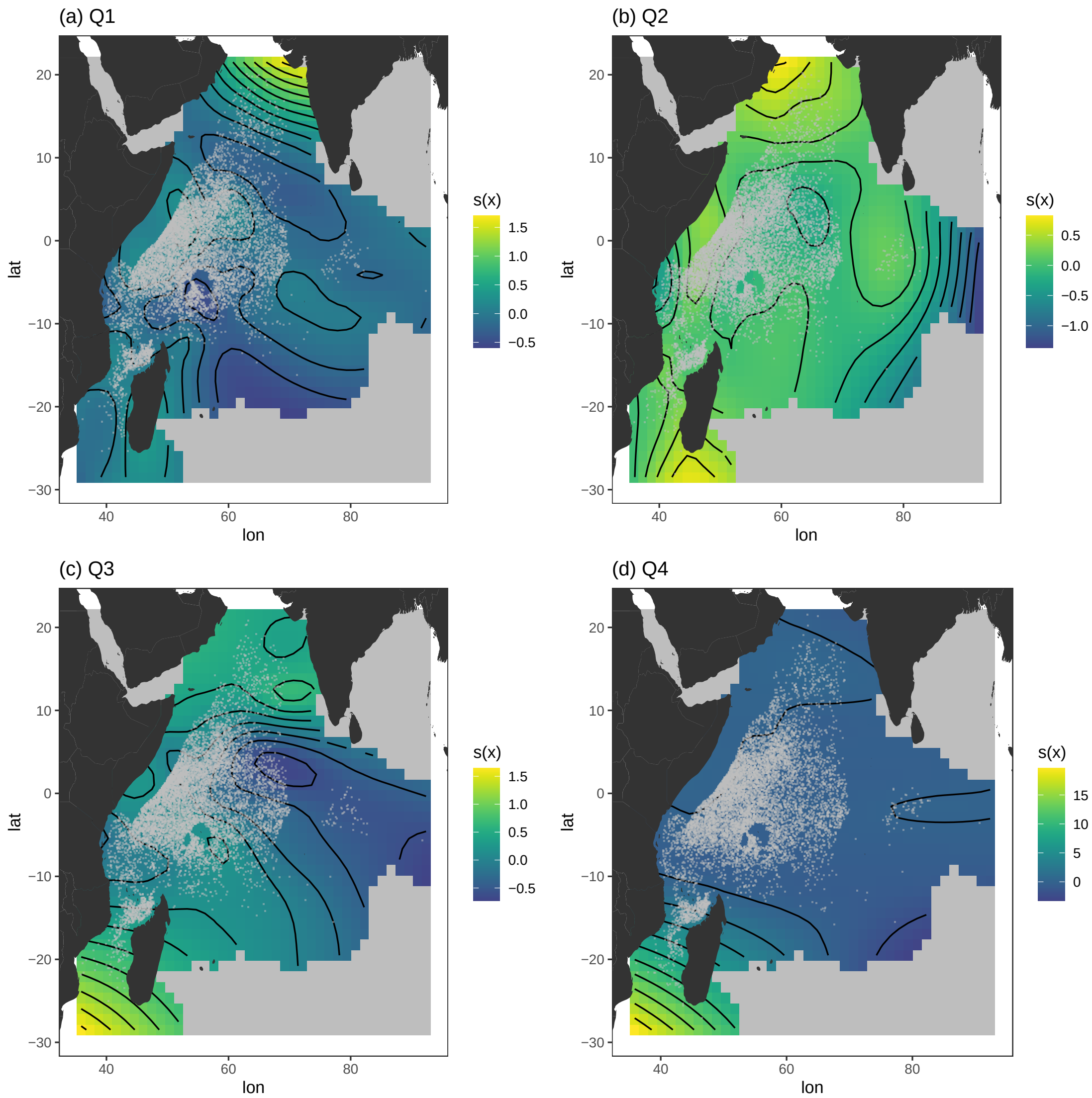


Figure 5: Marginal effect of lon,lat on log SKJ catch per FOB set for each of the four quarters.

