SPATIAL DISTRIBUTION OF MULTISPECIES LONGLINE CATCH PER UNIT EFFORT

Quang C. Huynh[1], Tom Carruthers[2]

SUMMARY

Longline catch per unit effort (CPUE) from the ICCAT Task 2 database were analyzed to develop spatial distribution models for six species: bigeye tuna (BET), blue marlin (BUM), blue shark (BSH), shortfin mako shark (SMA), swordfish (SWO), and white marlin (WHM). Historically, longline effort has increased linearly from the 1950s until 2000 followed by a decrease. Spatial distribution of effort expanded from the tropical ecoregion during the 1950s-1960s but has increasingly concentrated back to the tropical ecoregion during the 2000s. Spatial factor analysis is a multivariate ordination technique that can identify common spatial trends among species. CPUE from the Japanese and USA longline fleet were chosen to characterize species distribution based on their spatial coverage for the Eastern and Western Atlantic, respectively. The best model was a six-factor model that primarily modeled the distribution separately for each of the six species, but notable cross-correlations were estimated within taxa for marlin (BUM and WHM) and shark (BSH and SMA) species. The predicted spatial density can be used to inform stock distribution in multispecies spatial operating models.

RÉSUMÉ

La capture par unité d'effort (CPUE) de la base de données de la tâche 2 de l'ICCAT a été analysée afin de développer des modèles de distribution spatiale pour six espèces : le thon obèse (BET), le makaire bleu (BUM), le requin peau bleue (BSH), le requin-taupe bleu (SMA), l'espadon (SWO) et le makaire blanc (WHM). Historiquement, l'effort de pêche à la palangre a augmenté de façon linéaire entre les années 1950 et 2000 avant de connaître une diminution. La distribution spatiale de l'effort s'est étendue à partir de l'écorégion tropicale au cours des années 1950-1960, mais s'est de plus en plus concentrée vers l'écorégion tropicale au cours des années 2000. L'analyse des facteurs spatiaux est une technique d'ordination multivariée qui permet d'identifier les tendances spatiales communes entre les espèces. Les CPUE de la flottille palangrière japonaise et américaine ont été choisies pour caractériser la distribution des espèces sur la base de leur couverture spatiale dans l'Atlantique Est et Ouest, respectivement. Le meilleur modèle était un modèle à six facteurs qui modélisait principalement la distribution séparément pour chacune des six espèces, mais des corrélations croisées importantes ont été estimées au sein des taxons pour les espèces de makaires (BUM et WHM) et de requins (BSH et SMA). La densité spatiale prédite peut être utilisée pour informer la distribution des stocks dans les modèles opérationnels spatiaux multi-espèces.

RESUMEN

Se analizaron las capturas por unidad de esfuerzo (CPUE) de palangre de la base de datos de Tarea 2 de ICCAT para desarrollar modelos de distribución espacial de seis especies: patudo (BET), aguja azul (BUM), tiburón azul (BSH), marrajo dientuso (SMA), pez espada (SWO) y aguja blanca (WHM). Históricamente, el esfuerzo del palangre ha aumentado linealmente desde la década de 1950 hasta el año 2000, a lo que siguió un descenso del esfuerzo. La distribución espacial del esfuerzo se expandió fuera de la ecorregión tropical durante las décadas de 1950-1960, pero se ha concentrado cada vez más de nuevo en la ecorregión tropical durante la década de 2000. El análisis del factor espacial es una técnica de ordenación multivariante que puede identificar tendencias espaciales comunes entre especies. Se eligieron las CPUE de la flota palangrera japonesa y estadounidense para caracterizar la distribución de las especies basándose en su cobertura espacial para el Atlántico oriental y occidental, respectivamente. El mejor modelo fue un modelo de seis factores que modelaba principalmente la distribución por

[1] Blue Matter Science Ltd, Vancouver, British Columbia, Canada. Email: quang@bluematterscience.com
[2] Blue Matter Science Ltd, Vancouver, British Columbia, Canada.

separado para cada una de las seis especies, pero se estimaron notables correlaciones cruzadas dentro de los taxones para las especies de marlines (BUM y WHM) y tiburones (BSH y SMA). La densidad espacial predicha puede utilizarse para informar sobre la distribución de los stocks en modelos operativos espaciales multiespecies.

