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

David M. Kaplan[1,2,✉], Maitane Grande[3], María Lourdes Ramos Alonso[4], Jose Carlos Báez[4], Antoine Duparc[1,2], Jon Uranga[3], Taha Imzilen[1,2], Gorka Merino[3], and Giancarlo M. Correa[5]

[1] MARBEC (Univ. Montpellier, CNRS, Ifremer, IRD), 87 av. Jean Monnet, CS30171, 34203 Sète, France
[2] Institut de Recherche pour le Développement (IRD), 87 av. Jean Monnet, 34203 Sète, France
[3] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Herrera Kaia, Portualdea z/g, 20110 Pasaia (Gipuzkoa), Spain
[4] Instituto Español de Oceanografía, Centro Oceanográfico de Málaga, Puerto Pesquero, s/n Apdo., 29640, Fuengirola, Málaga, Spain
[5] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta (Bizkaia), Spain

✉ Correspondence: David M. Kaplan <david.kaplan@ird.fr>

*SUMMARY*

*In 2023, a 1-component GAMM model to standardize SKJ catch per FOB set of the Indian Ocean EU purse-seine fleet for the period 1991-2021 was presented (Kaplan et al. 2023a). This paper updates that model to include data for the period 2022-2024 for the SKJ management strategy evaluation (MSE) process. Results indicate a roughly stable trend 2018-2024 with non-negligible interannual variability.*

*KEYWORDS*

*Mathematical models, Stock assessment, General additive model (GAM), Random effects, Mixed layer, Catchability*

## 1. Introduction

In 2023, both a short (2010-2021; Kaplan et al. 2023b) and long (1991-2021; Kaplan et al. 2023a) time series standardized CPUE for skipjack tuna (SKJ) were produced based on data from the European purse-seine fishery on tropical tunas in the Indian Ocean. The long time series was the principal abundance index used for the stock assessment. It was requested that the same long time series model (Kaplan et al. 2023a) be rerun including more recent data (i.e., 2022-2024 for a total time series of 1991-2024) as part of the management strategy evaluation for SKJ. This analysis is performed here. For completeness, the same figures and text are included as in the 2023 paper, with any differences noted as needed.

## 2. Methods

### *2.1 Catch-effort dataset*

The catch-effort data in this study consisted of French and Spanish FOB sets over the period 1991-2024. The initial data consisted of 154,413 FOB sets corresponding to 151,912 fishing activity entries in the data set. The data was filtered to remove the following data entries (numbers of sets indicated are not exclusive):

- Null sets (9,699 sets)

- Fishing activities corresponding to multiple fishing sets (4,650 sets). Such multi-set fishing activities are concentrated in the early part of the time series, but never exceed 15% of all FOB sets in a given year.
- Sets by vessels in the bottom 5% of vessels in terms of number of positive sets or that were active less than 3 years or whose activities spanned less than 5 years (20 vessels corresponding to 3,199 sets)

After applying all of these filters, the final dataset used for building the CPUE standardization model consisted of 137,742 sets.

In addition to including data from 2022-2024, this dataset differs from that presented in 2023 (Kaplan et al. 2023a) in that improved SKJ catch estimates for the year 2020 have been used. Furthermore, French data from 2022 have been amended with respect to data used in a previous version of this update (Kaplan et al. 2025) and 2024 French and Spanish species compositions corrections have been carried out using the new R-based T3 algorithm detailed in Duparc et al. (2020).

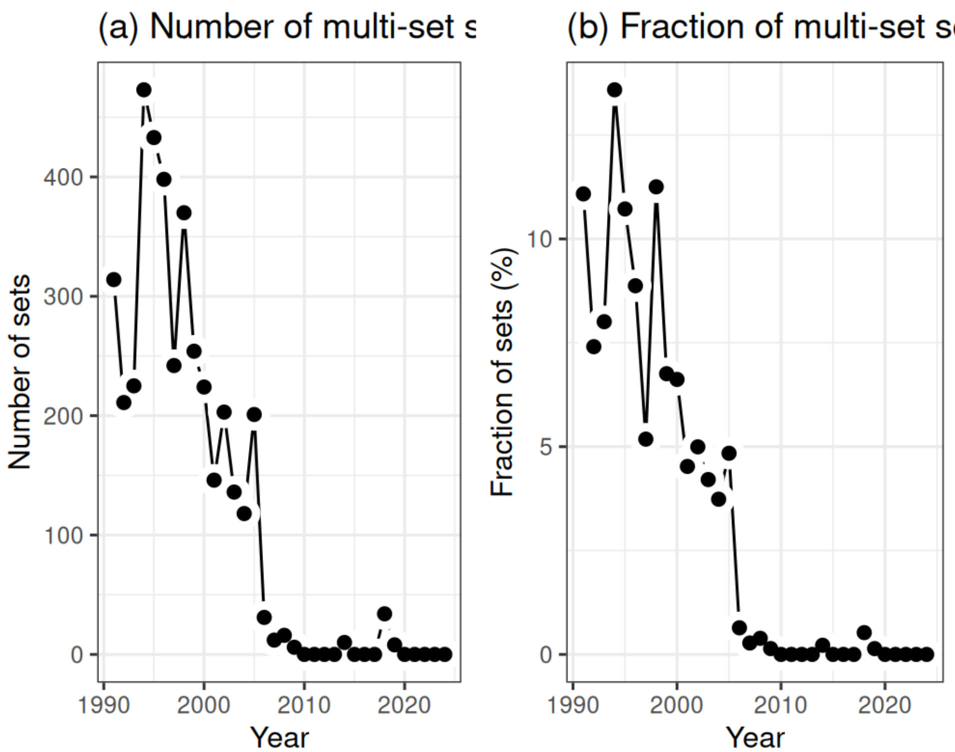


Figure 1: Number and fraction of FOB sets per year that are recorded in multi-set fishing activities.