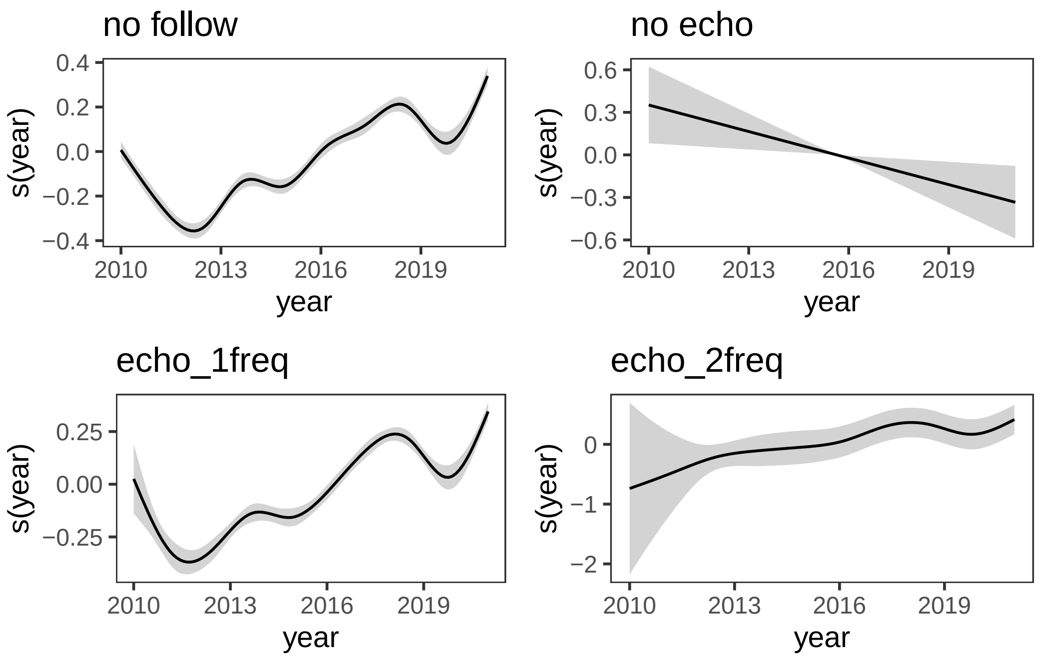


Figure 6: Marginal effect of year on log SKJ catch per FOB set by follow+echo category.

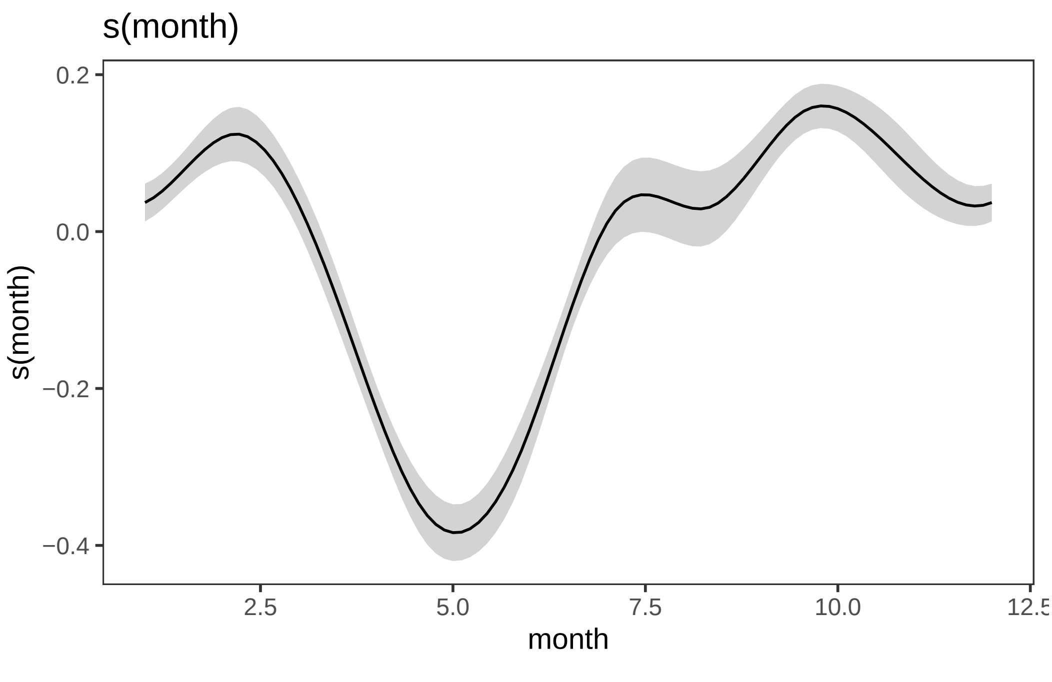


Figure 7: Marginal effect of month on log SKJ catch per FOB set.

