STANDARDIZED CATCH PER UNIT EFFORT OF YELLOWFIN TUNA IN THE ATLANTIC OCEAN FOR THE EUROPEAN PURSE SEINE FLEET OPERATING ON FLOATING OBJECTS

G. M. Correa[1,], D. M. Kaplan[2], M. Grande[3], J. Uranga[3],
M-L. Ramos Alonso[4], P. Pascual Alayón[5], V. Rojo[5], G. Merino[3], and J. Santiago[1]

SUMMARY

*Abundance indices for yellowfin tuna (*Thunnus albacares*) in the Atlantic Ocean were derived from the European purse seine CPUE series (2010-2022) for fishing operations made on floating objects. We used three modelling approaches for CPUE standardization: generalized linear mixed model (GLMM), generalized additive model (GAMst), and a spatiotemporal model (ST). Moreover, we implemented a hurdle method, which separates the probability of a positive set, and the catch (kg) per set in different models. These three CPUE series were compared to the nominal CPUE. To account for effort creep, several candidate variables were tested to be included as explanatory variables. We did not observe a temporal trend, but a high temporal variability in the standardized CPUE by all models. Also, all models predicted similar standardized CPUE series.*

RÉSUMÉ

*Les indices d'abondance de l'albacore (*Thunnus albacares*) dans l'océan Atlantique ont été calculés à partir de la série de CPUE des senneurs européens (2010-2022) pour les opérations de pêche effectuées sous des objets flottants. Nous avons utilisé trois approches de modélisation pour la standardisation des CPUE : un modèle linéaire mixte généralisé (GLMM), un modèle additif généralisé (GAMst) et un modèle spatio-temporel (ST). En outre, une méthode de type « haie», séparant la probabilité d'une opération positive de la prise (kg) par opération dans différents modèles, a été appliquée. Ces trois séries de CPUE ont été comparées à la CPUE nominale. Pour tenir compte de la progression de l'effort, plusieurs variables potentielles ont été testées afin de les inclure en tant que variables explicatives. Nous n'avons pas observé de tendance temporelle, mais une grande variabilité temporelle de la CPUE standardisée par tous les modèles. De plus, tous les modèles ont prédit des séries de CPUE standardisées similaires.*

RESUMEN

*Los índices de abundancia del rabil (*Thunnus albacares*) en el océano Atlántico se obtuvieron a partir de las series de CPUE de cerco europeo (2010-2022) para las operaciones de pesca realizadas sobre objetos flotantes. Utilizamos tres enfoques de modelación para la estandarización de la CPUE: modelo mixto lineal generalizado (GLMM), modelo aditivo generalizado (GAMst) y un modelo espaciotemporal (ST). Además, implementamos un método "hurdle", que separa la probabilidad de un lance positivo y la captura (kg) por lance en diferentes modelos. Estas tres series de CPUE se compararon con la CPUE nominal. Para tener en cuenta el incremento del esfuerzo, se probaron varias variables candidatas para ser incluidas como variables explicativas. No observamos una tendencia temporal, pero sí una gran variabilidad temporal en la CPUE estandarizada por todos los modelos. Además, todos los modelos predijeron series de CPUE estandarizadas similares.*

KEYWORDS

Catchability, Yellowfin tuna, catch, effort

[1] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta (Bizkaia), Spain. ✉ Correspondence: gmoron@azti.es
[2] MARBEC (Univ. Montpellier, CNRS, Ifremer, IRD), av. Jean Monnet, CS30171, 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 5] Instituto Español de Oceanografía, Centro Oceanográfico de Canarias, Calle La Farola del Mar N° 22, Dársena Pesquera, 38180 Santa Cruz de Tenerife, Spain

1. Introduction

An abundance index is a key data input in stock assessment models that can inform fluctuations in population abundance or biomass (Magnusson and Hilborn, 2007). Typically, an abundance index is obtained from fishery-independent (e.g., scientific surveys) and dependent sources. For highly migratory and large pelagic fishes (e.g., tunas), performing a scientific survey is impractical given the large extent of their distribution, therefore fishery-dependent abundance indices such as catch per unit effort (CPUE) are primarily used (Hoyle et al., 2024).

Using nominal CPUE is inappropriate since it is normally biased due to the spatial heterogeneity of fish populations, environmental factors, behavior of fishers, and features of fishing vessels (Wilberg et al., 2009). These factors may produce a disparity between the nominal CPUE and true population abundance trends. For this reason, a CPUE standardization process needs to be performed to remove the influence of external factors that can influence catch rates (Maunder and Punt, 2004).

The European (EU) tuna purse seine fishery operating in the Atlantic Ocean has experienced significant technological developments during the last years, which has increased their efficiency in finding and catching tunas (Torres-Irineo et al., 2014). The EU purse seine fleet is divided into two categories: 1) targeting free-swimming schools (FS), and 2) fishing around floating objects (LS). The latter category initially used natural objects (e.g., logs) that occurred naturally in the ocean; however, they now use artificial buoys known as fishing aggregating devices (a.k.a. FADs) with incorporated technology (e.g., satellite tracks, echo-sounders) (Lopez et al., 2014).

The EU purse seine fleet principally targets three tuna species: yellowfin (*Thunnus albacares*), bigeye (*Thunnus obesus*), and skipjack (*Katsuwonus pelamis*). Yellowfin tuna (YFT) is a fast-growing species widely distributed in the Atlantic Ocean, where the largest production zone is centered in equatorial waters off Africa in the Gulf of Guinea (Rooker et al., 2023). Based on the last stock assessment, YFT is not considered overfished (ICCAT, 2019a), although a continuous decline in YFT biomass has been observed since 1975. The purse seine fishery is the main fishing gear operating on this stock, contributing to more than 75% of the total YFT annual catch. The purse seine LS type catches mostly juveniles while the purse seine FS does mostly on adults (ICCAT, 2019a).

Three assessment platforms were used in the last assessment of the YFT stock in 2019 (ICCAT, 2019b): mpb (Kell et al., 2007), JABBA (Winker et al., 2018), and Stock Synthesis (Methot and Wetzel, 2013), which employed three different indices of abundance to inform biomass trend over time. One of these indices was developed using information from the EU purse seine operating on free schools (Guery et al., 2020), which principally informed variations in YFT adult abundance. The other indices were the joint index derived from the main longline fleets (Hoyle et al., 2019) and the Buoy-derived Abundance Index (Santiago et al., 2019). In this study, we use data from the EU purse seine operating on floating objects (LS) and diverse statistical techniques to derive standardized CPUE indices that can inform juvenile abundance in the assessment process and help to improve the stock assessment model estimates.

2. Methods

2.1 Data

We used logbook data from the EU purse seine fleet (Spain and France) targeting tropical tunas and operating on floating objects in the Atlantic Ocean from 2010 to 2022. The logbook data sets are managed by the Tuna Observatory (Ob7) and the Spanish Institute of Oceanography (IEO) for the French and Spanish fleets, respectively. The raw logbook data (Level 0) produced by the skippers were corrected in terms of total catch per set to account for the difference between reported catch at sea and landed catch. Likewise, the species composition per set was corrected based on port size sampling and the T3 methodology (Pallarés and Hallier, 1997) to generate Level 1 logbook data set.

We excluded observations from fishing sets that operated in areas ($1^\circ \times 1^\circ$) that were not fished for less than 5 years during the studied period in order to retain areas constantly sampled. **Figure 2** shows all the fishing sets used in the CPUE standardization process and **Figure 4** the yearly variation in the number of sets in the data. **Figure 1** shows the histogram of all the catch per set values, both in the original and log-transformed scale.

2.2 Spatial indicators

We used six indicators to summarize the spatial behavior of the fleet during the studied period. Diverse spatial indicators have previously been used for fishery-dependent (Russo et al., 2013; Sosa-López and Manzo-Monroy, 2002) and independent (Woillez et al., 2009; Woillez et al., 2007) sources to increase the probability of picking up changes in critical fleet-related factors over time. We calculated the following spatial indicators, which were calculated by year-quarter:

1. *Clark-Evans*: It is an index of point spatial aggregation (Clark and Evans, 1954), here represented by fishing sets, and provides information on how spatially aggregated the fishing sets took place. Smaller values indicate higher spatial aggregation of fishing sets.
2. *Covered area* (km^2): Represents the spatial expansion of the fishing sets. It was calculated assuming that each fishing set has an area of influence of 1 km^2, and then calculating the spatial union of those areas.
3. *Center of gravity (lon)*: Indicates the longitude where the YFT catches per set were centered.
4. *Center of gravity (lat)*: Indicates the latitude where the YFT catches per set were centered.
5. *Moran's autocorrelation coefficient*: It is a measure of spatial autocorrelation (Gittleman and Kot, 1990), which considers the YFT catch information per fishing set.
6. *Gini coefficient*: It is a measure of inequality (Cowell, 2011) among YFT catch per fishing set values.

2.3 Statistical models

We used three modelling approaches to standardize the observed catch rates. We used the two-part *delta* or *hurdle* approach (Aitchison, 1955), which models two processes: 1) probability of a positive set (catch > 0), modelled as a binomial response, and 2) catch per set, which was log-transformed and assumed a normal response.

2.3.1 Generalized linear mixed model (GLMM)

The GLMM approach is widely used for CPUE standardization. It extends the generalized linear model (GLM) approach by including random variables in the linear predictor, allowing the modelling of fixed and random effects simultaneously (Zuur et al., 2009). Scientists usually model the interaction between time and space as random effects.

The studied area was stratified by using a spatial cluster approach to identify strata that best match the population structure (Ono et al., 2015). To find the strata, we applied a k-medoids algorithm (Kaufman and Rousseeuw, 1990) using the $1^\circ \times 1^\circ$ catch information averaged over time (i.e., mean CPUE). We then calculated the Euclidean distance between pairs of grids, considering the mean CPUE values but also the longitude and latitude information of the grids, and then ran the cluster analysis. Finally, we found the optimal number of clusters by using the average silhouette width method (Rousseeuw, 1987). The identified clusters were used as the *cluster* variable in the GLMM (see **Table 1**).

The model can be represented as:

$$\eta = g(\mu) = \mathbf{X\beta} + \alpha + \epsilon \qquad (1)$$

Where η is the linear predictor. μ is either the expected probability of presence with a logit link function (g) for the first model component, or log-transformed YFT catch (kg) per set for positive CPUE values with an identity link function for the second model component. $\mathbf{X}$ is the design matrix of fixed effects, and $\boldsymbol{\beta}$ is a vector of estimated parameters. α is the interaction between year, quarter, and cluster, which was treated as random effects ($\alpha \sim N(0, \sigma_\alpha^2)$), and only included for the second model component. ϵ represents the random error. We implemented the GLMM model (**Equation 1**) in R using the package *glmmTMB* (Brooks et al., 2017).

We tested different candidate variables (see **Table 1**) to include them as fixed effects for each model component. To do so, we ran an automated model selection in R using the package *MuMIn* (Barton, 2023), which runs different combinations of fixed effect terms, based on the Akaike information criterion (AIC).

Once we obtain the best model (smallest AIC), we then used the *DHARMa* R package (Hartig, 2022) to evaluate the model residuals. Standard raw residuals are not always appropriate when using GLMM, and other types of residuals (e.g., Pearson, deviance residuals) are commonly used instead. *DHARMa* uses a simulation-based approach to create readily interpretable scaled (quantile) residuals for generalized linear mixed models. We analyzed two plots produced by *DHARMa*: 1) the QQ plot residuals, which detects overall deviations from the expected distribution, and 2) the residual vs. predicted plot, which detects trends in residuals along model predictions and simulation outliers.