### *2.2 Predictor variables*

The predictor variables consisted of the the standard temporal, spatial, fleet and vessel identifier predictor variables included in previous standardization efforts (Guéry et al. 2021, Akia et al. 2022, Kaplan et al. 2023a):

- lon,lat spatial variables
- year, month temporal variables
- quarter for stratifying spatial smooths and prediction grids
- vessel country, capacity and year of initiation of activity
- vessel unique identifier

### *2.3 Modeling approach*

Only a 1-part GAMM model was evaluated for this long time series CPUE.

*2.3.1 1-part GAMM model*

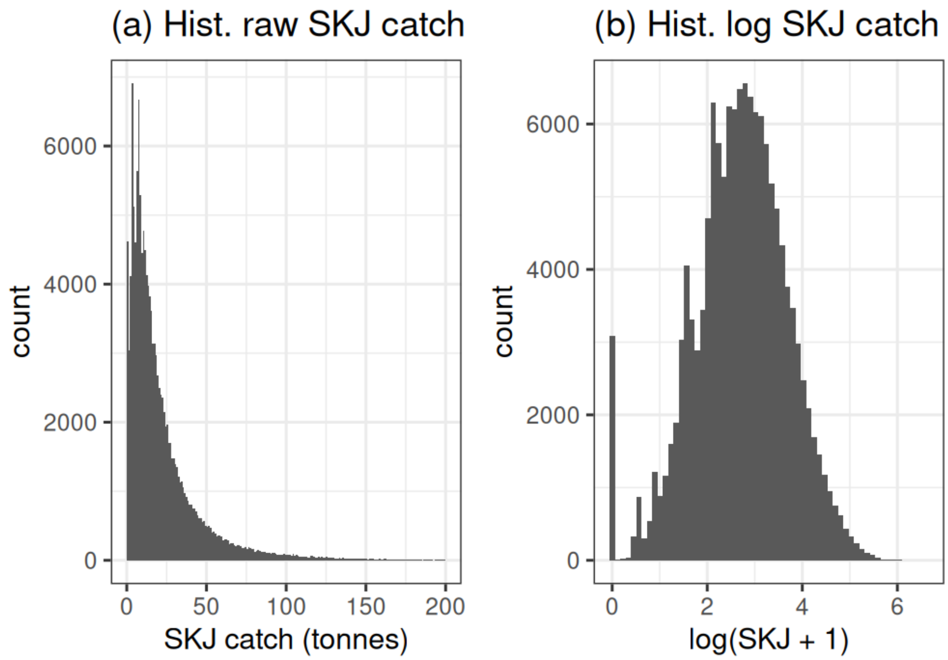


Figure 2: Histograms of SKJ catch data in model training dataset before (a) and afer (b) log transformation.