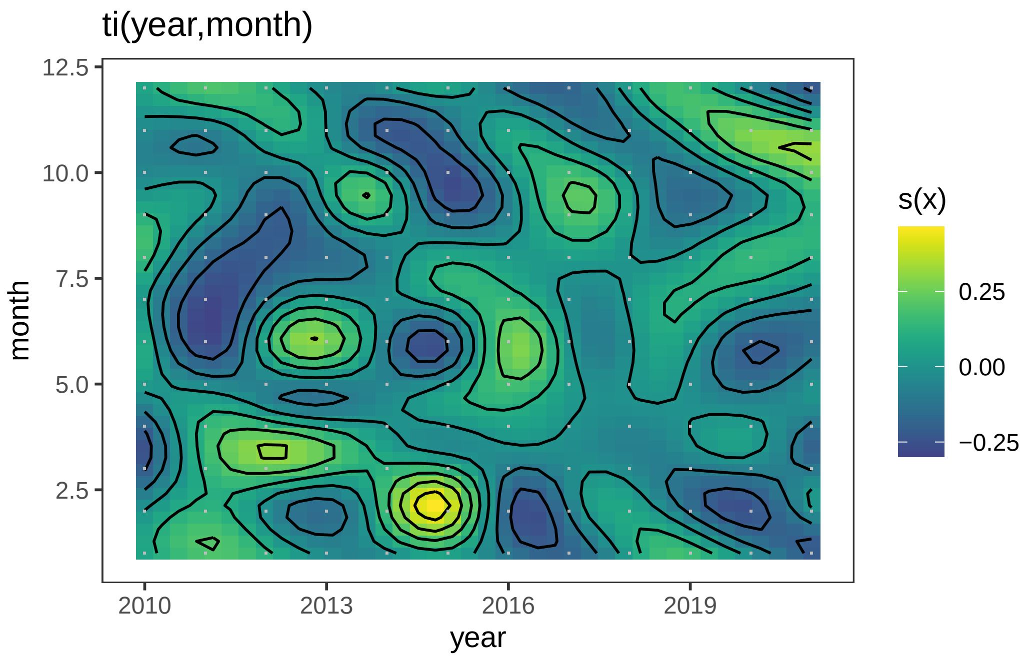


Figure 8: Tensor-interaction of year,month marginal effect on log SKJ catch per FOB set.