KEYWORDS

Ecosystem management, Catch/effort

1. Introduction

Fisheries catch per unit effort (CPUE) have frequently been used to describe abundance trends of exploited species. On a broad spatial scale, CPUE can provide information on general species distribution, whereas on a finer spatial scale, catch rates may be affected by a combination of abundance, availability, and fishing tactics. These fishing tactics include choice of hook type, location, depths fished, and season timing to search for target species and potentially avoid non-target species. While CPUE have analysed to inform single species stock assessments, many species overlap in distribution and there is information on broader habitat dynamics when CPUE is evaluated on a multispecies basis. These dynamics can have broader implications for management, especially with respect to bycatch species. For example, bycatch will be difficult to reduce without strong avoidance behaviour if there is significant overlap in encounter rates and catch rates with target species.

Statistical techniques that analyse multispecies data can quantify the extent to which abundance overlaps among species. Here we describe the result of the spatial factor analysis (SFA), a tool for joint species distribution modelling (Thorson et al. 2015), for the U.S. and Japanese longline fishery for tuna, swordfish, marlin, and shark species.

2. Methods

2.1 Data

The ICCAT Task 2 database (T2CE) consists of catch and effort records reported by ICCAT Contracting Parties. Catch and effort in the T2CE database are mostly aggregated by spatially by 5 x 5 degree coordinate squares and temporally in monthly time steps, but considerable portions of the database have historically use alternative stratification schemes, e.g., by 1 x 1 degree squares or quarterly time steps. Reporting coverage also by species and CPC.

Aggregation of catch and effort into time and spatial strata, in lieu of individual set by set data alongside important catchability covariates (hook type, depths fished, etc.), present difficulties in using the T2CE database for developing an index of abundance. Therefore, the approach here was to use the T2CE data to characterize spatial distribution of catch rates from the most recent decade (2010-2019) to minimize the impacts of changes in fishery behavior over time.

Catch rates were obtained from the database that met all of the following criteria, as denoted in R code:

```
FlagCode %in% c("JPN", "USA")
TimePeriodID <= 12
YearC %in% 2010:2019
Eff1Type == "NO.HOOKS"
SquareTypeCode == "5x5"
ifelse(DSetTypeID == "nw", CatchUnit == "nr", CatchUnit == "kg" | CatchUnit == "nr")
```

Some data are reported twice (in both numbers and weight for catch when DSetTypeID = "nw"). Therefore, the last criterion ensured that no duplicated records were used. While a large portion of the database has reported catch over 1x1 squares, almost all records over 2010-2019 are reported using 5x5 squares.

For this initial analysis, we used data from the Japanese and U.S. longline fleets. The Japanese fleet had a broad spatial distribution in the Eastern Atlantic which made it a good candidate fleet for describing abundance over a broad basin scale. The U.S. fleet was also used because it had a separate spatial footprint (Western Atlantic) distinct from the Japanese fleet. Both fleets report catch in the same units (numbers).

Catch rates for six species were analyzed: bigeye tuna (BET), blue shark (BSH), blue marlin (BUM), shortfin mako shark (SMA), swordfish (SWO), and white marlin (WHM). These represent a variety of taxa caught in the longline fishery. Some taxa such as marlin and sharks are more likely to be discarded compared to more commercially important species such as BET and SWO. To support the EcoTest framework for developing multispecies operating models, these species were identical to those in the operating model presented in Huynh et al. (2022). The CPUE data are presented in the Appendix.

2.2 Spatial factor analysis

Spatial factor analysis (SFA) has been proposed as a tool for joint species distribution modelling (Thorson et al. 2015). SFA is part of a broader technique of factor analysis, in which a multivariate dataset, i.e., multispecies longline CPUE, is modelled as a linear combination of latent variables ("factors"), common to all response variables, and scalars ("loadings"), corresponding to each categorical response variable. The spatial component to SFA is accommodated by modelling the factors as a Gaussian random spatial field. In addition to describing the spatial distribution of individual species, the estimated loadings matrix has a statistical interpretation and can provide estimate of the spatial correlation in distribution among species.

SFA is time-invariant and estimates spatial patterns within a single time step. Here, using CPUE over 2010-2019 would represent average distribution during that decade. In essence, SFA is a multivariate, spatial generalized linear mixed model (GLMM) where the random effects contribute to the predicted response variable for all categories, i.e., species. If there are fewer factors than categories in the parsimonious model, this implies that the categories are not fully independent and that there are important correlations.