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, Duparc et al. 2020) catch of skipjack per purse seine FOB set. As for a small number of sets (3,089 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 2) 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, vessel capacity and years of service at the time of fishing as individual smooths and vessel country as a categorical predictor. 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=c(20,11),bs=c("cr","cc")) +
            s(yr_serv,k=10) + s(capacity,k=10) + country,
          data=data,random=list(vessel_id=~1))
```

The model was verified using the `gam.check` function of the `mgcv` package to assure that the numbers of splines used for each smooth (i.e., `k`) were sufficient.

### *2.4 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 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 3.

Variables related to fishing efficiency and catchability were fixed at their median values from the training data set. Specifically, when calculating standardized CPUEs, vessel capacity was fixed at 1850 and vessel initial year of activity was fixed at 1996. Predictions were made for all levels of categorical predictor variable vessel country 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):

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

where $\mu_X$ is the expected value predicted by the GAMM model, $\sigma_X^2$ is the residual variance of the GAMM model (i.e., the `scale` parameter of the model outputs) and $\mu_Y$ 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).

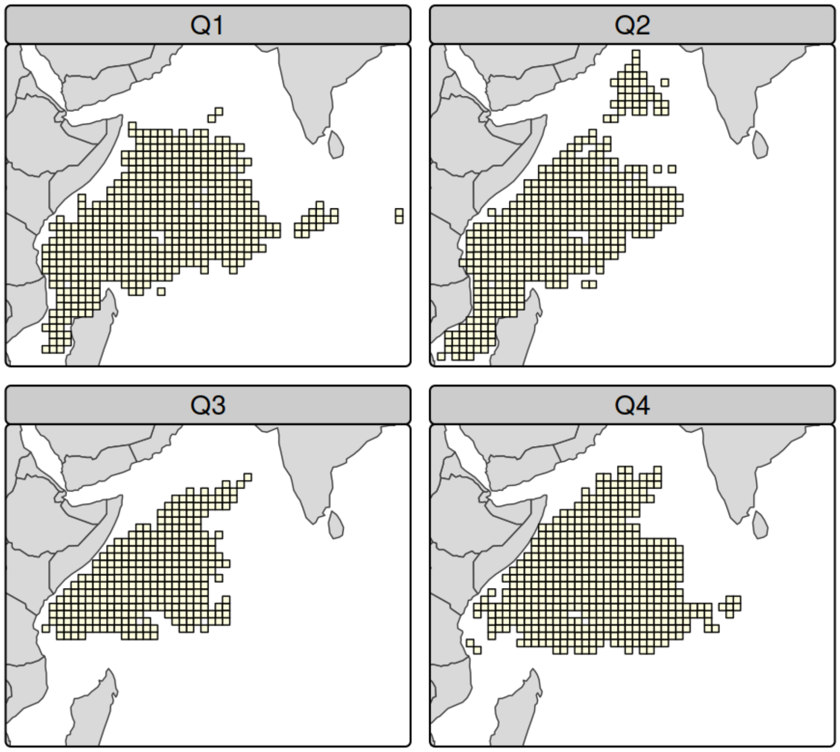


Figure 3: The 1° × 1° grid cells used for model prediction for each quarter. The quarter number is indicated at the top of each panel.

## 3. Results

### *3.1 Nominal catch data*

Boxplots of catch per set find that nominal SKJ CPUEs for the period 2018-2024 have been more or less stable, with values being somewhat lower for the period ~2020-2024 than for 2018-2019 (Figure 4, Figure 5).

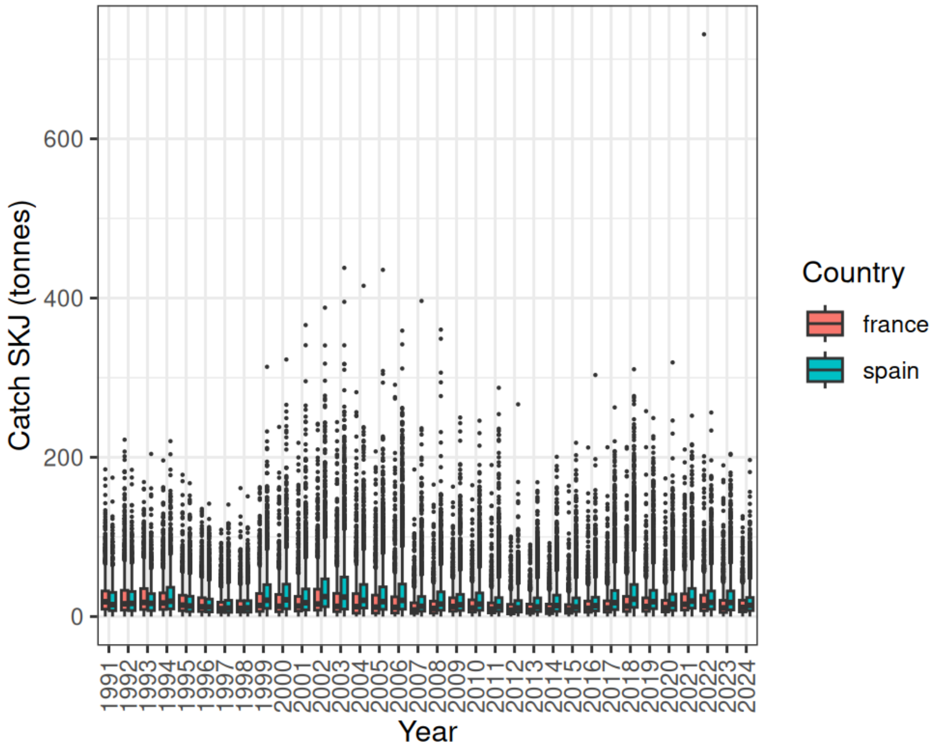


Figure 4: Boxplot of nominal catch per set by year and country.

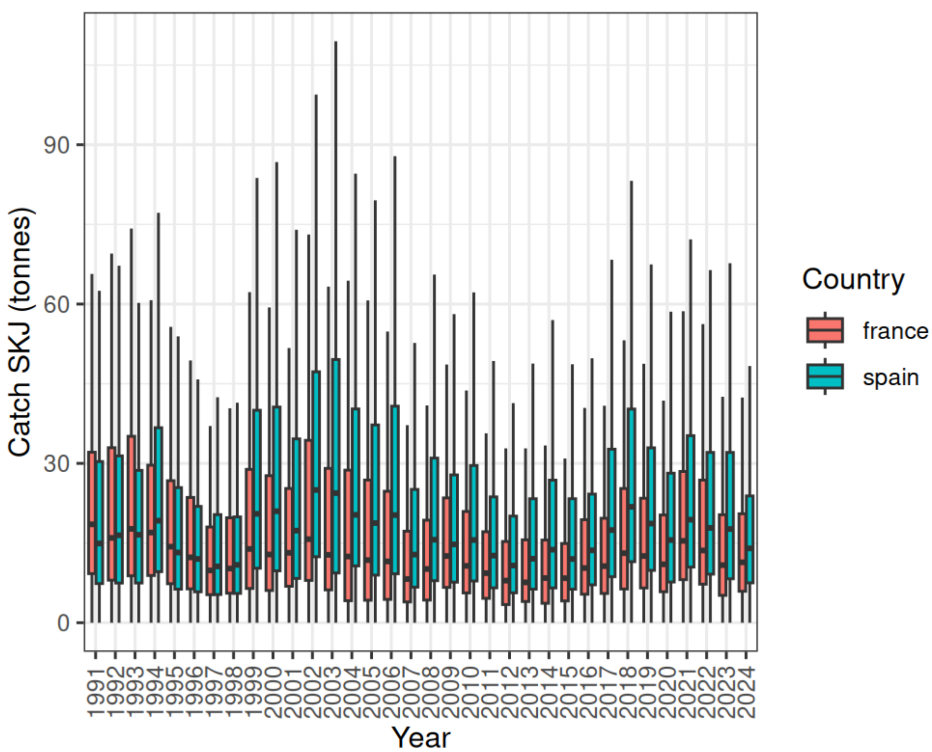


Figure 5: Boxplot of nominal catch per set by year and country. Outliers are not shown in this boxplot.

### *3.2 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 6) and a QQ-plot for the GAM (Figure 7). 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 except for year of vessel initiation of activity (see model summaries below and Table 1).

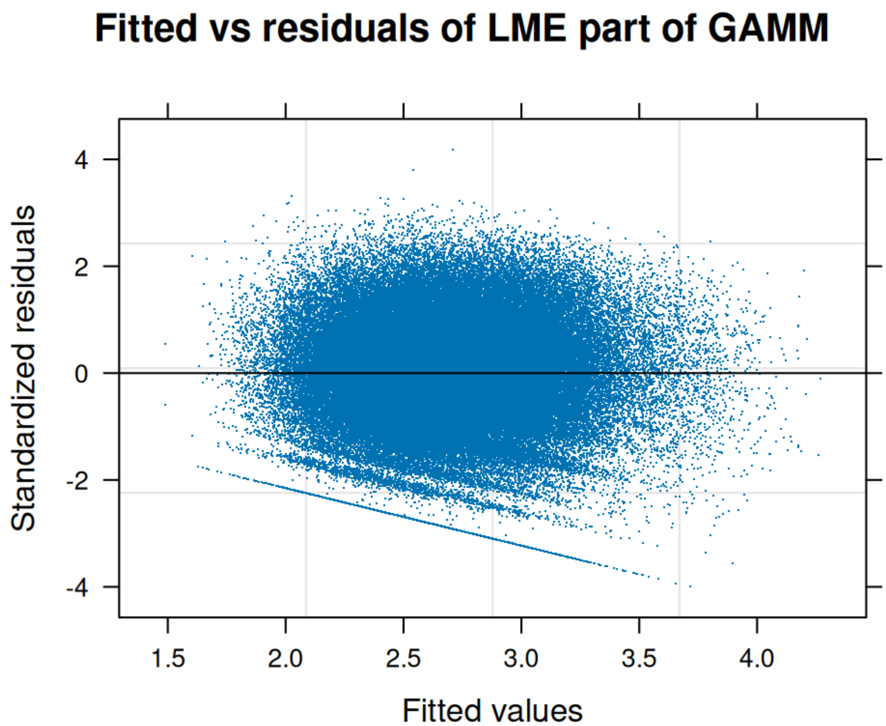


Figure 6: Fitted values versus residuals for LME part (i.e., random part) of GAMM.

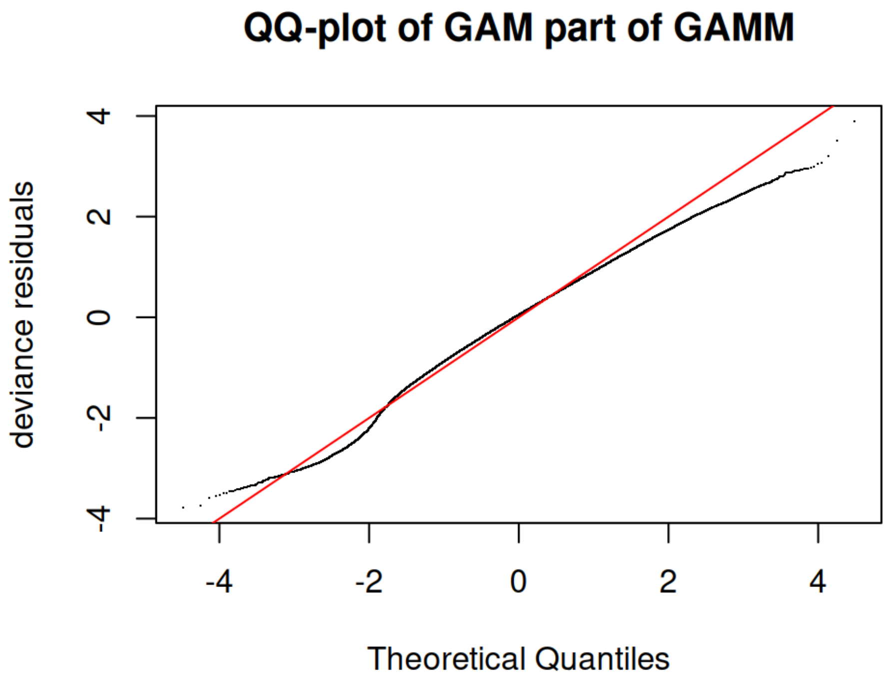


Figure 7: QQ-plot of GAM part (i.e., non-random part) of GAMM.

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