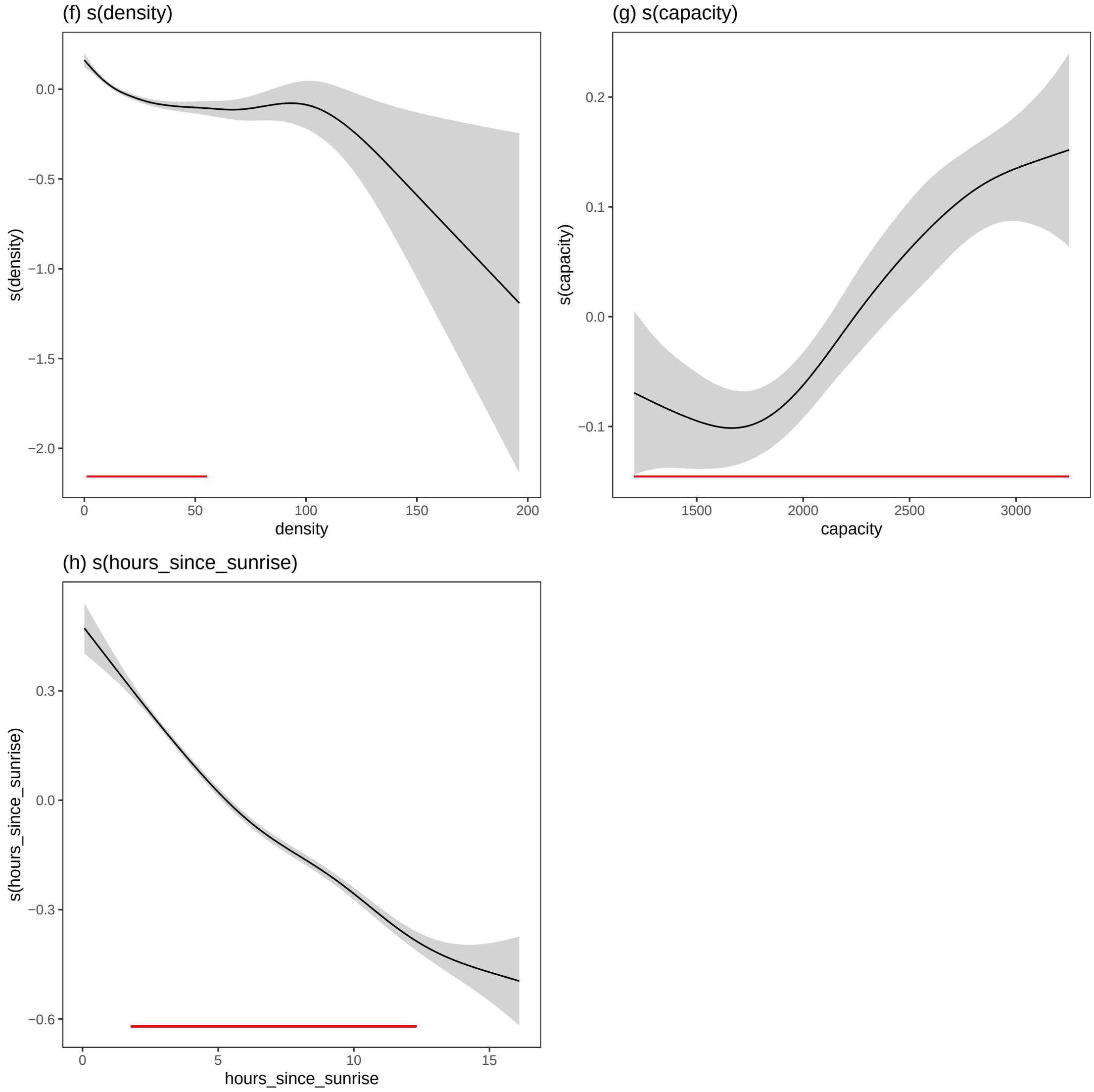


Figure 9: Marginal effects of fishing-efficiency-related individual smooths on log SKJ catch per FOB set. The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.

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 9; note the different scales on the panels). The impacts of spatial (Figure 5) and temporal (Figure 6, Figure 7 & Figure 8) predictors on log SKJ catch are more difficult to interpret.

Table 5: 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 units of tonnes per set.

Year	Unweighted, Mean	Unweighted, 2.5%	Unweighted, 97.5%	Weighted, Mean	Weighted, 2.5%	Weighted, 97.5%
2010	17.60	14.09	24.32	18.57	15.09	25.35
2011	14.03	11.54	17.89	14.38	12.02	18.07
2012	12.93	10.79	15.66	13.13	11.14	15.65
2013	14.47	12.42	16.86	14.87	12.98	17.03
2014	15.91	13.73	18.44	16.59	14.56	18.89
2015	16.08	13.76	18.79	16.67	14.49	19.17
2016	17.95	15.52	20.77	18.62	16.36	21.18
2017	20.13	17.49	23.18	21.19	18.71	23.99
2018	22.85	19.98	26.15	23.85	21.20	26.84
2019	21.22	18.47	24.39	22.10	19.56	25.00
2020	18.91	15.97	22.41	20.01	17.11	23.42
2021	25.47	22.01	29.48	26.95	23.65	30.72

3.1.3 Standardized CPUEs

Table 6: Quarterly 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 units of tonnes per set.

Year	Quarter	Unweighted, Mean	Unweighted, 2.5%	Unweighted, 97.5%	Weighted, Mean	Weighted, 2.5%	Weighted, 97.5%
2010	1	18.56	14.85	25.66	19.16	15.49	26.26
2010	2	14.45	11.37	20.28	14.67	11.73	20.30
2010	3	21.83	17.70	29.82	21.94	18.01	29.67
2010	4	16.95	13.65	23.32	17.39	14.15	23.71
2011	1	17.69	14.54	22.56	18.27	15.19	23.04
2011	2	10.93	8.81	14.21	11.10	9.10	14.19
2011	3	13.36	11.13	16.85	13.43	11.35	16.72
2011	4	13.78	11.44	17.42	14.14	11.89	17.68
2012	1	15.51	12.95	18.77	16.02	13.54	19.16
2012	2	11.29	9.23	13.92	11.45	9.54	13.89
2012	3	11.73	9.92	14.03	11.80	10.13	13.91
2012	4	12.69	10.69	15.23	13.02	11.11	15.45
2013	1	15.52	13.31	18.10	16.03	13.92	18.45
2013	2	13.46	11.34	15.97	13.66	11.71	15.92
2013	3	15.16	13.19	17.43	15.24	13.46	17.27
2013	4	13.91	12.06	16.04	14.27	12.54	16.25
2014	1	19.27	16.62	22.34	19.90	17.38	22.77
2014	2	12.65	10.66	14.99	12.84	11.02	14.94
2014	3	18.03	15.79	20.59	18.13	16.10	20.39
2014	4	14.25	12.42	16.35	14.62	12.91	16.56
2015	1	21.47	18.38	25.09	22.16	19.19	25.60
2015	2	12.14	10.13	14.53	12.32	10.46	14.50
2015	3	17.38	15.02	20.10	17.47	15.31	19.93
2015	4	13.49	11.70	15.54	13.84	12.15	15.75
2016	1	16.97	14.64	19.67	17.53	15.31	20.06
2016	2	17.88	15.18	21.05	18.14	15.66	21.00
2016	3	20.83	18.26	23.77	20.94	18.62	23.55
2016	4	17.13	14.95	19.62	17.57	15.53	19.89
2017	1	20.18	17.49	23.27	20.83	18.28	23.72
2017	2	17.44	14.93	20.38	17.70	15.42	20.31
2017	3	24.79	21.81	28.18	24.92	22.25	27.91
2017	4	19.71	17.21	22.57	20.21	17.86	22.88
2018	1	25.82	22.55	29.56	26.64	23.57	30.12
2018	2	18.75	16.13	21.81	19.02	16.65	21.73
2018	3	26.11	23.08	29.56	26.24	23.55	29.26
2018	4	21.72	19.12	24.68	22.28	19.86	24.99
2019	1	22.68	19.73	26.08	23.41	20.62	26.58
2019	2	17.74	15.12	20.83	18.00	15.59	20.78
2019	3	23.67	20.83	26.91	23.79	21.24	26.66
2019	4	21.61	19.00	24.58	22.16	19.74	24.89
2020	1	18.70	15.88	22.01	19.30	16.58	22.47
2020	2	14.61	12.04	17.71	14.82	12.40	17.71
2020	3	22.76	19.17	27.03	22.88	19.48	26.87
2020	4	21.02	17.97	24.60	21.56	18.63	24.96
2021	1	25.30	21.77	29.41	26.11	22.71	30.03
2021	2	19.69	16.82	23.06	19.97	17.36	22.99
2021	3	30.78	26.93	35.21	30.94	27.45	34.89
2021	4	28.05	24.33	32.35	28.76	25.25	32.78