2.3.2 Generalized additive model (GAMst)

Generalized additive models (GAMs) is an extension of a linear model that allows the inclusion of nonlinear terms (Wood, 2017). GAMs are commonly used to standardize CPUE by modelling the interaction between longitude and latitude through a smooth function to account for spatial structure at a broad scale (Grüss et al., 2018). Therefore, clustering methods or other techniques to define areas are not required. In our study, we allowed the interaction between latitude and longitude to vary per year and modelled as random effect (GAMst).

The model can be represented as:

$$\eta = g(\mu) = \mathbf{X\beta} + s(lon, lat) + ti(lon, lat, time) + \epsilon \qquad (2)$$

Where $s(lon, lat)$ is product smooth fitted to longitude and latitude information, which accounts for spatial autocorrelation, and $s(lon, lat, time)$ is the interaction between space and time modelled using a tensor product spline.

We implemented the GAMst model (**Equation 2**) in R using the package *mgcv* (Wood, 2017). We used the *mgcViz* R package (Fasiolo et al., 2020) to analyze the model results, *MuMIn* to do the model selection, and the *DHARMa* package to evaluate the model residuals.

2.3.3 Spatiotemporal model (ST)

Geostatistical generalized linear mixed effects models can account for unmeasured variables (e.g., population biomass) that cause observations (e.g., catch) to be correlated over space and time through random effects (Anderson et al., 2024). A Gaussian random field (GRF) is multidimensional spatial process, where the random effects that describe the spatial pattern follow a multinomial distribution with mean $\mu = [\mu(s_1), \ldots, \mu(s_n)]$ and spatially structured covariance matrix Σ (Blangiardo and Cameletti, 2015).

For CPUE standardization, the *VAST* (Thorson, 2019) and *R-INLA* (Lindgren and Rue, 2015) R packages have been used in previous studies for distinct fish stocks (Grüss et al., 2019; e.g., Zhou et al., 2019). Recently, Anderson et al. (2024) developed the *sdmTMB* R package that implements geostatistical spatial and spatiotemporal GLMMs in TMB (Kristensen et al., 2016) for model fitting such as done in *VAST*, but also provides a user-friendly interface, especially for users familiar with the *glmmTMB* package. For this reason, we decided to use *sdmTMB* to implement a spatiotemporal model for CPUE standardization.

sdmTMB approximates the GRF by relying on the Stochastic Partial Differential Equation (SPDE) approach using the Integrated Nested Laplace Approximation in *R-INLA* to reduce computational costs. The first step to using the SPDE approach is to construct the mesh, which was composed of triangles covering the studied area with a minimum allowed triangle edge length of 60 km (**Figure 17**). We assumed the spatial correlation is Matérn and bilinearly interpolated over the prediction grid using the values at the mesh vertices. Following Anderson et al. (2024), our model can be mathematically represented as:

$$\eta = g(\mu) = \mathbf{X\beta} + \omega_s + \varepsilon_{s,t} \qquad (3)$$
$$\omega \sim MVN(0, \Sigma_\omega)$$
$$\varepsilon_{t=1} \sim MVN(0, \Sigma_\varepsilon)$$
$$\varepsilon_{t>1} \sim \rho\varepsilon_{t-1} + \sqrt{1-\rho^2}\delta_t, \delta_t \sim MVN(0, \Sigma_\varepsilon)$$

Where ω is the spatial random field (i.e., constant across time), which represents the effect of latent spatial variables that are not otherwise accounted for in the model. ε_t represents the latent spatiotemporal effects. ρ is the autoregressive parameter to allow temporal autocorrelation of the spatial random field with deviations created by δ, and Σ is the covariance matrix of the multivariate normal (MVN) distribution.

2.4 Standardized CPUE calculation

We calculated three standardized CPUE indices by year and quarter for each optimal model (Equation 1**,** Equation 2**,** Equation 3). To do so, we made predictions ($\hat{\cdot}$) on the response scale for each model component for all combinations of years y, quarters q, and areas a. For other covariates, we assumed the mean value of the continuous covariates, or the level with the largest sample size for discrete covariates. The area a represents a cluster for the GLMM or a 1° × 1° prediction grid for the GAMst and ST models (**Figure 3**). For the second model component, we back-transformed the predicted values using $\hat{d} = exp(\hat{\eta} + 0.5\hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the estimated variance of residuals.

Then, the predicted values of both model components were multiplied to produce the CPUE per year, quarter, and area ($\widehat{CPUE}_{y,q,a} = \hat{p}_{y,q,a}\hat{d}_{y,q,a}$). Finally, we calculated the area-weighted CPUE by year and quarter:

$$\widehat{CPUE}_{y,q} = \sum_{a} A_a \times \widehat{CPUE}_{y,q,a} \qquad (4)$$

Where A_a is the area (km^2) of a, excluding the area on land.

2.5 Uncertainty calculation

The standard error ($SE(.)$) of predictions was approximated based on Taylor expansion for each model component. For the first model component:

$$SE(\hat{p}) \approx \frac{exp(-\hat{\eta})}{\left(1 + exp(-\hat{\eta})\right)^2} SE(\hat{\eta})$$

Where $\hat{\eta}$ represents the predictions in the linear predictor scale (logit). For the second model component, we used:

$$SE(\hat{d}) \approx exp(\hat{\eta}) SE(\hat{\eta})$$

Where $\hat{\eta}$ represents the predictions in linear predictor scale (identity).

Then, we applied the delta-method (Lo et al., 1992) to calculate the standard error of the predicted CPUE ($\widehat{CPUE}$):

$$SE(\widehat{CPUE}) = \sqrt{SE(\hat{p})^2 \hat{d}^2 + SE(\hat{d})^2 \hat{p}^2 + SE(\hat{p})^2 SE(\hat{d})^2}$$

3. Results

We observed that the number of sets in the data increased from 2010 to 2017, and decreased since then (**Figure 4**). The values of catch per set were skewed to the left, with values generally smaller than 10 kg and rarely above 50 kg. In log-scale, we did not notice a clear temporal trend. Moreover, the proportion of null sets fluctuated around 5% over time.

The fishing sets were more frequent along the equator and in coastal areas off Gabon and Senegal (**Figure 6**). On the other hand, catches were larger in coastal areas in the southern hemisphere (**Figure 7**). We did not observe a clear spatial pattern of areas with high proportion of null sets (**Figure 8**).

3.1 Spatial indicators

We noticed that the covered area expanded progressively over the years and the fishing set locations tended to be more aggregated (**Figure 5**). The center of gravity fluctuated around 5°W (longitude) and 0° (latitude). The Moran index indicated that the catch per set values increased their spatial autocorrelation from 2010 to 2012, and displayed a slight decrease in the recent years. Moreover, the Gini index indicated that the catch per set values tended to be more heterogeneous over the years, especially after 2018.

3.2 GLMM

We found an optimal number of clusters of three (**Figure 9**), which was used as an explanatory variable in the GLMM model. The optimal model based on AIC showed different explanatory variables for each model component (**Table 2**). Also, the residual check for the first model component did not detect overall deviations from the expected distribution nor outliers (**Figure 10**). On the other side, the second model component showed evidence of dispersion, but not for outliers (**Figure 11**). Predictions are shown in **Figure 12**.

3.3 GAMst

The optimal model based on AIC showed different explanatory variables for each model component (**Table 3**). Also, the residual check for the first model component did not detect overall deviations from the expected distribution or outliers (**Figure 13**). On the other side, the second model component showed evidence of dispersion, but not for outliers (**Figure 14**). Predictions are shown in **Figure 16**.

We showed the predicted CPUE over time in **Figure 18**.

3.4 ST

Due to time constraints, we did not show results for this model. We expect to add these results in a future version of this working document.

References

Aitchison, J., 1955. On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin. Journal of the American Statistical Association 50, 901. https://doi.org/10.2307/2281175

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

Barton, K., 2023. MuMIn: Multi-Model Inference.

Blangiardo, M., Cameletti, M., 2015. Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley and Sons, Inc, Chichester, West Sussex.

Brooks, M., E., Kristensen, K., Benthem, van, J., Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., Bolker, B., M., 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal 9, 378. https://doi.org/10.32614/RJ-2017-066

Clark, P.J., Evans, F.C., 1954. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. Ecology 35, 445–453. https://doi.org/10.2307/1931034

Cowell, F., 2011. Measuring Inequality. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199594030.001.0001

Fasiolo, M., Nedellec, R., Goude, Y., Wood, S.N., 2020. Scalable Visualization Methods for Modern Generalized Additive Models. Journal of Computational and Graphical Statistics 29, 78–86. https://doi.org/10.1080/10618600.2019.1629942

Gittleman, J.L., Kot, M., 1990. Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. Systematic Zoology 39, 227. https://doi.org/10.2307/2992183

Grüss, A., Chagaris, D.D., Babcock, E.A., Tarnecki, J.H., 2018. Assisting Ecosystem-Based Fisheries Management Efforts Using a Comprehensive Survey Database, a Large Environmental Database, and Generalized Additive Models. Marine and Coastal Fisheries 10, 40–70. https://doi.org/10.1002/mcf2.10002

Grüss, A., Walter, J.F., Babcock, E.A., Forrestal, F.C., Thorson, J.T., Lauretta, M.V., Schirripa, M.J., 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fisheries Research 213, 75–93. https://doi.org/10.1016/j.fishres.2019.01.008

Guery, L., Kaplan, D., Deslias, C., Marsac, F., Abascal, F., Pascual, P., Gaertner, D., 2020. Accounting For Fishing Days Without A Fishing Set In The Cpue Standardisation Of Yellowfin Tuna In Free Schools For The Eu Purse Seine Fleet Operating In The Eastern Atlantic Ocean During The 1993-2018 Period (No. SCRS/2019/066). ICCAT (International Commission for the Conservation of Atlantic Tunas), Madrid, Spain.

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

Hoyle, S.D., Campbell, R.A., Ducharme-Barth, N.D., Grüss, A., Moore, B.R., Thorson, J.T., Tremblay-Boyer, L., Winker, H., Zhou, S., Maunder, M.N., 2024. Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research 269, 106860. https://doi.org/10.1016/j.fishres.2023.106860

Hoyle, S.D., Lauretta, M.V., Lee, M.K., Matsumoto, K., Sant'Ana, R., Yokoi, H., 2019. Collaborative study of yellowfin tuna CPUE from multiple Atlantic Ocean longline fleets in 2019 (No. SCRS/2019/081). ICCAT (International Commission for the Conservation of Atlantic Tunas).

ICCAT, 2019b. REPORT OF THE 2019 ICCAT YELLOWFIN TUNA STOCK ASSESSMENT MEETING. ICCAT (International Commission for the Conservation of Atlantic Tunas), Grand-Bassam, Cote d'Ivoire.

ICCAT, 2019a. Report of the 2019 Standing Committee on Research and Statistics (SCRS). ICCAT Report for Biennial Period ({{ICCAT Report}} for {{Biennial Period}}, 2018-2019 No. 2). ICCAT (International Commission for the Conservation of Atlantic Tunas), Madrid, Spain.

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data: An introduction to cluster analysis, Wiley series in probability and mathematical statistics. Wiley, New York.