```
  numDF  denDF  F-value p-value
X    17 137654 2668.057  <.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 = c(20, 11), bs = c("cr", "cc")) + s(yr_serv, k = 10) +
    s(capacity, k = 10) + country

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.67166    0.02136 125.069   <2e-16 ***
countryspain  0.06224    0.02927   2.127   0.0334 *
---
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  49.95  49.96 24.728  < 2e-16 ***
te(lon,lat):quarter2  31.07  31.07 12.674  < 2e-16 ***
te(lon,lat):quarter3  37.70  37.70 42.638  < 2e-16 ***
```

```
te(lon,lat):quarter4  45.35  45.35 24.481  < 2e-16 ***
te(year,month)       180.96 180.96 35.029  < 2e-16 ***
s(yr_serv)             1.00   1.00  1.665    0.197    
s(capacity)            1.00   1.00 26.202 3.18e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.112
  Scale est. = 0.86307   n = 137742
```

(a) Parametric terms

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 2.672 | 0.021 | 125.069 | 0.000 |
| countryspain | 0.062 | 0.029 | 2.127 | 0.033 |

(b) Smoothed terms

| | edf | Ref.df | F | p-value |
|---|---|---|---|---|
| te(lon,lat):quarter 1 | 49.955 | 49.955 | 24.728 | 0.000 |
| te(lon,lat):quarter 2 | 31.070 | 31.070 | 12.674 | 0.000 |
| te(lon,lat):quarter 3 | 37.699 | 37.699 | 42.638 | 0.000 |
| te(lon,lat):quarter 4 | 45.348 | 45.348 | 24.481 | 0.000 |
| te(year,month) | 180.964 | 180.964 | 35.029 | 0.000 |
| s(yr_serv) | 1.000 | 1.000 | 1.665 | 0.197 |
| s(capacity) | 1.000 | 1.000 | 26.202 | 0.000 |

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

### *3.3 Marginal effects of predictor variables*

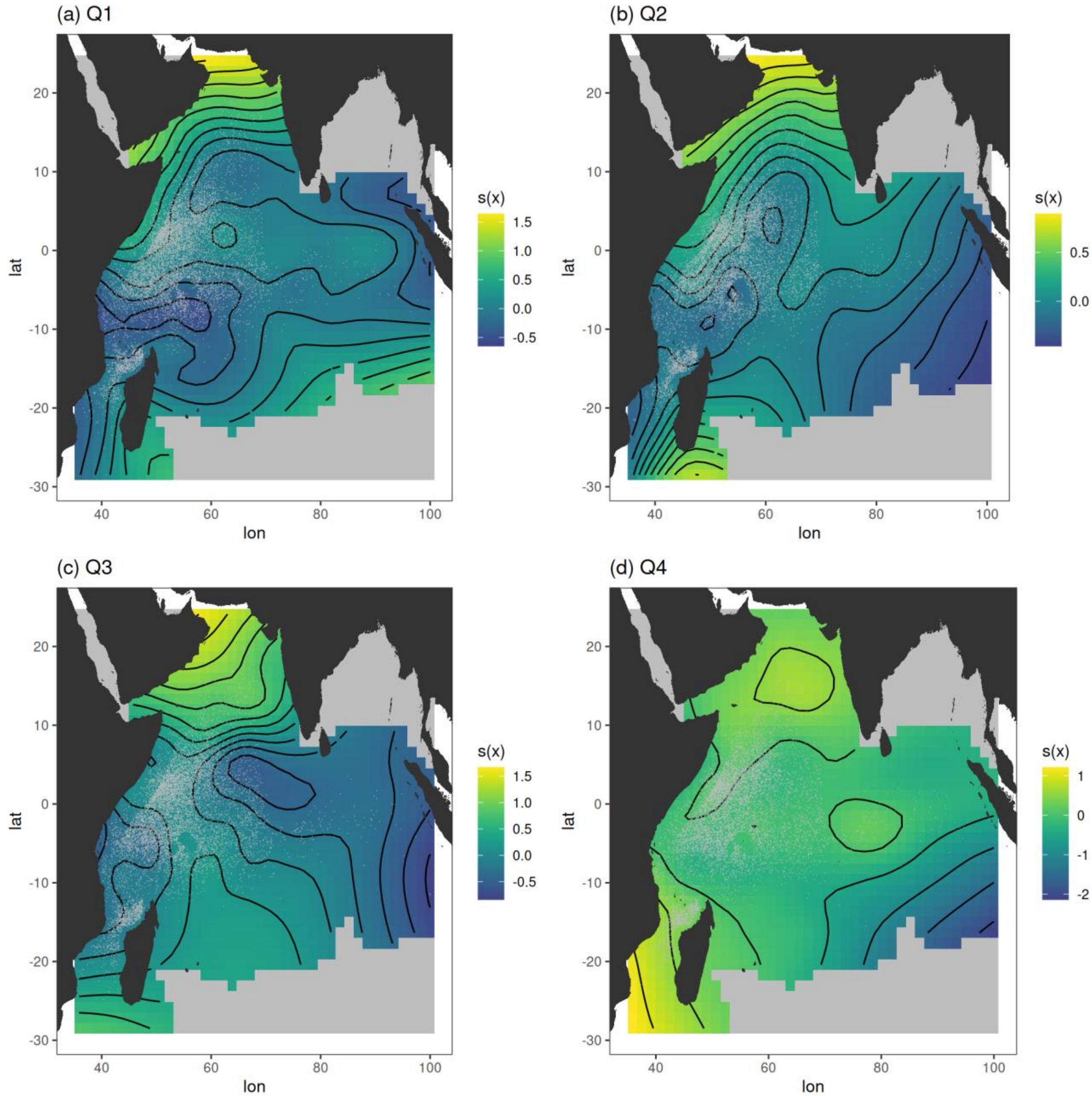


Figure 8: Marginal effect of lon,lat on log SKJ catch per FOB set for each of the four quarters.

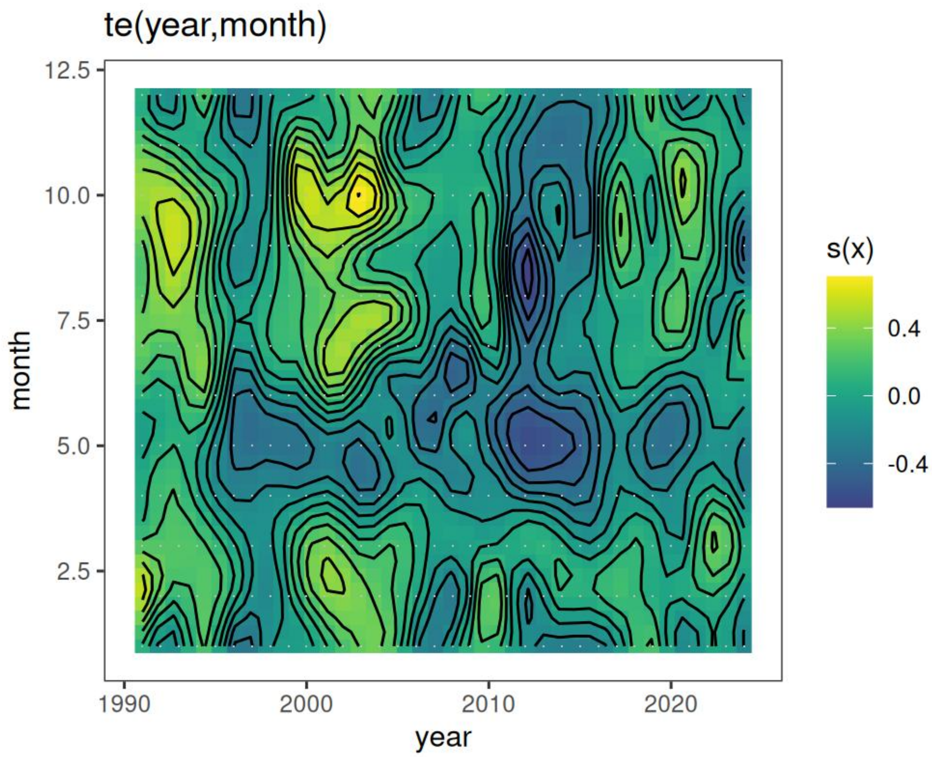


Figure 9: Marginal effect of year,month on log SKJ catch per FOB set.

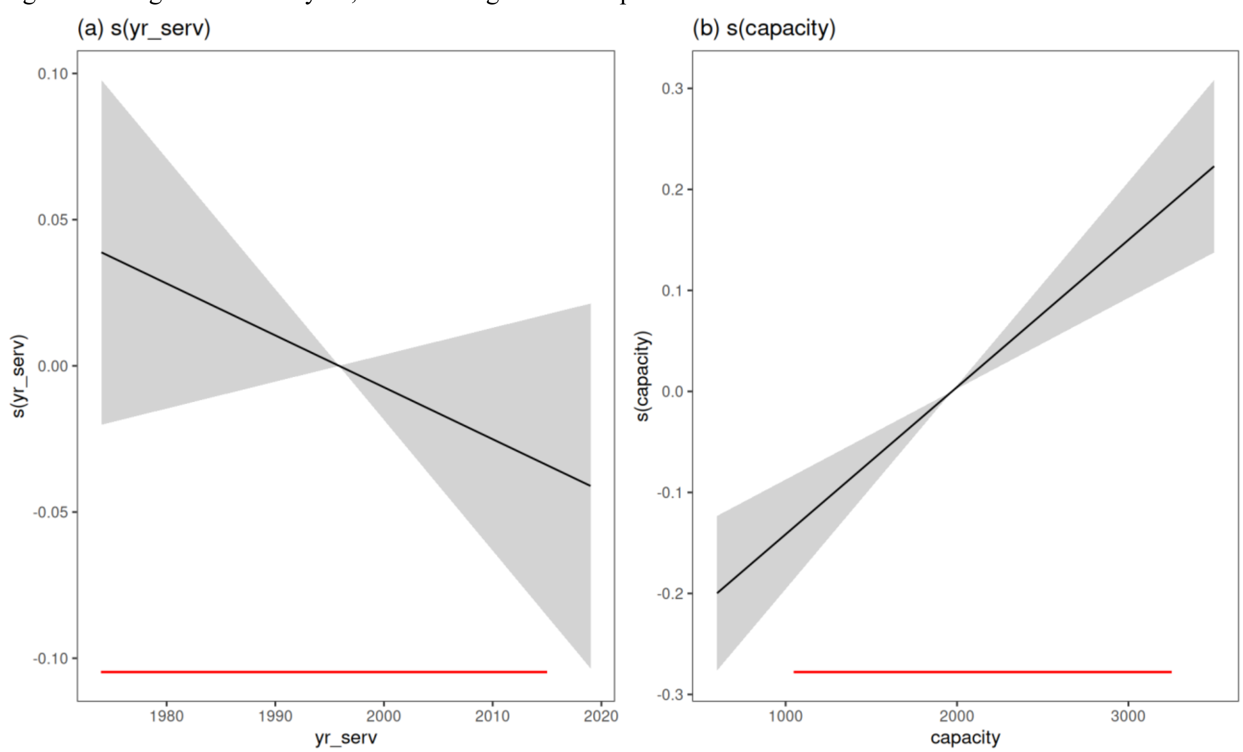


Figure 10: Marginal effects of individual smooths on log SKJ catch per FOB set. The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.

Vessel carrying capacity had a linearly increasing impact on log SKJ catch per set, whereas the effect of year of entry into the fishery was insignificant. The impacts of spatial (Figure 8) and temporal (Figure 9) predictors on log SKJ catch are more difficult to interpret.