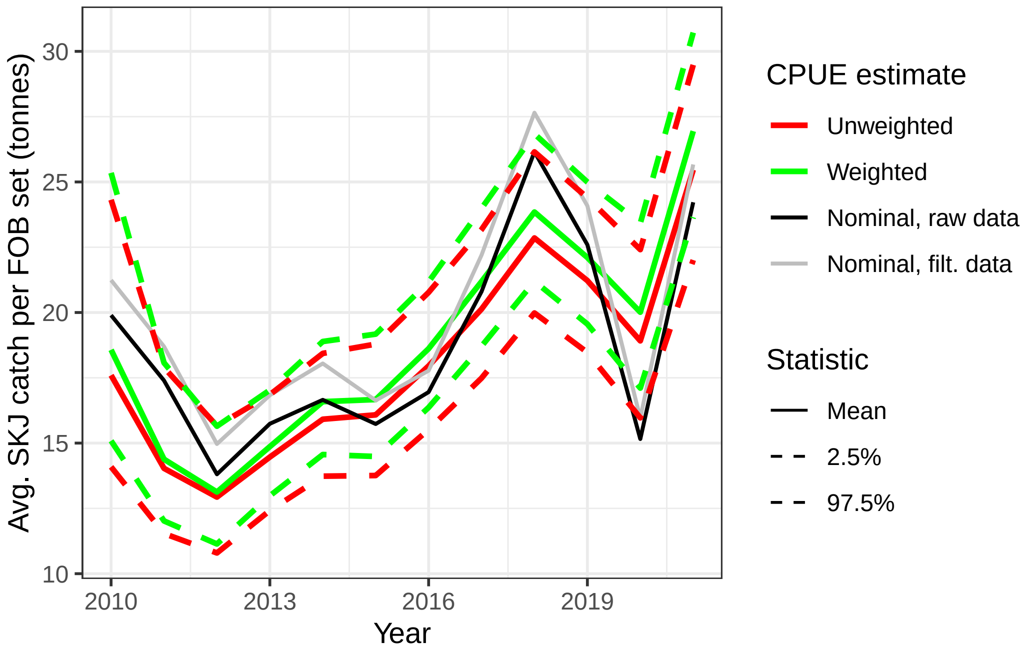


Figure 10: Yearly standardized CPUE predictions from the single-component GAMM model. CPUEs are in units of tonnes of SKJ catch per PS FOB set in the Indian Ocean. Solid curves indicate mean tendencies, whereas dashed curves indicate the upper and lower limits of the 95% confidence interval. Red curves 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.