Kell, L.T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., Scott, R.D., 2007. FLR: An open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science 64, 640–646. https://doi.org/10.1093/icesjms/fsm012

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

Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with *R* - INLA. Journal of Statistical Software 63. https://doi.org/10.18637/jss.v063.i19

Lo, N.C., Jacobson, L.D., Squire, J.L., 1992. Indices of Relative Abundance from Fish Spotter Data based on Delta-Lognormal Models. Canadian Journal of Fisheries and Aquatic Sciences 49. https://doi.org/10.1139/f92-278

Lopez, J., Moreno, G., Sancristobal, I., Murua, J., 2014. Evolution and current state of the technology of echo-sounder buoys used by Spanish tropical tuna purse seiners in the Atlantic, Indian and Pacific Oceans. Fisheries Research 155, 127–137. https://doi.org/10.1016/j.fishres.2014.02.033

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish and Fisheries 8, 337–358. https://doi.org/10.1111/j.1467-2979.2007.00258.x

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: A review of recent approaches. Fisheries Research 70, 141–159. https://doi.org/10.1016/j.fishres.2004.08.002

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012

Ono, K., Punt, A.E., Hilborn, R., 2015. Think outside the grids: An objective approach to define spatial strata for catch and effort analysis. Fisheries Research 170, 89–101. https://doi.org/10.1016/j.fishres.2015.05.021

Pallarés, P., Hallier, J.P., 1997. Analyse du schéma d'échantillonnage multispécifique des thonidés tropicaux (Rapport Scientifique No. OEP/ORSTOM, Programme n° 95/37 réalisé avec le soutien financier de la Commission des Communautés Européennes). Madrid, Spain.

Rooker, J.R., Sluis, M.Z., Kitchens, L.L., Dance, M.A., Falterman, B., Lee, J.M., Liu, H., Miller, N., Murua, H., Rooker, A.M., Saillant, E., Walter, J., David Wells, R.J., 2023. Nursery origin of yellowfin tuna in the western Atlantic Ocean: Significance of Caribbean Sea and trans-Atlantic migrants. Scientific Reports 13, 16277. https://doi.org/10.1038/s41598-023-43163-1

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Russo, T., Parisi, A., Cataudella, S., 2013. Spatial indicators of fishing pressure: Preliminary analyses and possible developments. Ecological Indicators 26, 141–153. https://doi.org/10.1016/j.ecolind.2012.11.002

Santiago, J., Uranga, J., Quincoes, I., Orue, B., Grande, M., Murua, H., Merino, G., Boyra, G., 2019. A novel index of abundance of juvenile yellowfin tuna in the Atlantic Ocean derived from echosounder buoys (No. SCRS/2019/075). ICCAT (International Commission for the Conservation of Atlantic Tunas), Madrid, Spain.

Sosa-López, A., Manzo-Monroy, H., 2002. Spatial patterns of the yellowfin tuna (Thunnus albacares) in the Eastern Pacific Ocean: An exploration of concentration profiles. Ciencias Marinas 28, 331–346. https://doi.org/10.7773/cm.v28i4.241

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

Torres-Irineo, E., Gaertner, D., Chassot, E., Dreyfus-León, M., 2014. Changes in fishing power and fishing strategies driven by new technologies: The case of tropical tuna purse seiners in the eastern Atlantic Ocean. Fisheries Research 155, 10–19. https://doi.org/10.1016/j.fishres.2014.02.017

Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2009. Incorporating Time-Varying Catchability into Population Dynamic Stock Assessment Models. Reviews in Fisheries Science 18, 7–24. https://doi.org/10.1080/10641260903294647

Winker, H., Carvalho, F., Kapur, M., 2018. JABBA: Just Another Bayesian Biomass Assessment. Fisheries Research 204, 275–288. https://doi.org/10.1016/j.fishres.2018.03.010

Woillez, M., Poulard, J.-C., Rivoirard, J., Petitgas, P., Bez, N., 2007. Indices for capturing spatial patterns and their evolution in time, with application to European hake (Merluccius merluccius) in the Bay of Biscay. ICES Journal of Marine Science 64, 537–550. https://doi.org/10.1093/icesjms/fsm025

Woillez, M., Rivoirard, J., Petitgas, P., 2009. Notes on survey-based spatial indicators for monitoring fish populations. Aquatic Living Resources 22, 155–164. https://doi.org/10.1051/alr/2009017

Wood, S.N., 2017. Generalized additive models: An introduction with R, Second edition. ed, Chapman & Hall/CRC texts in statistical science. CRC Press/Taylor & Francis Group, Boca Raton.

Zhou, S., Campbell, R.A., Hoyle, S.D., 2019. Catch per unit effort standardization using spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. ICES Journal of Marine Science 76, 1489–1504. https://doi.org/10.1093/icesjms/fsz034

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R, Statistics for Biology and Health. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-87458-6

Table 1. Candidate explanatory variables for the tested CPUE standardization models.

Variable code	*Variable description*	*Variable type*
year	Year	Factor (levels: 2010,…,2022)
quarter	Quarter of the year	Factor (levels: 1,2,3,4)
cluster	Clustered area (only for GLMM)	Factor (levels: 1,2,3)
lon	Longitude	Numeric
lat	Latitude	Numeric
time	Time as continuous (calculated from year and quarter values)	Numeric
country	Fleet country	Factor (levels: France, Spain)
hold_cap	Vessel hold capacity	Numeric
vessel_op	Age of vessel	Numeric
follow	Followed a FAD with echosounder capacity?	Factor (levels: No, Yes_No-Echo, Yes_Echo)
num_buoys_20nm	Number of buoys within 20 nm	Numeric
num_buoys_250km	Number of buoys within 250 km	Numeric
avg_density	Monthly average density of buoys in a 1x1 grid	Numeric

Table 2. Summary of the GLMM model.

Model	*Term*	*Est*	*Std.err*	*p.val*
Component 1	(Intercept)	2.4497	0.2417	<0.01
	countrySpain	0.3515	0.0660	<0.01
	vessel_op	0.0123	0.0029	<0.01
	num_buoys_250 km	-0.0027	0.0005	<0.01
	avg_density	0.0225	0.0047	<0.01
	followYes_No-echo	-0.2784	0.2263	0.22
	followYes_Echo	0.1458	0.0676	0.03
Component 2	(Intercept)	-0.0452	0.2982	0.88
	countrySpain	0.0458	0.0156	<0.01
	hold_cap	0.0003	0.0000	<0.01
	vessel_op	-0.0042	0.0007	<0.01
	num_buoys_20n m	0.0024	0.0010	0.02
	num_buoys_250 km	-0.0011	0.0001	<0.01
	followYes_No-echo	-0.1734	0.0650	0.01
	followYes_Echo	-0.0578	0.0151	<0.01
	sd__(Intercept)	0.4774		
	sd__Observation	1.0310		

Table 3. Summary of the GAMst model.

Model	*Term*	*Est*	*Std.err*	*p.val*
Component 1	(Intercept)	2.4497	0.2417	<0.01
	countrySpain	0.3515	0.0660	<0.01
	vessel_op	0.0123	0.0029	<0.01
	num_buoys_250 km	-0.0027	0.0005	<0.01
	avg_density	0.0225	0.0047	<0.01
	followYes_No-echo	-0.2784	0.2263	0.22
	followYes_Echo	0.1458	0.0676	0.03
Component 2	(Intercept)	-0.0452	0.2982	0.88
	countrySpain	0.0458	0.0156	<0.01
	hold_cap	0.0003	0.0000	<0.01
	vessel_op	-0.0042	0.0007	<0.01
	num_buoys_20nm	0.0024	0.0010	0.02
	num_buoys_250 km	-0.0011	0.0001	<0.01
	followYes_No-echo	-0.1734	0.0650	0.01
	followYes_Echo	-0.0578	0.0151	<0.01
	sd__(Intercept)	0.4774		
	sd__Observation	1.0310		

Table 4. Predicted CPUE, 95% confidence interval, and coefficient of variation (CV) by the GLMM model by year and quarter (Time column).

Time	*Est*	*Lower*	*Upper*	*CV*
2010.00	0.64	0.52	0.75	0.09
2010.25	1.05	0.94	1.16	0.05
2010.50	1.28	1.11	1.45	0.07
2010.75	1.03	0.92	1.14	0.05
2011.00	1.09	0.94	1.24	0.07
2011.25	1.12	0.98	1.26	0.06
2011.50	0.97	0.84	1.10	0.07
2011.75	0.82	0.73	0.91	0.06
2012.00	0.81	0.72	0.91	0.06
2012.25	1.08	0.92	1.24	0.07
2012.50	0.68	0.59	0.78	0.07
2012.75	1.43	1.27	1.58	0.06
2013.00	0.89	0.78	1.00	0.06
2013.25	0.98	0.87	1.09	0.06
2013.50	1.11	0.99	1.24	0.06
2013.75	1.02	0.91	1.13	0.06
2014.00	0.66	0.57	0.75	0.07
2014.25	1.18	1.05	1.31	0.06
2014.50	1.14	1.04	1.24	0.04
2014.75	1.02	0.92	1.11	0.05
2015.00	0.64	0.58	0.71	0.05
2015.25	1.02	0.92	1.11	0.05
2015.50	1.38	1.15	1.61	0.08
2015.75	0.96	0.88	1.04	0.04
2016.00	0.48	0.41	0.54	0.07
2016.25	1.35	1.20	1.51	0.06
2016.50	1.08	0.98	1.18	0.05
2016.75	1.09	1.01	1.18	0.04
2017.00	0.79	0.70	0.89	0.06
2017.25	0.93	0.83	1.02	0.05
2017.50	1.01	0.90	1.11	0.05
2017.75	1.27	1.16	1.38	0.04
2018.00	1.00	0.90	1.11	0.05
2018.25	0.97	0.83	1.10	0.07
2018.50	0.89	0.79	1.00	0.06
2018.75	1.14	1.01	1.26	0.06
2019.00	1.09	0.93	1.26	0.08
2019.25	1.26	1.13	1.39	0.05
2019.50	0.81	0.67	0.95	0.09
2019.75	0.84	0.75	0.92	0.05
2020.00	0.94	0.79	1.09	0.08
2020.25	1.08	0.95	1.22	0.06
2020.50	0.92	0.81	1.02	0.06
2020.75	1.06	0.97	1.14	0.04
2021.00	0.66	0.13	1.20	0.41
2021.25	1.32	1.16	1.48	0.06
2021.50	1.03	0.92	1.13	0.05
2021.75	0.99	0.92	1.07	0.04
2022.00	0.91	0.71	1.12	0.11
2022.25	1.03	0.93	1.12	0.05
2022.50	0.78	0.70	0.86	0.05
2022.75	1.27	1.18	1.37	0.04

Table 5. Predicted CPUE, 95% confidence interval, and coefficient of variation (CV) by the GAMst model by year and quarter (Time column).