### *3.4 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 units of tonnes per set.

| Year | Unweighted, Mean | Unweighted, 2.5% | Unweighted, 97.5% | Weighted, Mean | Weighted, 2.5% | Weighted, 97.5% |
|---|---|---|---|---|---|---|
| 1991 | 27.79 | 23.60 | 32.76 | 27.64 | 23.77 | 32.18 |
| 1992 | 26.93 | 23.83 | 30.42 | 27.33 | 24.49 | 30.51 |
| 1993 | 27.15 | 23.93 | 30.80 | 27.70 | 24.72 | 31.04 |
| 1994 | 27.49 | 24.41 | 30.96 | 27.65 | 24.89 | 30.72 |
| 1995 | 23.93 | 21.32 | 26.86 | 23.76 | 21.46 | 26.31 |
| 1996 | 18.96 | 16.84 | 21.35 | 18.80 | 16.91 | 20.89 |
| 1997 | 17.75 | 15.87 | 19.85 | 17.79 | 16.12 | 19.63 |
| 1998 | 19.69 | 17.37 | 22.31 | 19.95 | 17.81 | 22.34 |
| 1999 | 22.89 | 20.43 | 25.64 | 23.21 | 21.00 | 25.64 |
| 2000 | 26.03 | 23.10 | 29.32 | 26.32 | 23.68 | 29.25 |
| 2001 | 27.57 | 24.50 | 31.03 | 27.89 | 25.13 | 30.94 |
| 2002 | 28.29 | 25.24 | 31.71 | 28.69 | 25.97 | 31.69 |
| 2003 | 28.34 | 25.09 | 32.00 | 28.69 | 25.73 | 31.98 |
| 2004 | 26.68 | 23.82 | 29.89 | 26.75 | 24.21 | 29.56 |
| 2005 | 23.91 | 21.20 | 26.96 | 23.80 | 21.40 | 26.46 |
| 2006 | 20.80 | 18.65 | 23.20 | 20.81 | 18.93 | 22.87 |
| 2007 | 18.93 | 16.91 | 21.20 | 18.92 | 17.13 | 20.88 |
| 2008 | 19.15 | 17.09 | 21.46 | 18.92 | 17.11 | 20.92 |
| 2009 | 21.55 | 19.31 | 24.03 | 21.38 | 19.43 | 23.52 |
| 20 | 23.02 | 20.55 | 25.79 | 22.91 | 20.72 | 25.33 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | | | | | | |
| 2011 | 19.05 | 17.09 | 21.23 | 18.48 | 16.81 | 20.31 |
| 2012 | 16.02 | 14.22 | 18.04 | 15.26 | 13.72 | 16.97 |
| 2013 | 16.91 | 15.15 | 18.88 | 16.40 | 14.89 | 18.05 |
| 2014 | 17.67 | 15.79 | 19.75 | 17.41 | 15.78 | 19.19 |
| 2015 | 16.94 | 15.14 | 18.95 | 16.67 | 15.11 | 18.39 |
| 2016 | 18.74 | 16.87 | 20.82 | 18.57 | 16.95 | 20.33 |
| 2017 | 22.43 | 20.12 | 25.01 | 22.45 | 20.40 | 24.69 |
| 2018 | 23.23 | 21.00 | 25.69 | 23.05 | 21.13 | 25.14 |
| 2019 | 22.40 | 20.13 | 24.92 | 22.28 | 20.29 | 24.46 |
| 2020 | 22.08 | 19.89 | 24.51 | 22.53 | 20.56 | 24.68 |
| 2021 | 22.74 | 20.44 | 25.28 | 23.09 | 21.04 | 25.35 |