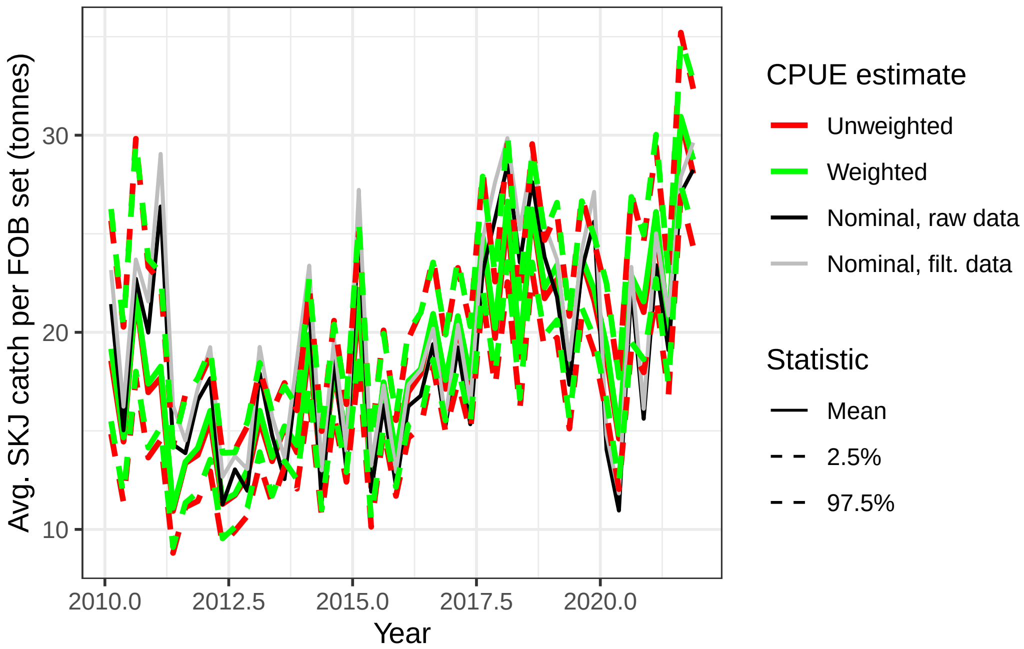
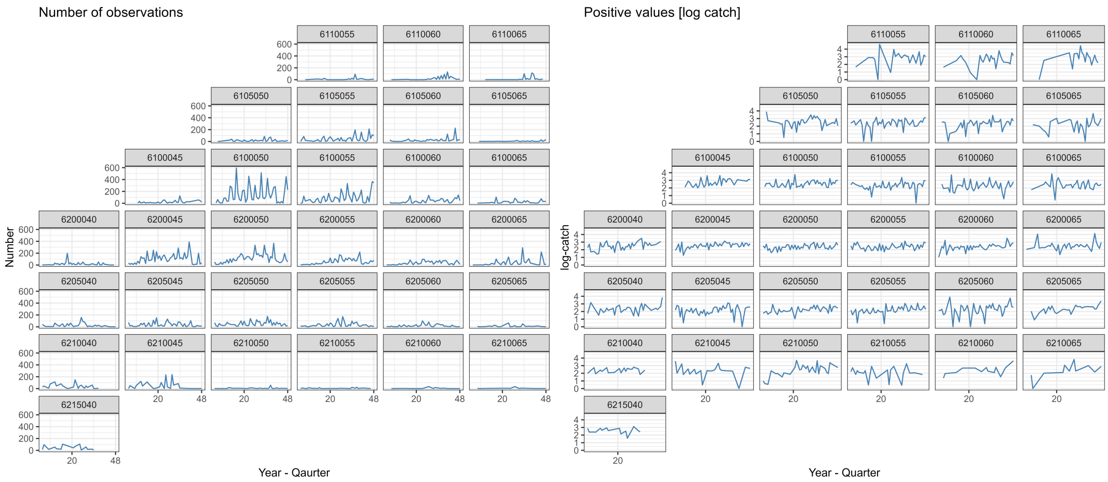


Figure 11: Quarterly standardized CPUE predictions from the single-component GAMM model. CPUEs are in units of tonnes of SKJ catch per PS FOB set in the Indian Ocean. Solid curves indicate mean tendencies, whereas dashed curves indicate the upper and lower limits of the 95% confidence interval. Red curves 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 by year in Figure 10 and Table 5, and by quarter in Figure 11 and Table 6. The weighted and unweighted standardized CPUE curves are generally similar to each other and similar to the nominal CPUE curves. 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.

3.2 GLMM



(a) (b)

Figure 12: The number of FOB sets as a function of $1^\circ \times 1^\circ$ grid cell and year-quarter in number (left panel) and on a log scale (right panel).

ANOVA results for the GLMM model indicate that all predictors had significant effects, though deviance explained by `density` and `follow_echo` were small (Table 7). The specific formula for the final model including random effects was (in R formula notation):

```
logskj ~ year:quarter + area + country + hours_since_sunrise +
  follow_echo + density + (1 | numbat) + (1 | (year:quarter:area))
```

The proportion of deviance explained by the final GLMM model was 19% and model diagnostics were satisfactory (Figure 13). Standardized CPUE values are shown in Figure 14 and Table 8.

Table 7: Analysis of deviance for the GLMM Lognormal.