The predicted catch is represented by $\boldsymbol{\Lambda}$, an I x J matrix, where the catch of species *j* in sample *i* is

$$\boldsymbol{\Lambda} = \exp(\boldsymbol{X\beta} + \widetilde{\boldsymbol{\Omega}}\, \boldsymbol{\psi}^{\boldsymbol{T}})$$

where $\boldsymbol{X}$ is a I x P design matrix, $\boldsymbol{\beta}$ is a P x J matrix of fixed effect coefficients, $\widetilde{\boldsymbol{\Omega}}$ is a I x K matrix of factors, and $\boldsymbol{\psi}$ is a J x K matrix of loadings. The design matrix here included intercepts by country and species (interpreted as catchability coefficients), quarterly effects (for seasonality effects, with no interactions with species or country), and log-effort (number of hooks) as an offset (fixed coefficient of 1).

The spatial field of factors $\boldsymbol{\Omega}$ is estimated from a grid and the corresponding value of the field for sample *i* is bilinearly interpolated based on geographical coordinates of the sample. Each factor $\boldsymbol{\Omega}_{\mathbf{k}}$ is a Gaussian Markov random field

$$\boldsymbol{\Omega}_{\mathbf{k}} \sim MVN(0, \boldsymbol{\Sigma})$$

where the covariance matrix is a Matern function where the correlation between two grid points is a function of their Euclidean distance. The marginal variance of $\boldsymbol{\Omega}_{\mathbf{k}}$ is fixed to one and the range parameter κ determines the distance at which the correlation is effectively close to zero.

The likelihood for the observed catch $\boldsymbol{C}$ (same dimensions as $\boldsymbol{\Lambda}$) follows a Tweedie distribution,

$$\boldsymbol{C} \sim \text{Tweedie}(\boldsymbol{\Lambda}, \boldsymbol{\phi}, \boldsymbol{p})$$

where $\boldsymbol{\phi}, \boldsymbol{p}$ are vectors of dispersion parameters corresponding to each country-species combination. The variance of the catch is a power function of the mean, i.e., $\text{Var}(\boldsymbol{C}) = \boldsymbol{\phi}\boldsymbol{\Lambda}^{\boldsymbol{p}}$.

The number of factors k (up to the number of species) is specified by the user, and model diagnostics can be used to identify the number of factors that generates the best model. Intuitively, the extent to which species distribution are correlated can inform k. For example, if all species have identical distributions, then one factor should be sufficient to fit the parsimonious model. On the other hand, if all J species have disparate distributions, then each factor describes distribution of a different species if K = J.

Given the structure of the random effects, the covariance of $\log(\boldsymbol{\Lambda})$ among species can be derived from $\boldsymbol{\psi}^{\boldsymbol{T}}\boldsymbol{\psi}$ and used to obtain the correlation. To ensure identifiability, elements above the diagonal of the loading matrix $\boldsymbol{\psi}$ is zero, with $\sum_{n=1}^{J} n - \sum_{n=0}^{J-K} n$ estimated parameters. Varimax rotation of $\boldsymbol{\psi}$ and $\boldsymbol{\Omega}$ is used be used to maximize contrast in the magnitude of the loadings among species (Thorson et al. 2015).

The SFA was implemented in the VAST R package (Thorson 2019) for spatiotemporal modeling. Parameters $\boldsymbol{\beta}, \boldsymbol{\psi}, \kappa, \boldsymbol{\phi}, \boldsymbol{p}$ are fixed effects and $\boldsymbol{\Omega}$ are random effects.

3. Results

3.1 Database summary

A historical summary of the number of T2CE records by Flag is reported in **Figure 1**. Reported effort (number of hooks) increases linearly over time from the 1950s until 2000 followed by a decrease over the next decade (**Figure 2**). Since 2014-2015, effort has been relatively constant if not slightly increasing. Effort was also described across the ecoregions proposed by the Ecosystems Sub-Committee (Juan-Jordá et al. 2022). Historically, the majority of longline effort has occurred in the Tropical Ecoregion (**Figure 3**). After initial expansion of the longline fishery from the Tropical Ecoregion into temperate waters, the effort has been increasingly concentrated back into the Tropical Ecoregion.

Coverage of catches in T2CE may be lower relative to the Task I catch database. For example, reporting of shark catches to the species level, e.g., for blue shark (BSH) and shortfin mako (SMA), appear to be limited prior to the early 2000s, and some species catch may not be reported (**Figure 4**).