| 2022 | 23.18 | 20.77 | 25.87 | 22.74 | 20.64 | 25.06 |
| 2023 | 22.03 | 19.80 | 24.51 | 21.36 | 19.44 | 23.47 |
| 2024 | 19.60 | 17.27 | 22.23 | 19.36 | 17.22 | 21.75 |

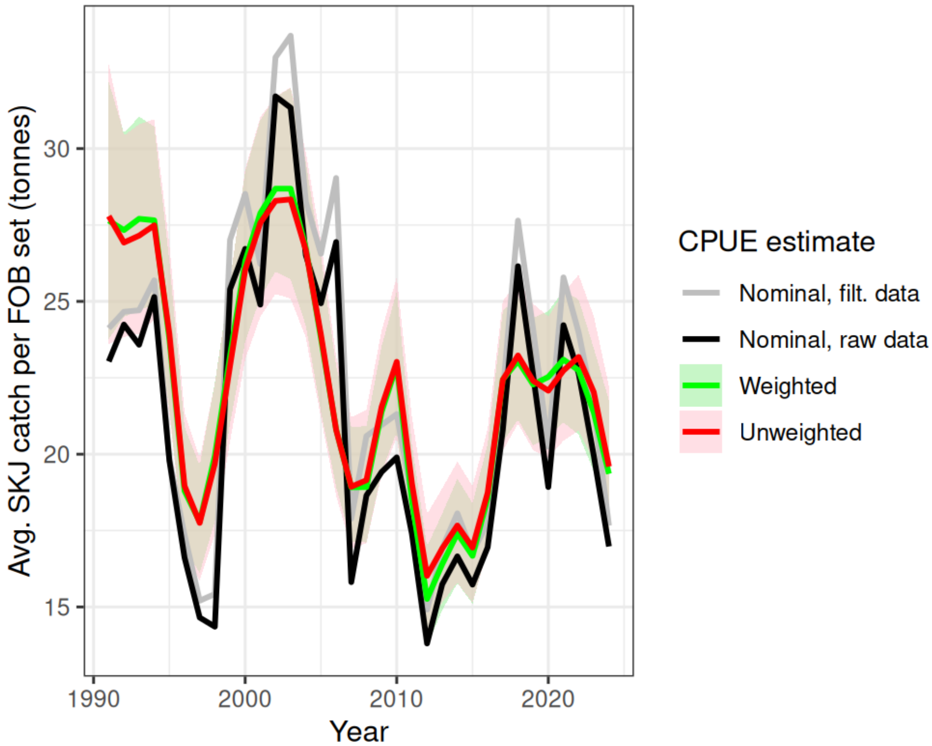


Figure 11: Yearly standardized CPUE predictions from the single-component GAMM model. CPUEs are in units of tonnes of SKJ catch per PS FOB set in the Indian Ocean. Solid curves indicate mean tendencies, whereas shaded areas indicate the upper and lower limits of the 95% confidence interval. Red curves 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.

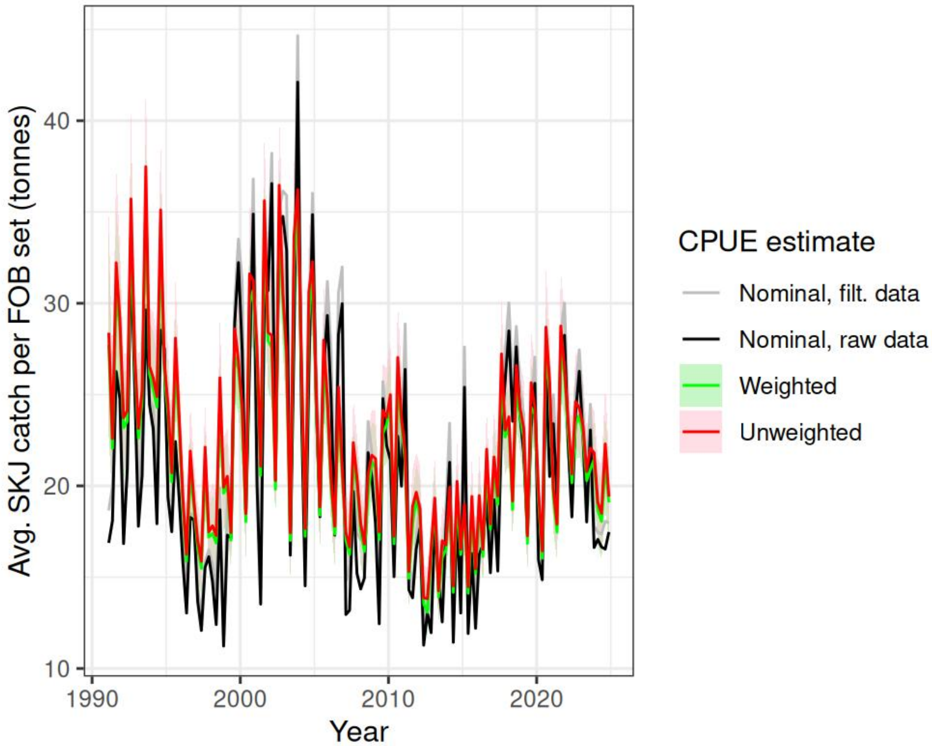


Figure 12: Quarterly standardized CPUE predictions from the single-component GAMM model. CPUEs are in units of tonnes of SKJ catch per PS FOB set in the Indian Ocean. Solid curves indicate mean tendencies, whereas shaded areas indicate the upper and lower limits of the 95% confidence interval. Red curves 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 annual CPUE curves are shown in Figure 11 and Table 2, whereas quarterly CPUEs are shown in Figure 12. 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 occur in 3 specific periods: (a) 1991-1997: standardized CPUEs are consistently above nominal CPUEs, perhaps due to the balancing of catch between the different fleets; (b) 2000-2007: standardized CPUEs are generally below nominal CPUEs, perhaps related to the effect of the "golden years"; and (c) 2007-2017: this period corresponding to the most important impacts of Somali piracy is also characterized by standardized CPUEs being somewhat superior to nominal CPUE. Nominal CPUEs based on the original, unfiltered data and the filtered data used for GAMM model training are generally quite close, but show some differences particularly 2003-2007.

The standardized time series is generally quite close to that presented in 2023 (Kaplan et al. 2023a) except for 2020 which has changed due to the use of revised data. This standardized time series differs from that prevented in a previous version of this update (Kaplan et al. 2025) in that French data from 2022 has been updated, increasing CPUEs for 2022.