Time	*Est*	*Lower*	*Upper*	*CV*
2010.00	0.69	0.55	0.84	0.11
2010.25	1.00	0.81	1.18	0.09
2010.50	1.11	0.92	1.31	0.09
2010.75	1.19	1.01	1.37	0.08
2011.00	1.08	0.92	1.24	0.07
2011.25	1.09	0.94	1.25	0.07
2011.50	0.88	0.76	0.99	0.07
2011.75	0.95	0.83	1.07	0.06
2012.00	0.97	0.85	1.09	0.06
2012.25	0.96	0.84	1.07	0.06
2012.50	0.81	0.71	0.92	0.07
2012.75	1.26	1.10	1.42	0.06
2013.00	0.95	0.83	1.06	0.06
2013.25	0.89	0.78	1.00	0.06
2013.50	1.10	0.96	1.25	0.07
2013.75	1.06	0.93	1.18	0.06
2014.00	0.57	0.50	0.64	0.06
2014.25	1.09	0.96	1.21	0.06
2014.50	1.34	1.20	1.48	0.05
2014.75	1.00	0.90	1.10	0.05
2015.00	0.64	0.57	0.70	0.05
2015.25	1.02	0.92	1.12	0.05
2015.50	1.38	1.23	1.52	0.05
2015.75	0.97	0.87	1.06	0.05
2016.00	0.53	0.47	0.59	0.06
2016.25	1.39	1.23	1.55	0.06
2016.50	1.06	0.95	1.17	0.05
2016.75	1.02	0.92	1.12	0.05
2017.00	0.65	0.59	0.72	0.05
2017.25	1.05	0.94	1.15	0.05
2017.50	1.09	0.98	1.19	0.05
2017.75	1.21	1.10	1.32	0.05
2018.00	0.92	0.83	1.00	0.05
2018.25	0.96	0.87	1.06	0.05
2018.50	1.11	1.00	1.22	0.05
2018.75	1.02	0.91	1.12	0.05
2019.00	0.86	0.75	0.96	0.06
2019.25	1.17	1.04	1.29	0.06
2019.50	0.97	0.86	1.08	0.06
2019.75	1.01	0.90	1.12	0.05
2020.00	0.93	0.82	1.05	0.06
2020.25	1.26	1.12	1.41	0.06
2020.50	0.84	0.75	0.93	0.05
2020.75	0.96	0.87	1.06	0.05
2021.00	0.76	0.65	0.87	0.07
2021.25	1.14	1.03	1.25	0.05
2021.50	1.07	0.96	1.19	0.06
2021.75	1.03	0.92	1.13	0.05
2022.00	0.87	0.74	0.99	0.07
2022.25	0.91	0.80	1.03	0.07
2022.50	0.97	0.83	1.11	0.07
2022.75	1.25	1.06	1.44	0.08

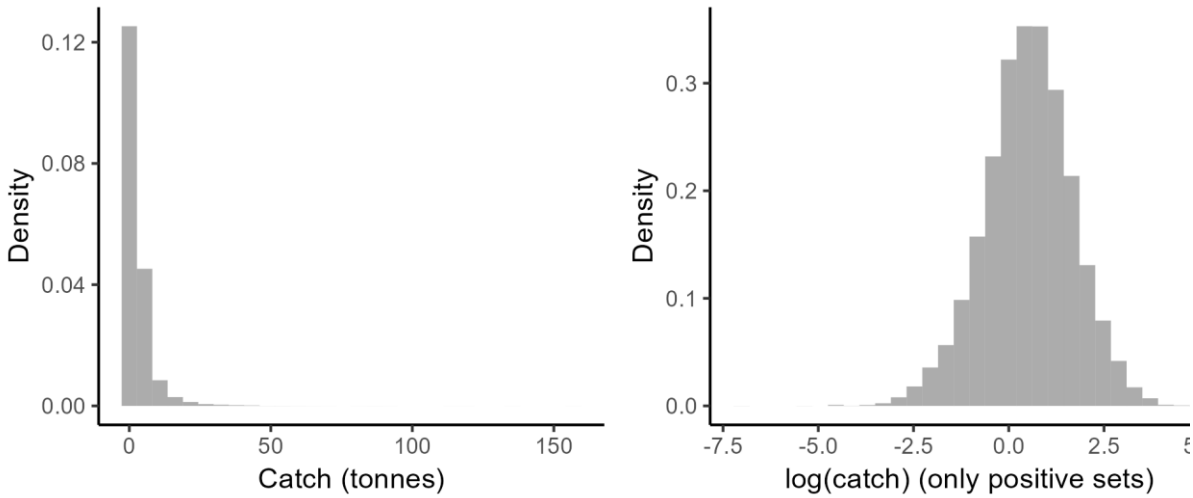

Figure 1. Distribution of observed catch per set values (only positive fishing sets).

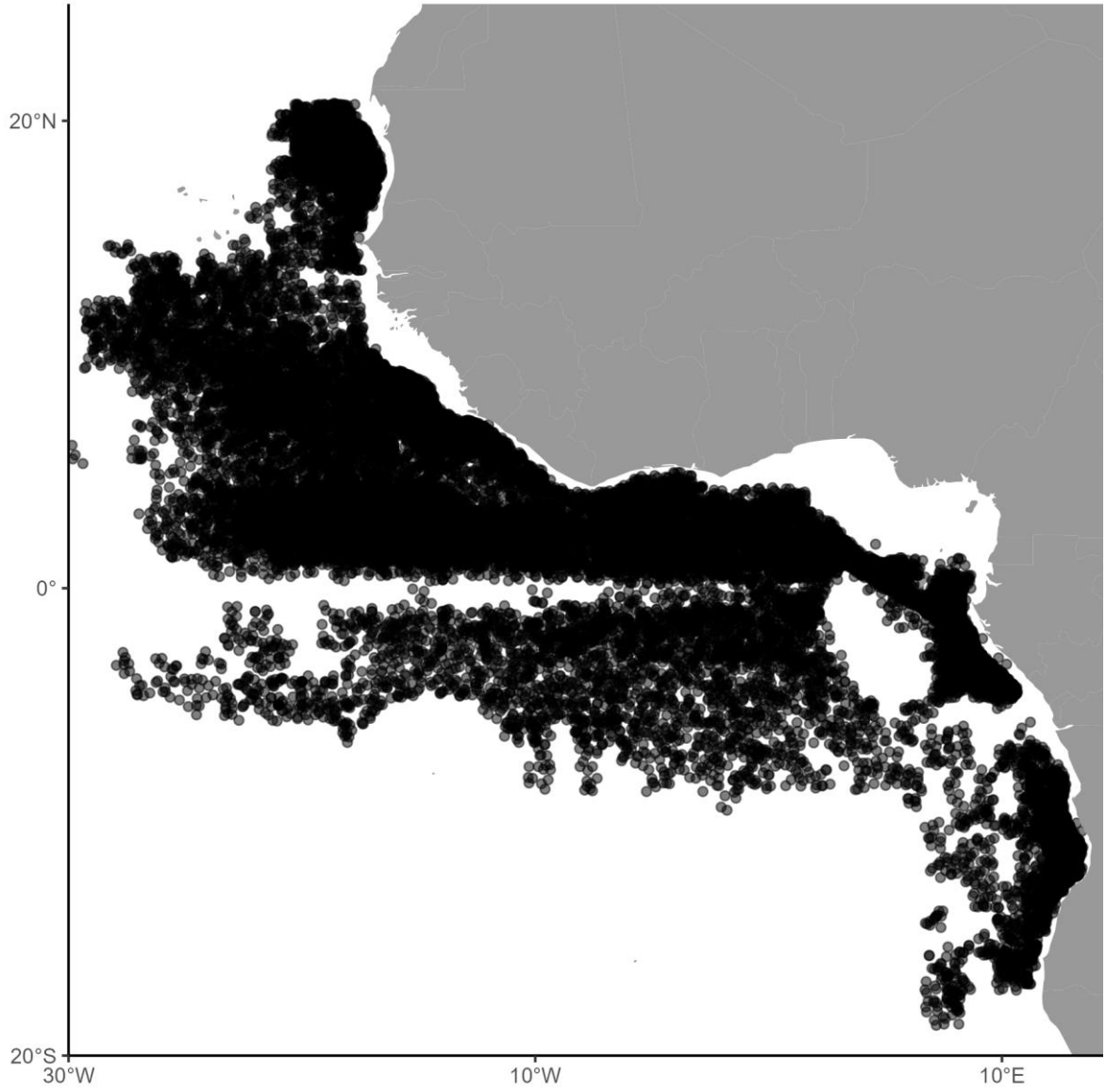

Figure 2. Fishing sets included in the standardization process.

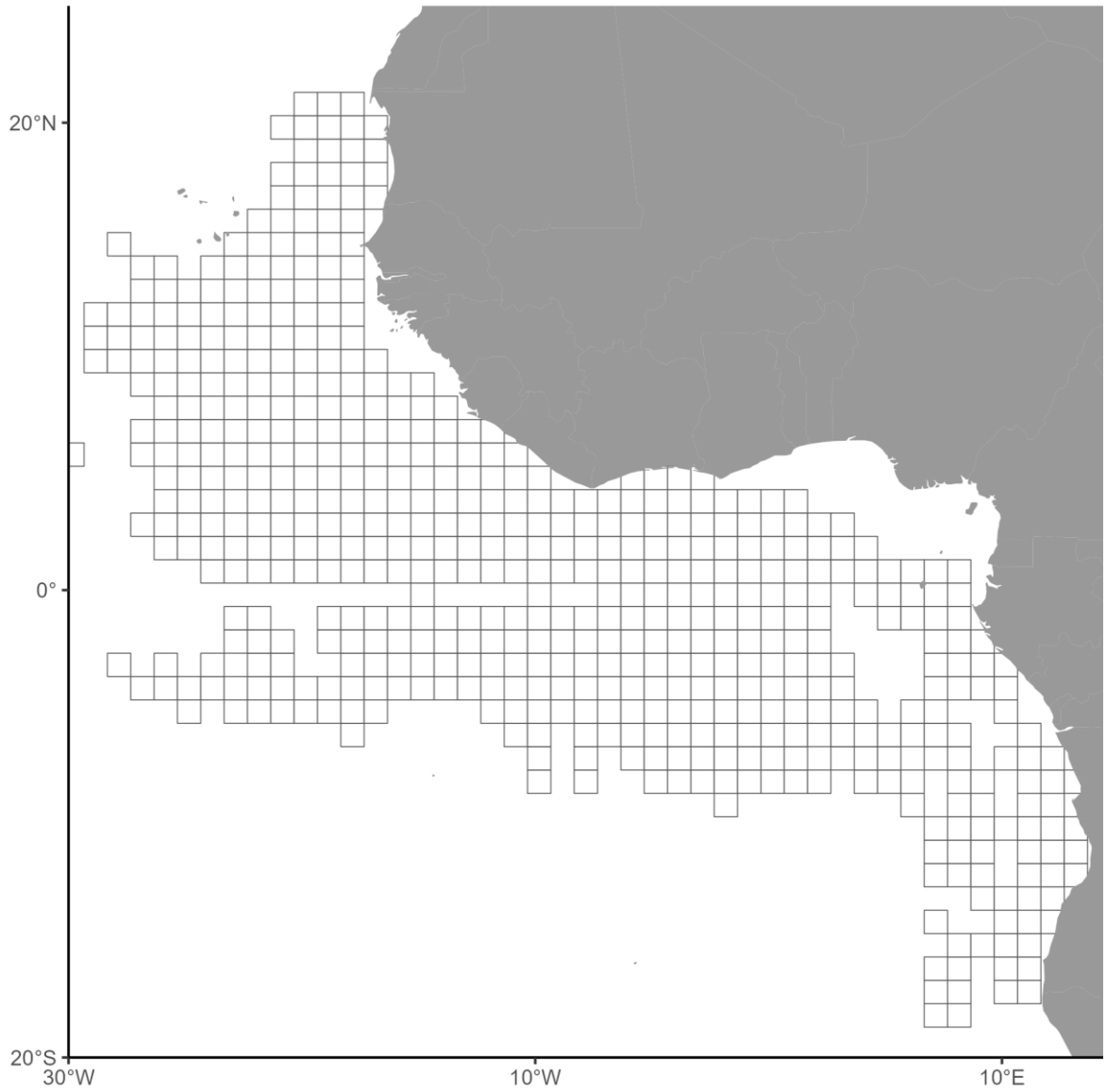

Figure 3. Prediction grid used for the GAMst and ST models.