3.2 Spatial factor analysis

A summary of the SFA model fits of varying complexity is presented in **Table 1**. The number of parameters in $\boldsymbol{\psi}$ and $\boldsymbol{\Omega}$ increases with the number of factors. An additional model fit was generated where the spatial field was independent for each species (k = 6 and $\boldsymbol{\psi}$ is a diagonal matrix). AIC selected the model with the most parameters (regardless of whether only fixed or both fixed and random parameters are considered), with six factors shared among species. This model had well-behaved residuals generated from the DHARMa R package (**Figure 5**).

The six species modeled here have geographically distinct distributions, as estimated in the six-factor model (**Figure 6**). BET and BUM abundance is high in tropical and subtropical regions, although the latter has higher density in the Caribbean compared to the rest of the Atlantic basin. BSH density is higher in subpolar waters of both hemispheres compared to the tropical areas. SMA density is higher in the NW Atlantic and SE Atlantic off the coast of Africa. SWO distribution is fairly panmictic. Finally, WHM density is higher in the Western Atlantic than in the Eastern Atlantic.

Since the model is of full rank, each factor generally follows the distribution of a single species, and the loadings matrix is sparse with highest magnitude in the diagonals (**Figures 7, 8**). Nonetheless, there are notable cross-correlations among species and all correlations were positive (**Figure 9**). BET generally was correlated to some degree with the other five species (correlation ranging between 0.2 – 0.4). There is high overlap in distribution (and catch rates) between BSH and SMA (correlation = 0.60).

The lowest correlations occur across taxa. For example, BSH is most correlated the most with SMA followed by BET and has low correlation (< 0.10) with the other four species. This appears to be driven by the more subpolar distribution of BSH compared to the tropical distribution of other species. Similarly, the two marlin species BUM and WHM have the third highest correlation (0.48) in this analysis, with the second between BUM and BET (0.5).

4. Discussion

Fishery CPUE that is aggregated across area and time strata (monthly time steps) can broadly inform species distribution as shown here as well as in previous studies looking at seasonal movements (Sculley and Die 2016). Broadly, SFA is an ordination technique that can be used to group species with similar characteristics together. A model fit with 2 factors provides a convenient number of dimensions for plotting to demonstrate clustering. For example, the SFA with 2 factors show a cluster between SMA and BSH and another for BUM and WHM based on the Euclidean proximity of the loadings (**Figure 10**).

With the available data, this type of analysis can inform spatial correlations in fishery catch rates (as a snapshot in time). There is a desire to understand how fishery catch rates change over time due to abundance and targeting. Spatial dynamic factor analysis (SDFA) expands upon SFA to estimates changes in abundance over time (Thorson et al. 2017). Initial attempts to use SDFA using the dataset in this analysis were not fruitful as catch rates were constant over time.

There is also loss of information when data are aggregated (along with the lack of catchability covariates). The covariates alongside fine scale CPUE data can provide the necessary contrast to explain high and low catches within time-area strata. With single species index standardization models that are not spatially explicit, time-area effects are typically attributed as habitat (abundance) effects and other covariates controlling catchability. In spatiotemporal models such as SFA and SDFA, the spatial and spatiotemporal fields are analogous to those time-area effects.

The current analysis only uses two fleets with little spatial overlap. By considering more fleets, there may be conflicting data with high and low catch rates within the same strata. Typically, this is believed to be caused by different targeting behavior among and within country fleets. With multispecies analysis, it may be possible to detect changes in targeting behavior. Catch ratios between species that change over time may be attributable to avoidance behavior. A fleet effect needs to be estimated, with one serving as the 'control' variable to estimate abundance while another fleet switches targeting in the same area.

The goal of this analysis was to use the available ICCAT data to inform stock and fleet spatial dynamics for multispecies operating models such as EcoTest to support ecosystem-based modeling (Huynh et al. 2022). While the Task II database does not have complete catch coverage, efforts to raise the catch and effort to the Task I database, considered to be more comprehensive, have shown similar trends in effort as seen here (for example, comparing Figure 2 vs. Figure 12 of Beare et al. 2016). Stock distribution can be assigned to broad spatial scales, e.g., western Atlantic vs. eastern Atlantic, tropical vs. temperate latitudes, with the area specification based on the species modeled and their home range, or by ecoregion (Juan-Jordá et al. 2022).