	Variable	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)	Dev. Exp.
year:quarter	47	3143	49564	58323	65	0	5.11 %
area	61	1149	49503	57174	18	0	1.87 %
country	1	1136	49502	56037	1105	0	1.85 %
density	1	123	49501	55915	119	0	0.2 %
hours_since_sunrise	1	2531	49500	53384	2461	0	4.12 %
numbat	32	947	49468	52436	29	0	1.54 %
follow_echo	3	28	49465	52409	9	0	0.04 %
year:quarter:area	1094	2667	48371	49742	2	0	4.34 %

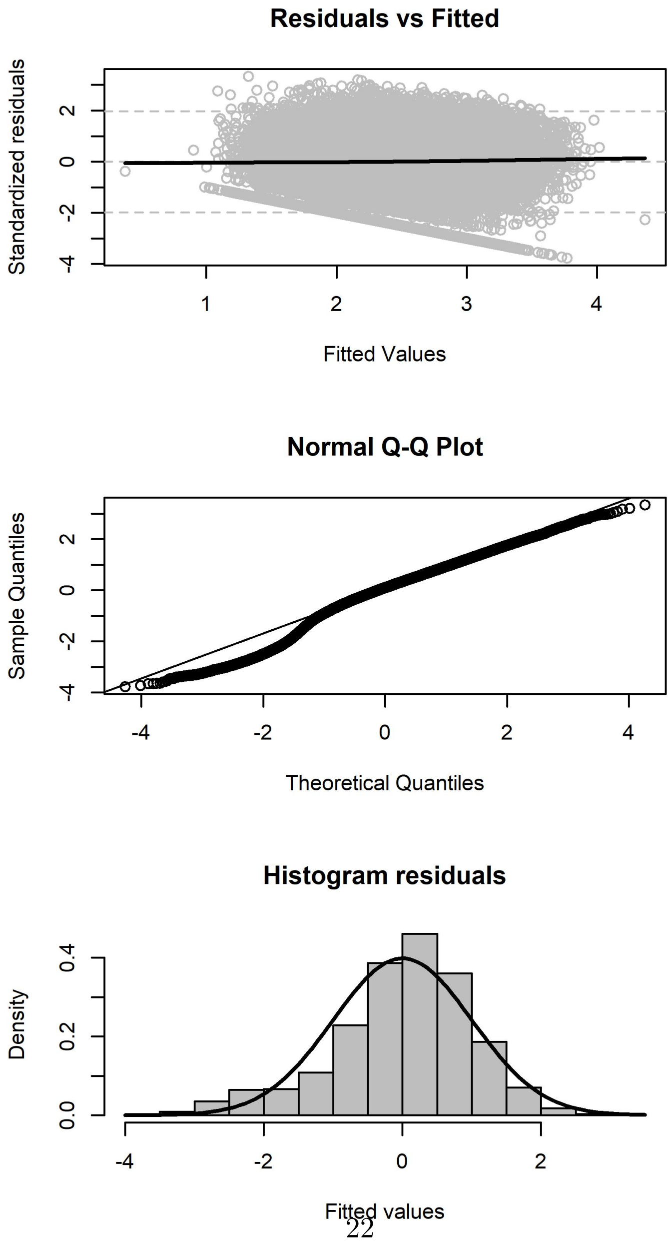


Figure 13: Standard diagnostic plots for the GLMM model.