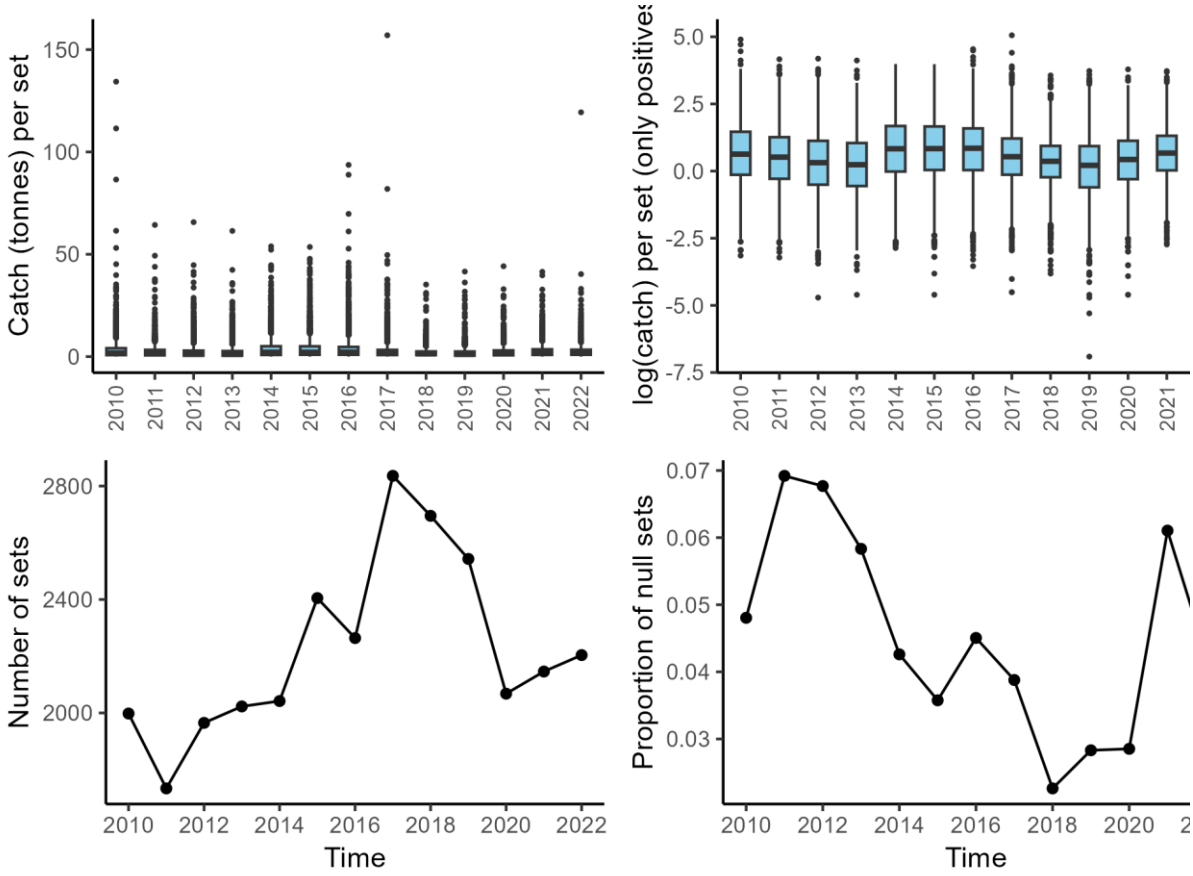

Figure 4. Distribution of catch per set, number of sets, and proportion of null sets in the data per year**.**

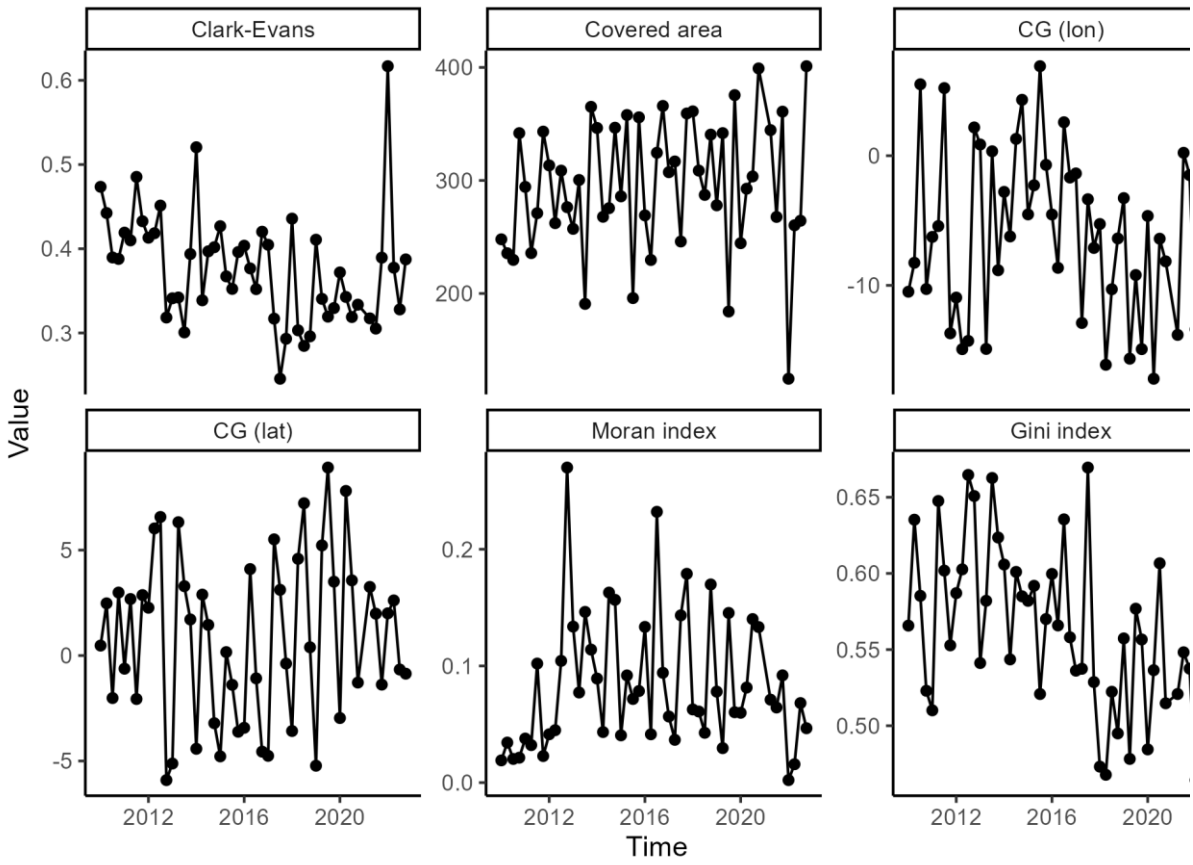

Figure 5. Spatial indicators calculated by quarter.

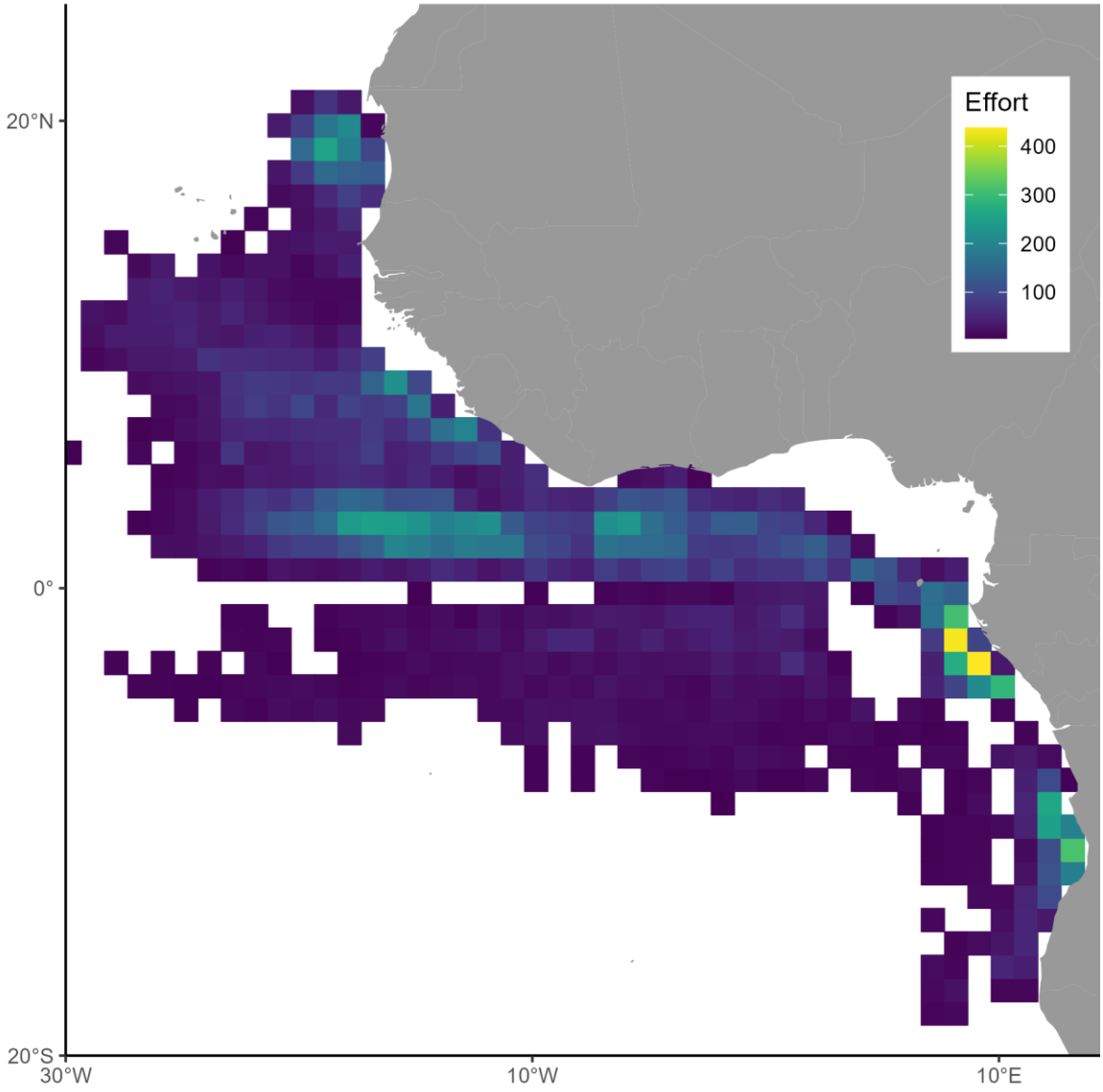

Figure 6. Aggregated number of fishing sets (effort) per grid.