5. Acknowledgements

The authors acknowledge the Ocean Foundation for funding support.

6. References

Beare, D.J., Palma, C., de Bruyn, P., and Kell, L. 2016. A modeling approach to estimate overall Atlantic fishing effort by time-area strata (EFFDIS). SCRS/2015/206. Collect. Vol. Sci. Pap. ICCAT 72(8): 2354-2370.

Huynh, Q.C., Carruthers, T., and Taylor, N.G. 2022. Ecotest, a proof of concept for evaluating ecological indicators in multispecies fisheries, with the Atlantic longline fishery case study. SCRS/2022/104. Collect. Vol. Sci. Pap. ICCAT 79(5): 165-177.

Juan-Jordá, M.J., Nieblas, A.E., Hanke, A., Tsuji, S., Andonegi, E., Murua, H. 2022. Report of the ICCAT workshop on the identification of regions in the ICCAT convention area for supporting the implementation of the ecosystem approach to fisheries management. SCRS/2022/107. Collect. Vol. Sci. Pap. ICCAT 79(5): 178-211.

Sculley, M. and Die, D. 2016. Inferring seasonal movements of tropical tunas between regions in the eastern Atlantic Ocean from catch per unit effort. SCRS/2015/094. Collect. Vol. Sci. Pap. ICCAT 72(3): 747-785.

Thorson, J.T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) ackage in stock, ecosystem, habitat and climate assessments. Fisheries Research 210: 143-161.

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., and Winker, H. 2017. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Can. J. Fish. Aquat. Sci. 74: 1794-1807.

Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., and Kristensen, K. 2015. Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. Methods in Ecology and Evolution 6: 627-637.

Table 1. Summary of SFA models and spatial model with independent factors. The independent model has a diagonal loadings matrix, where each factor (spatial field) predicts the abundance of only a single species. AIC was calculated with either the number of fixed parameters or the number of total parameters (fixed and random effects).

Number of factors (k)	*Objective function*	*Fixed Parameters*	*Random Effects*	*AIC (fixed)*	*AIC (random + fixed)*
2	139068	53	462	8485	6637
3	136643	57	693	3644	2258
4	135691	60	924	1745	821
5	135116	62	1155	600	138
6	134815	63	1386	0	0
6 (Independent)	134874	48	1386	88	88