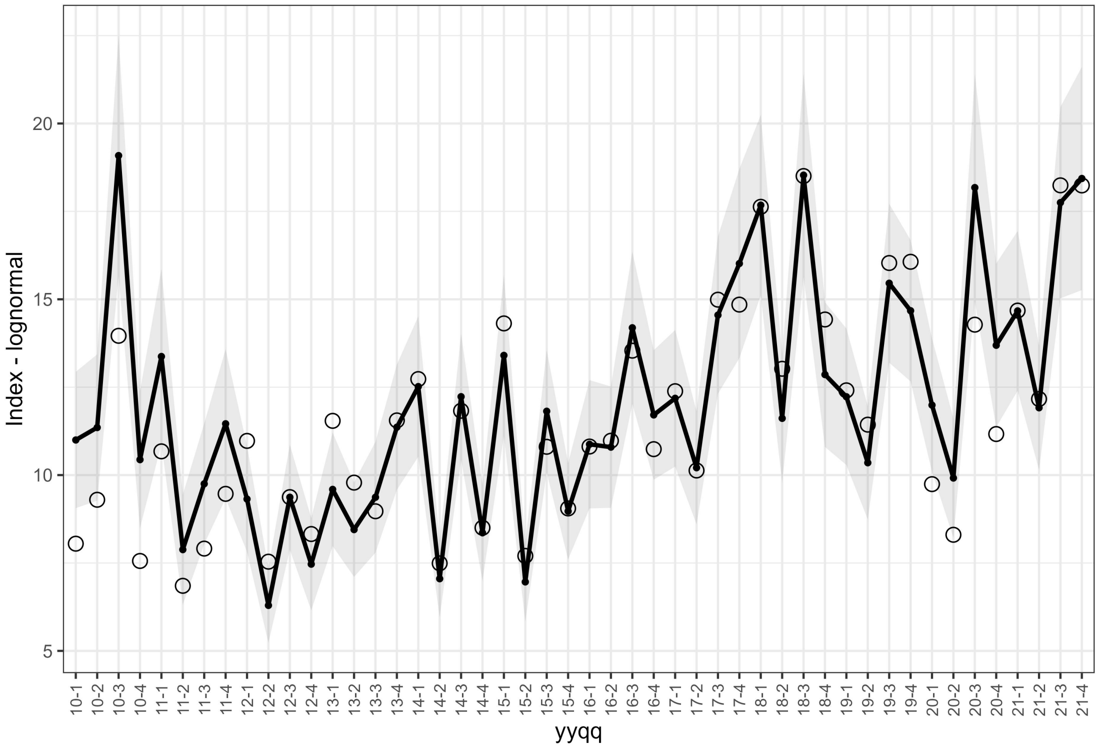


Figure 14: Standardized CPUE index by quarter from the GLMM model. Open circles show the nominal CPUE, the black curve and points shows the standardized CPUE and the gray area is one standard error around that curve.

4 Discussion

Model results are very similar to those presented at the WPTT data preparatory meeting (Kaplan et al. 2023) despite a number of significant changes in the data included in analyses, predictor variables used and model formulations. Both GLMM and GAMM results show an initial decrease in standardized skipjack catch per FOB set over the period 2010-2012, followed by an increase 2012-2018 and ending in a period of increased variability and stabilization or modest increast 2018-2021.

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.

Table 8: Quarterly standardized CPUE predictions from the GLMM.

Year	Quarter	Nominal	Standardized	se	cv
2010	1	8.05	11.00	0.99	0.090
2010	2	9.30	11.35	1.06	0.094
2010	3	13.96	19.09	1.75	0.092
2010	4	7.56	10.44	0.99	0.095
2011	1	10.68	13.37	1.27	0.095
2011	2	6.85	7.88	0.80	0.101
2011	3	7.91	9.75	0.87	0.090
2011	4	9.47	11.46	1.08	0.094
2012	1	10.97	9.32	0.77	0.082
2012	2	7.54	6.29	0.54	0.086
2012	3	9.37	9.37	0.76	0.081
2012	4	8.32	7.47	0.67	0.090
2013	1	11.54	9.59	0.82	0.086
2013	2	9.79	8.45	0.69	0.081
2013	3	8.97	9.37	0.80	0.086
2013	4	11.55	11.36	0.91	0.080
2014	1	12.73	12.52	1.02	0.081
2014	2	7.49	7.05	0.56	0.080
2014	3	11.83	12.23	0.91	0.075
2014	4	8.50	8.36	0.70	0.083
2015	1	14.31	13.40	1.17	0.087
2015	2	7.70	6.96	0.57	0.082
2015	3	10.80	11.82	0.89	0.075
2015	4	9.05	8.97	0.70	0.078
2016	1	10.81	10.88	0.93	0.086
2016	2	10.98	10.80	0.88	0.082
2016	3	13.54	14.19	1.11	0.078
2016	4	10.74	11.71	0.94	0.080
2017	1	12.39	12.18	0.99	0.081
2017	2	10.13	10.21	0.82	0.080
2017	3	14.99	14.55	1.14	0.078
2017	4	14.85	16.02	1.37	0.086
2018	1	17.63	17.68	1.31	0.074
2018	2	13.02	11.61	0.93	0.080
2018	3	18.50	18.53	1.48	0.080
2018	4	14.43	12.86	1.05	0.082
2019	1	12.41	12.23	0.99	0.081
2019	2	11.43	10.35	0.82	0.079
2019	3	16.03	15.46	1.15	0.075
2019	4	16.07	14.67	1.03	0.070
2020	1	9.74	11.99	0.97	0.081
2020	2	8.30	9.92	0.87	0.087
2020	3	14.28	18.18	1.65	0.091
2020	4	11.17	13.69	1.19	0.087
2021	1	14.68	14.67	1.16	0.079
2021	2	12.16	11.91	0.90	0.076
2021	3	18.24	17.75	1.39	0.078
2021	4	18.24	18.44	1.62	0.088

- Improved estimation of covariance (or lack thereof) in model prediction uncertainties when standardizing 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.

5 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

Kaplan DM, Grande M, Alonso MLR, Duparc A, Imzilen T, Floch L, Santiago J (2023) CPUE standardization for skipjack tuna (Katsuwonus pelamis) of the EU purse-seine fishery on floating objects (FOB) in the Indian Ocean. IOTC-2023-WPTT25(DP)-11_Rev1. IOTC 25th Working Party on Tropcial Tunas, Data Preparatory Meeting, Online

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