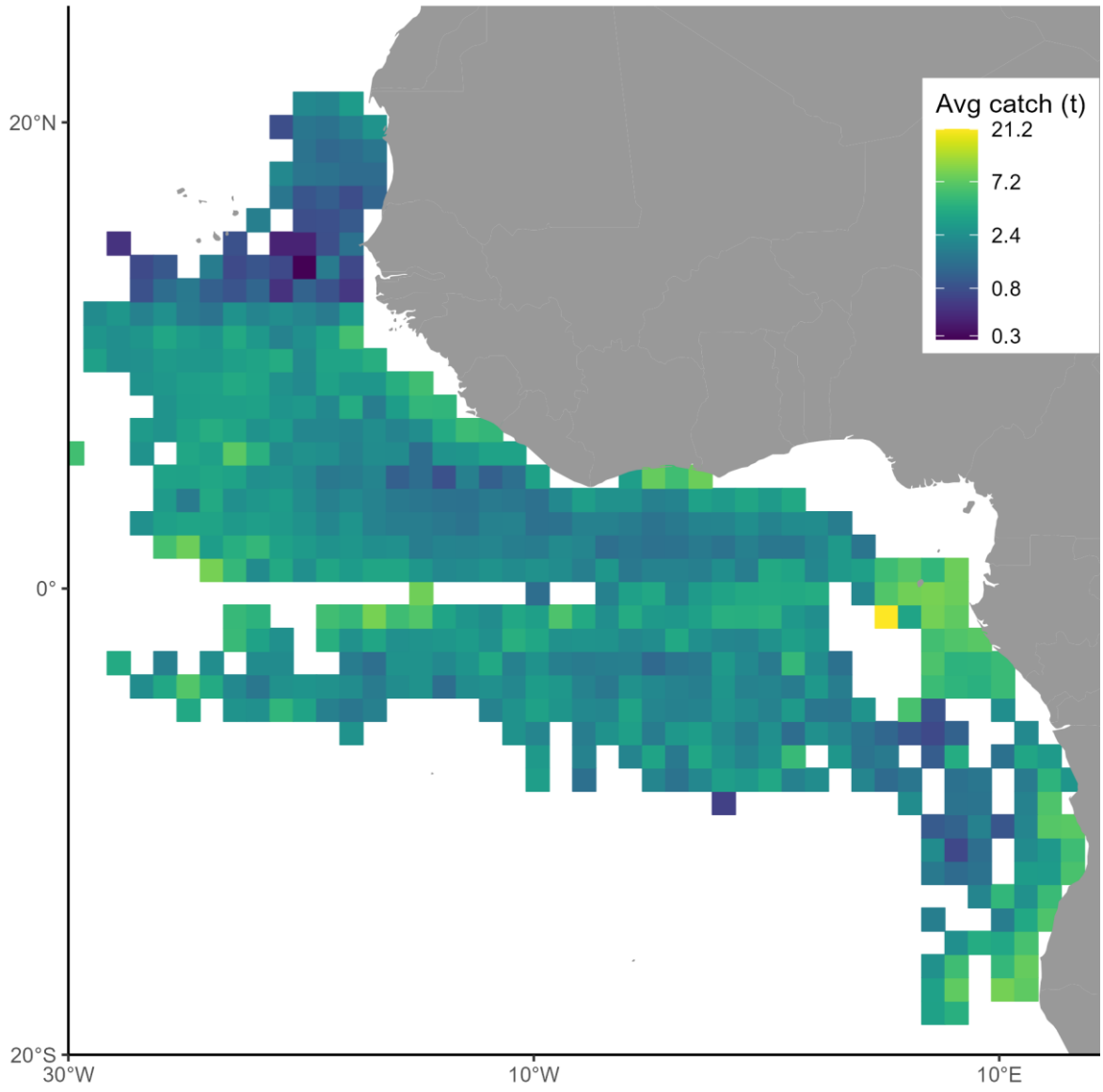

Figure 7. Average catch per grid.

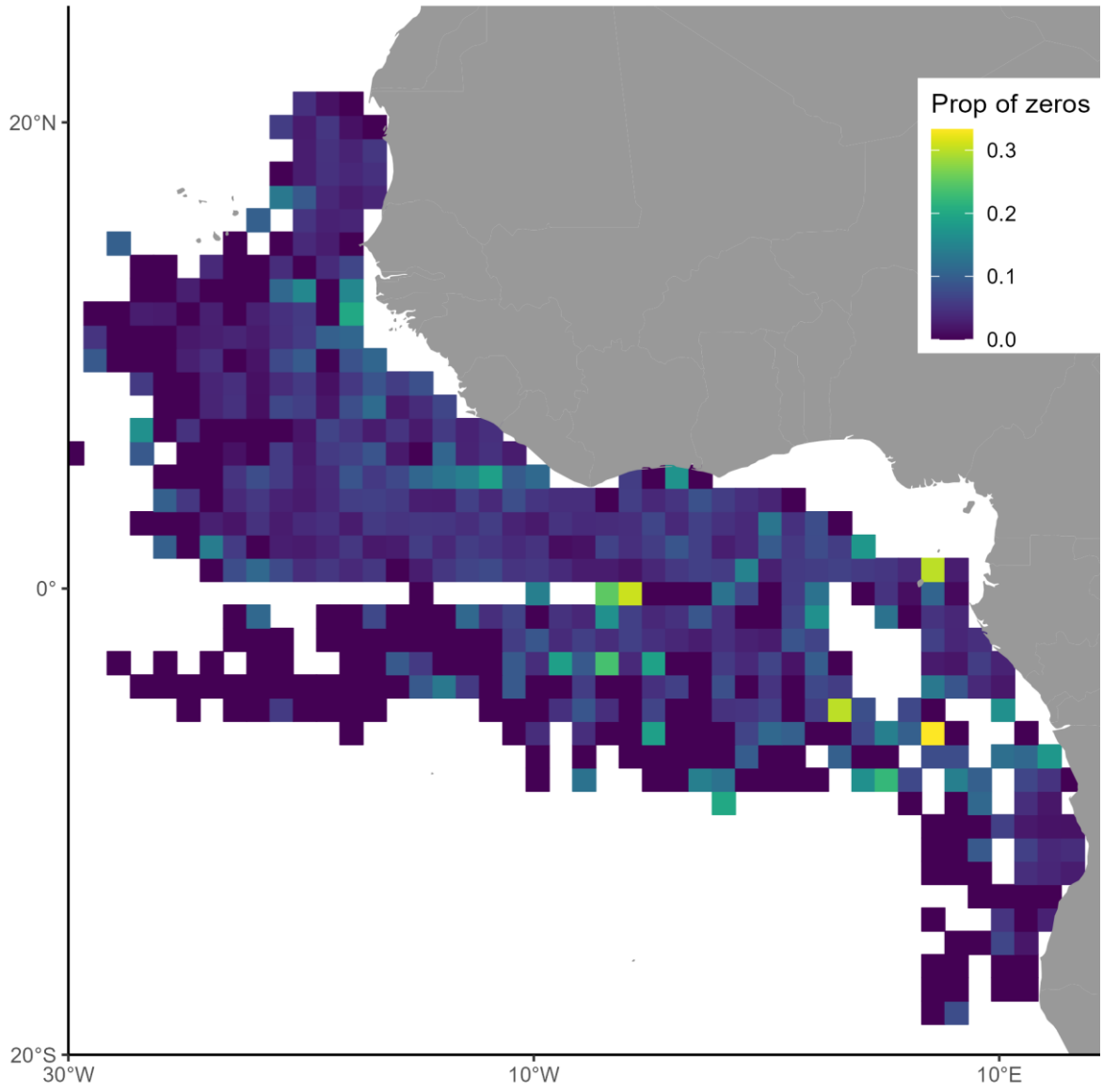

Figure 8. Proportion of null sets per grid.

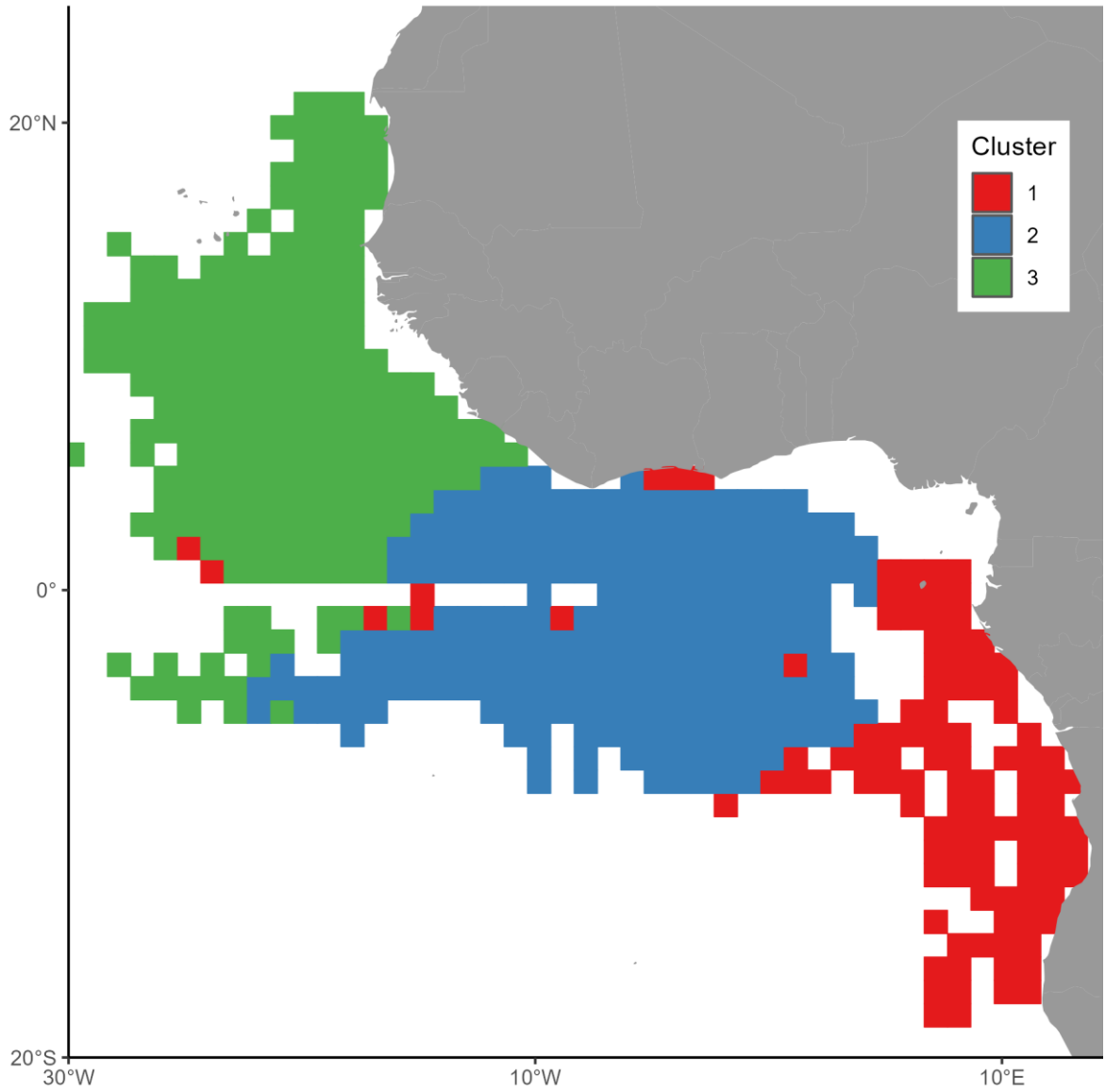

Figure 9. Strata identified by the clustering method and used in the GLMM.