### *3.5 Access to standarized time series*

The annual standardized CPUE time series can be download at: https://drive.ird.fr/s/Ccrny4a3qD7Q2wK

The quarterly standardized CPUE time series can be downloaded at: https://drive.ird.fr/s/jZcKz7GQH5AcrzT

## 4. Discussion

Results for 2018-2024 are variable, but roughly stable at values that are intermediate (20-25 tonnes/set) over the scale of the entire time series. 2023-2024 shows a somewhat decreasing trend, though considering interannual variability over the time series it is too early to tell if that simply represents interannual variability or a longer-term downward trend.

## 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.

## 6. 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, Depetris M, Cauquil P, Lebranchu J (2020) Improved version of the Tropical Tuna Treatment process: New perspectives for catch estimates of tropical purse seine fishery. IOTC-2020-WPTT22(AS)-13. 22nd IOTC Working Party on Tropical Tunas-Stock Assessment Meeting (WPTT22(AS))

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

Guéry L, Kaplan D, Marsac F, Grande M, Abascal F, Baez JC, Gaertner D (2021) Standardized purse seine CPUE of Yellowfin tuna in the Indian Ocean for the European fleet. IOTC-2021-WPTT23-10. IOTC 23rd Working Party on Tropical Tunas (WPTT)

Kaplan DM, Grande M, Alonso MLR, Duparc A, Imzilen T, Floch L, Santiago J (2023b) 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

Kaplan DM, Grande M, Alonso MLR, Baez JC, Duparc A, Imzilen T, Floch L, Santiago J (2023a) Long time series CPUE standardization for skipjack tuna (Katsuwonus pelamis) of the EU purse-seine fishery on floating objects (FOB) in the Indian Ocean. IOTC-2023-WPTT25-INF05. IOTC 25th Working Party on Tropcial Tunas, Online

Kaplan DM, Grande M, Ramos Alonso ML, Baez JC, Duparc A, Uranga J, Imzilen T, Merino G, Correa GM (2025) Update on the long time series CPUE standardization for skipjack tuna (Katsuwonus pelamis) of the EU purse-seine fishery on floating objects (FOB) in the Indian Ocean. IOTC-2025-WPTT27(DP)-13. IOTC 27th Working Party on Tropical Tunas, Data Preparation Meeting, Online/Virtual

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