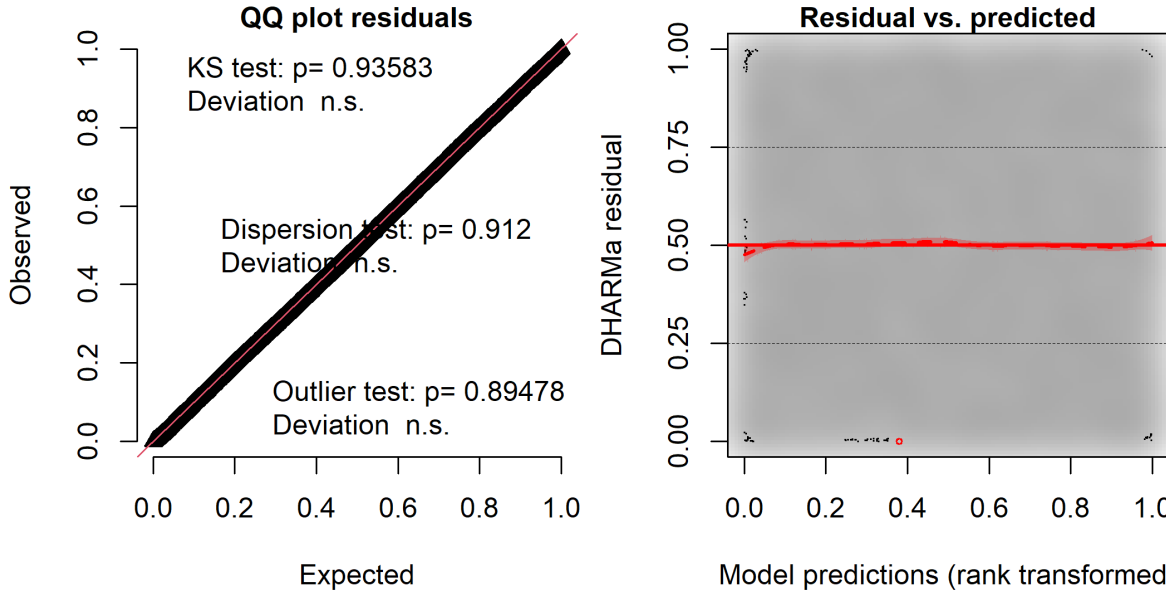

Figure 10. QQ-plot (left) and residual plot (right) for the first component of the GLMM model**.**

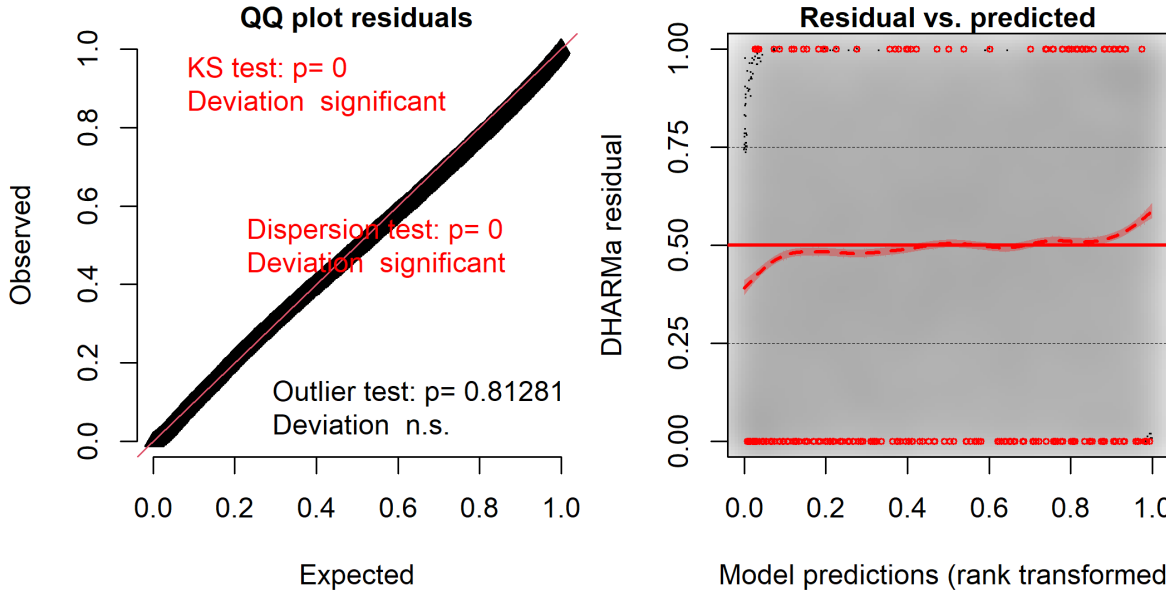

Figure 11. QQ-plot (left) and residual plot (right) for the second component of the GLMM model.

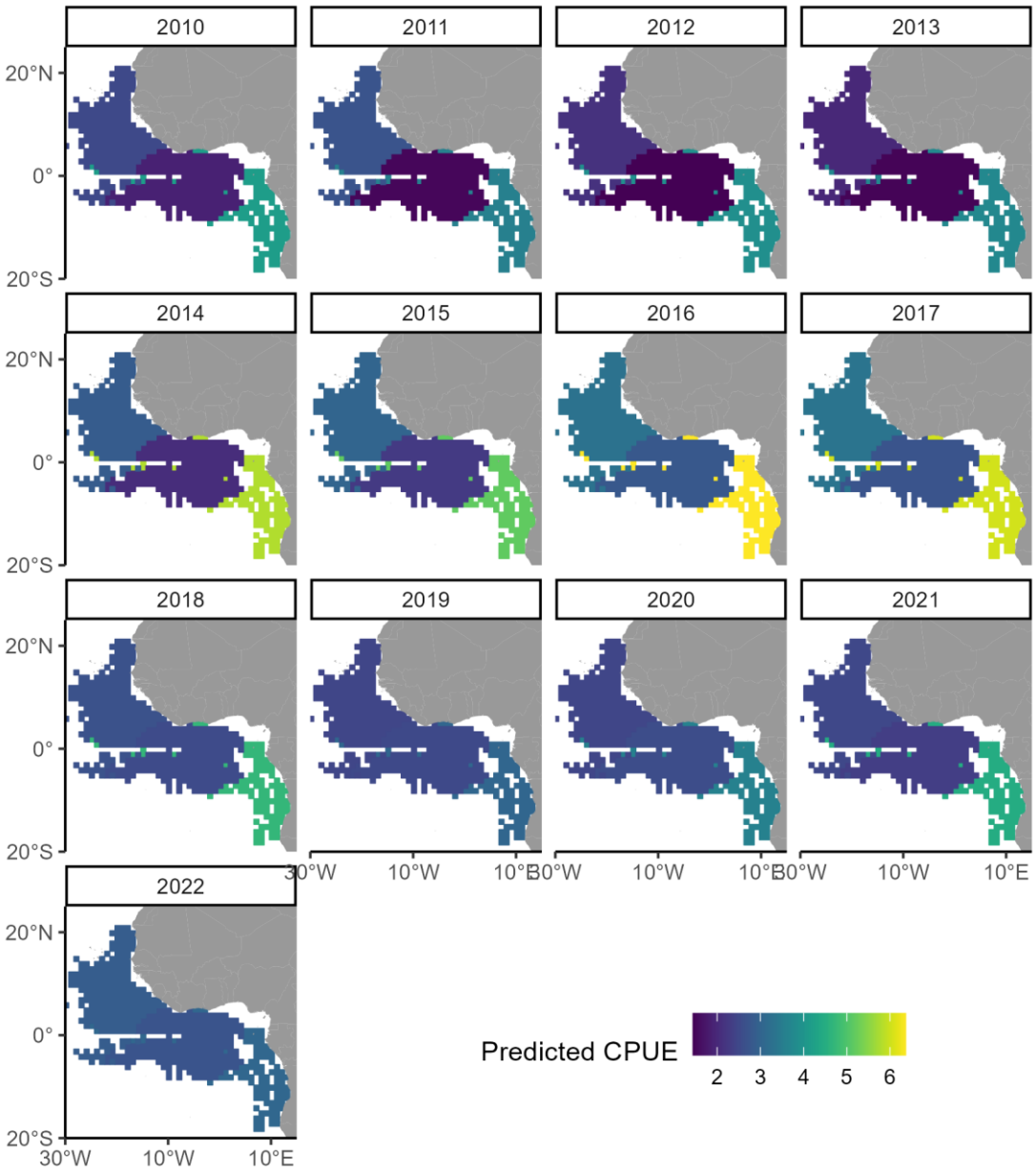

Figure 12. Predicted CPUE for each year-quarter-cluster combination by the GLMM model.

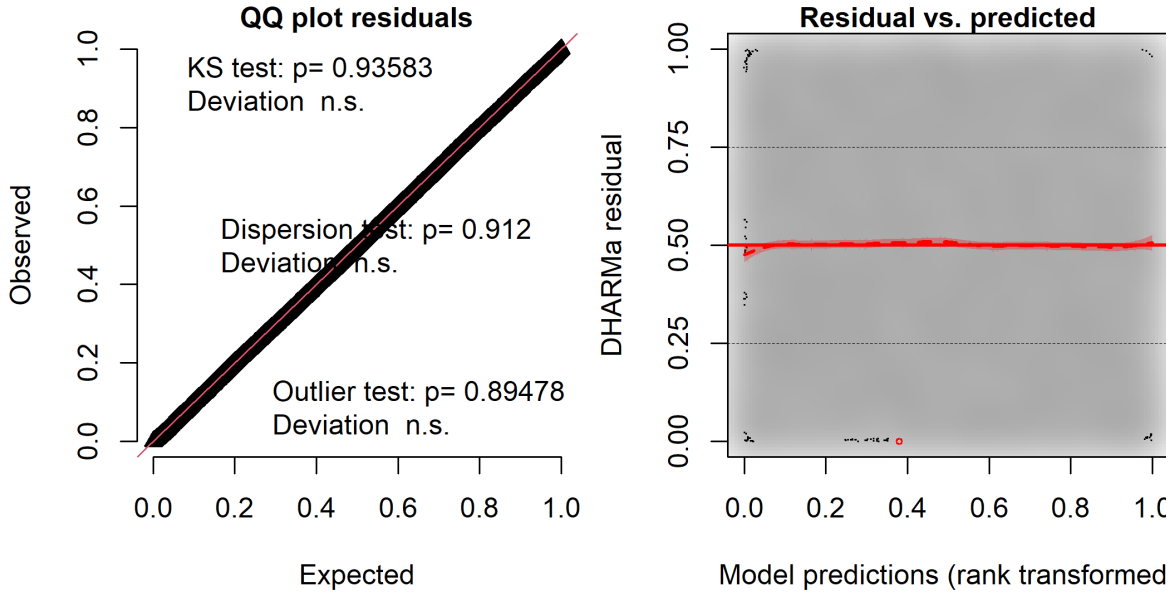

Figure 13. QQ-plot (left) and residual plot (right) for the first component of the GAMst model.

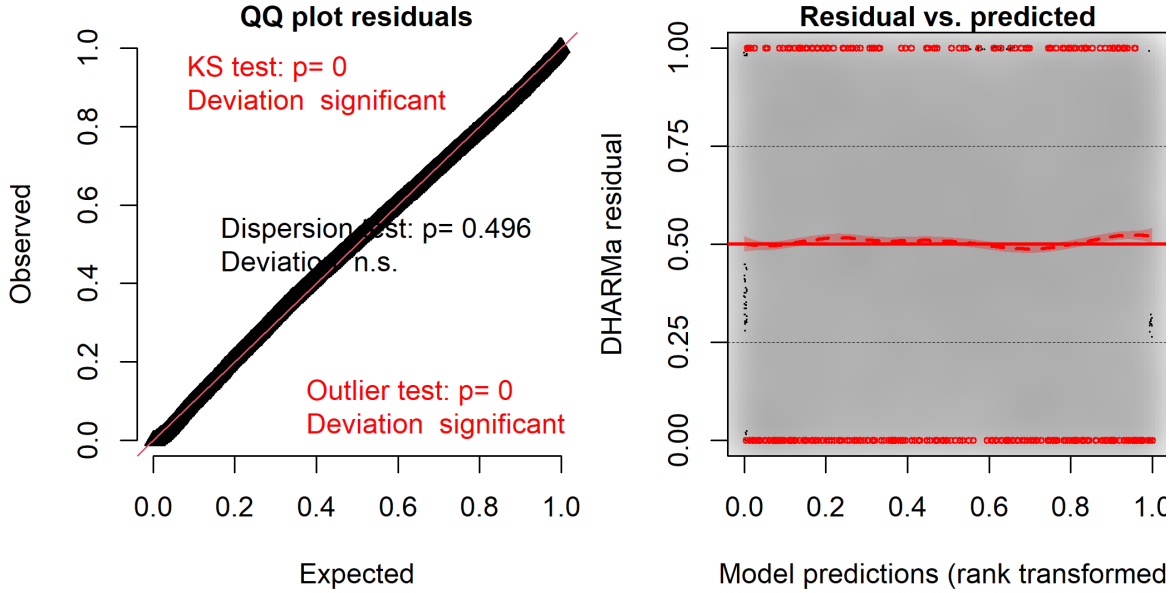

Figure 14. QQ-plot (left) and residual plot (right) for the second component of the GAMst model.

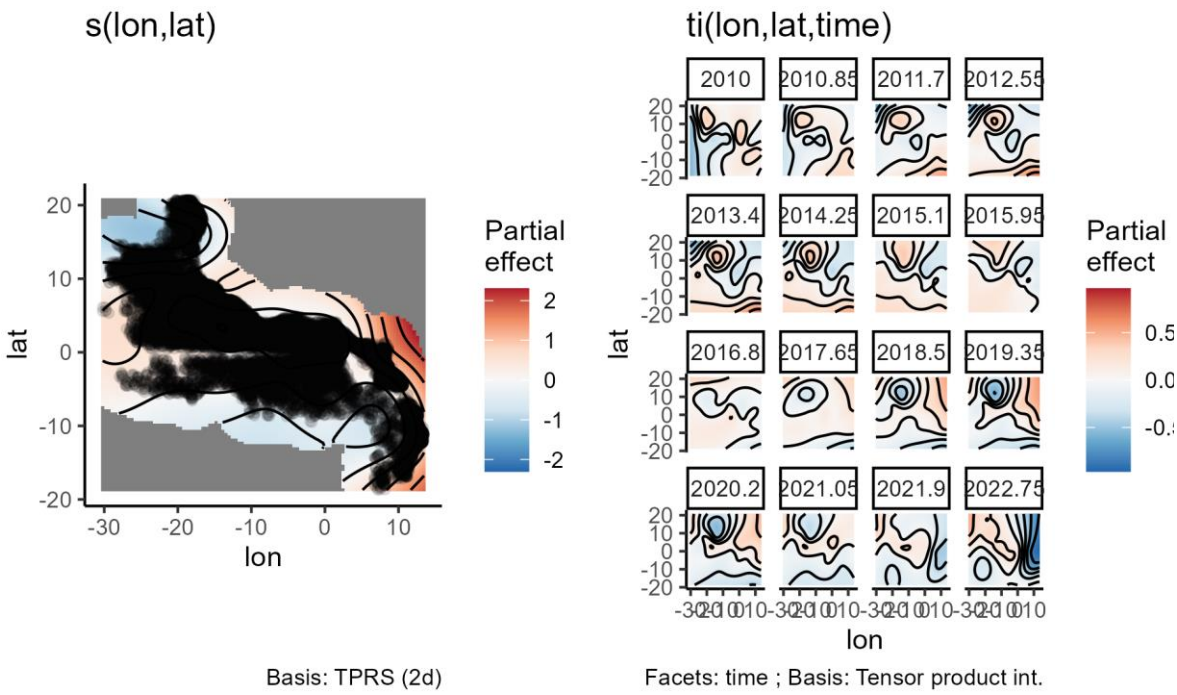

Figure 15. Product smooth terms for longitude and latitude interaction in the second component of the GAMst model.

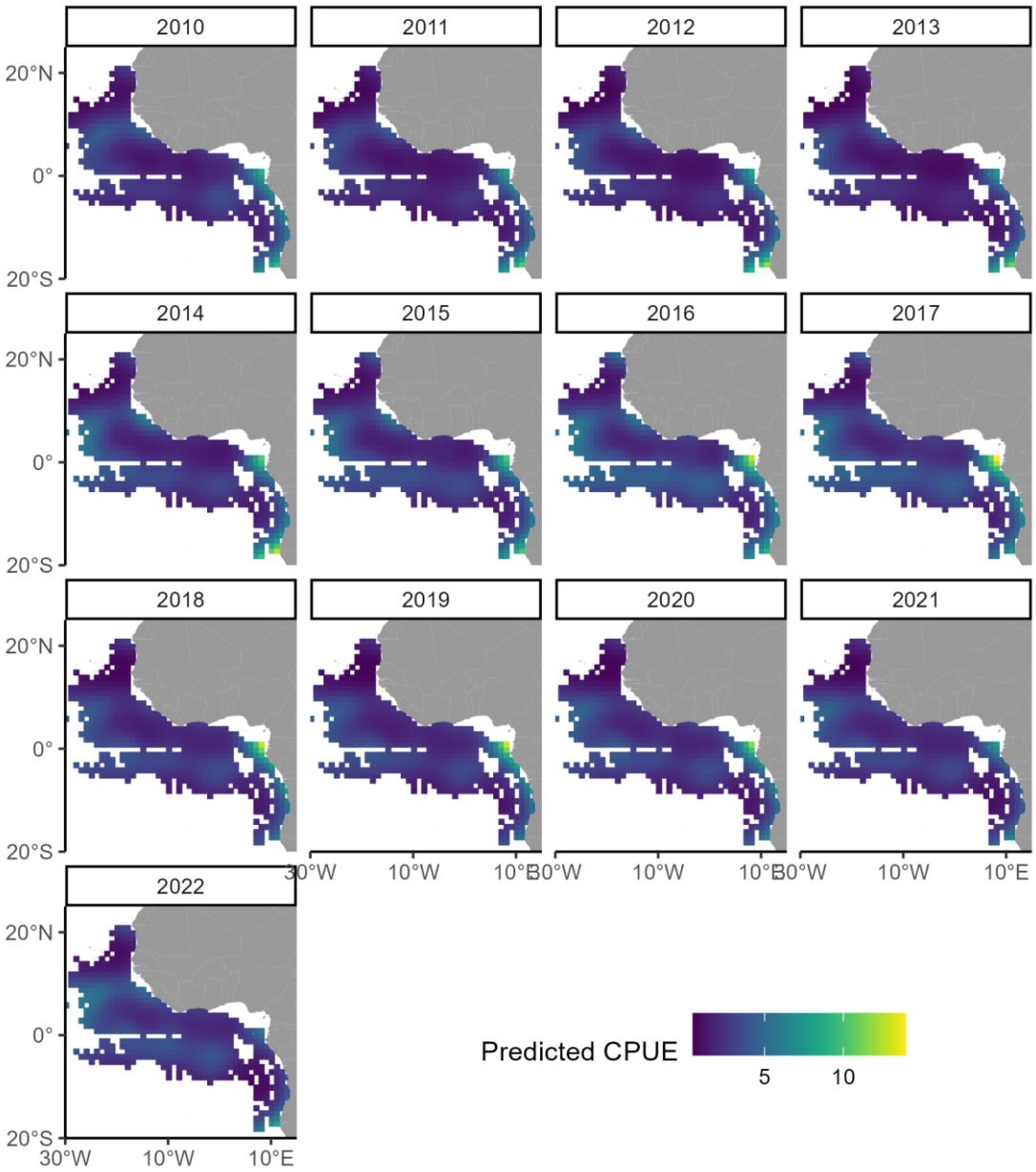

Figure 16. Predicted CPUE for each year-quarter-cluster combination by the GAMst model.

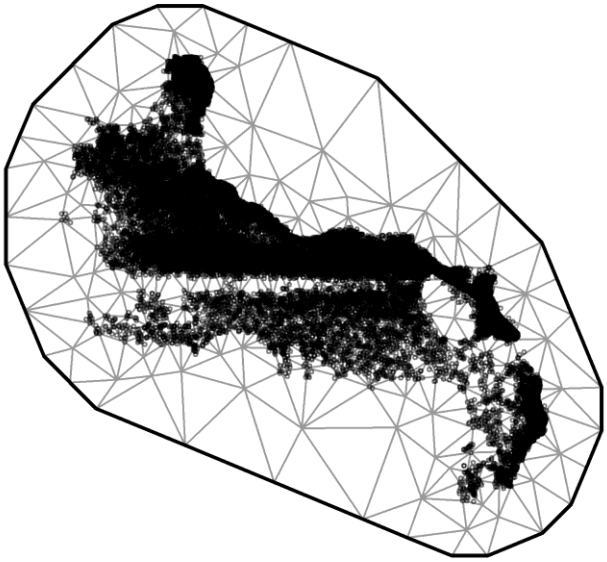

Figure 17. Mesh used in the spatiotemporal model.

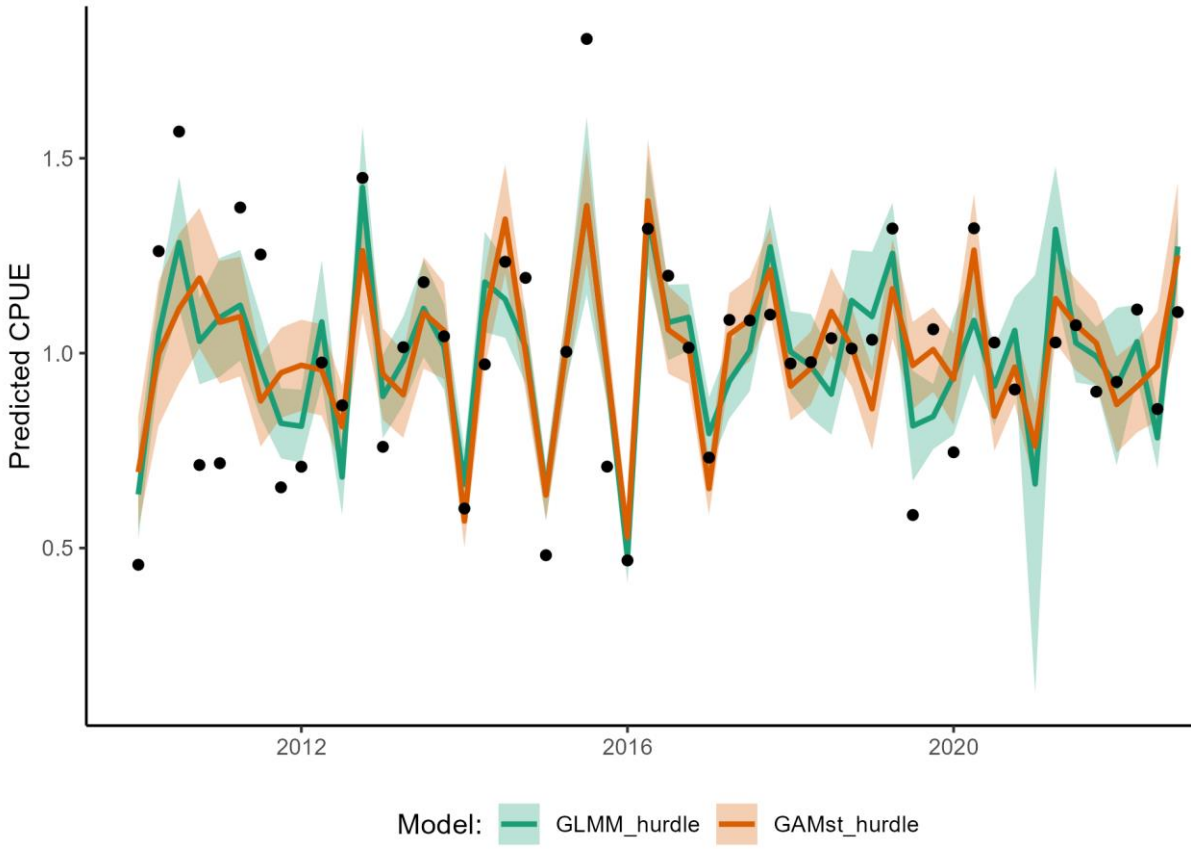

Figure 18. Predicted CPUE for year-quarter combination by the GLMM and GAMst models. Nominal CPUE is also shown in black points.