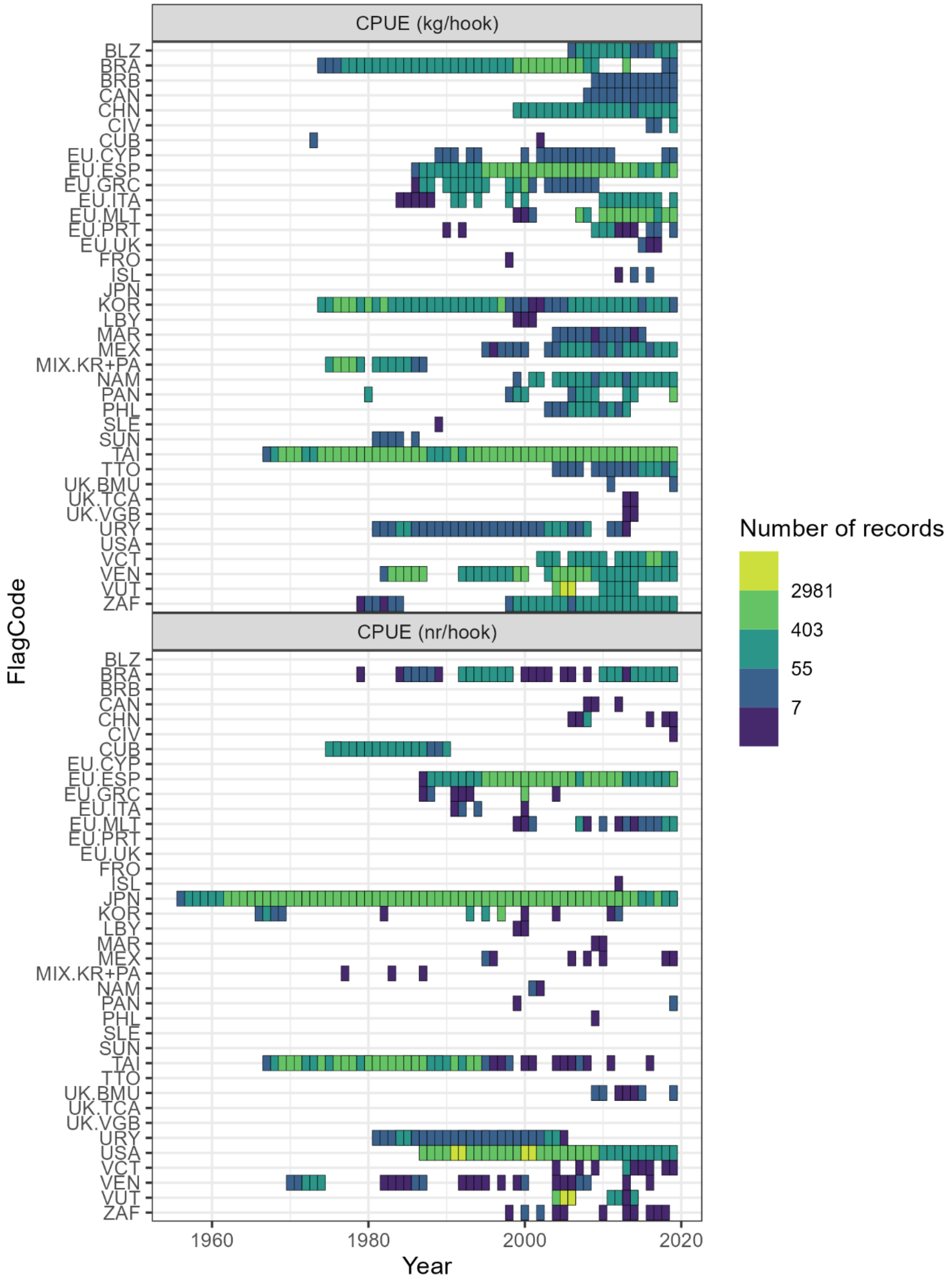

Figure 1. Number of longline records reported by FlagCode and units (either kg/hook or numbers/hook) in the Task 2 database. Only records that recorded catch by month and 5x5 or 1x1 squares are included here. This filter excluded approximately 8 percent of records.

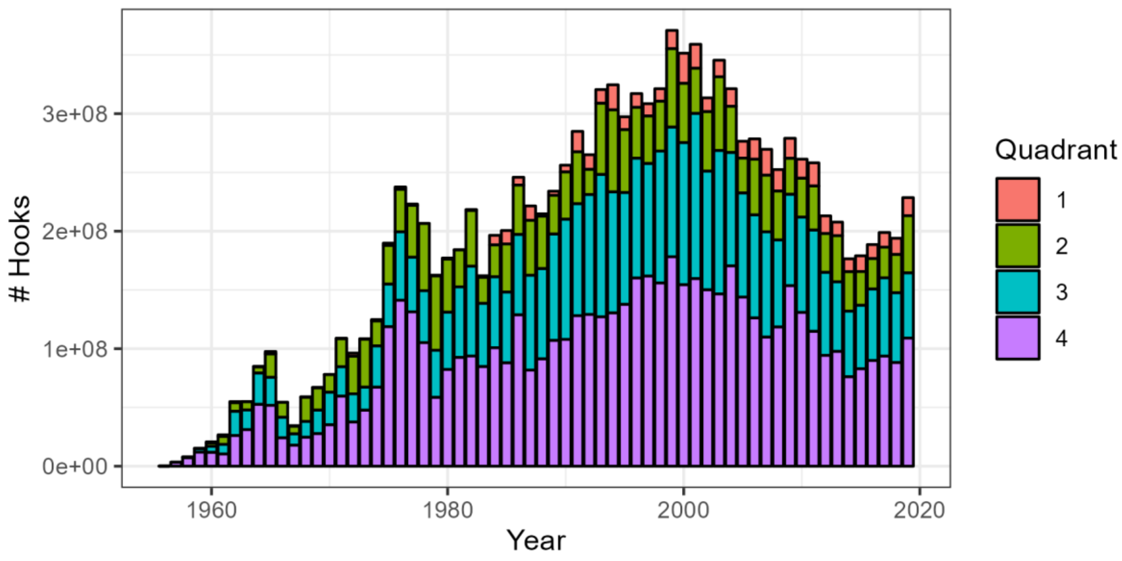

Figure 2. Total number of hooks in the longline gear by ICCAT quadrant in the Task 2 database.

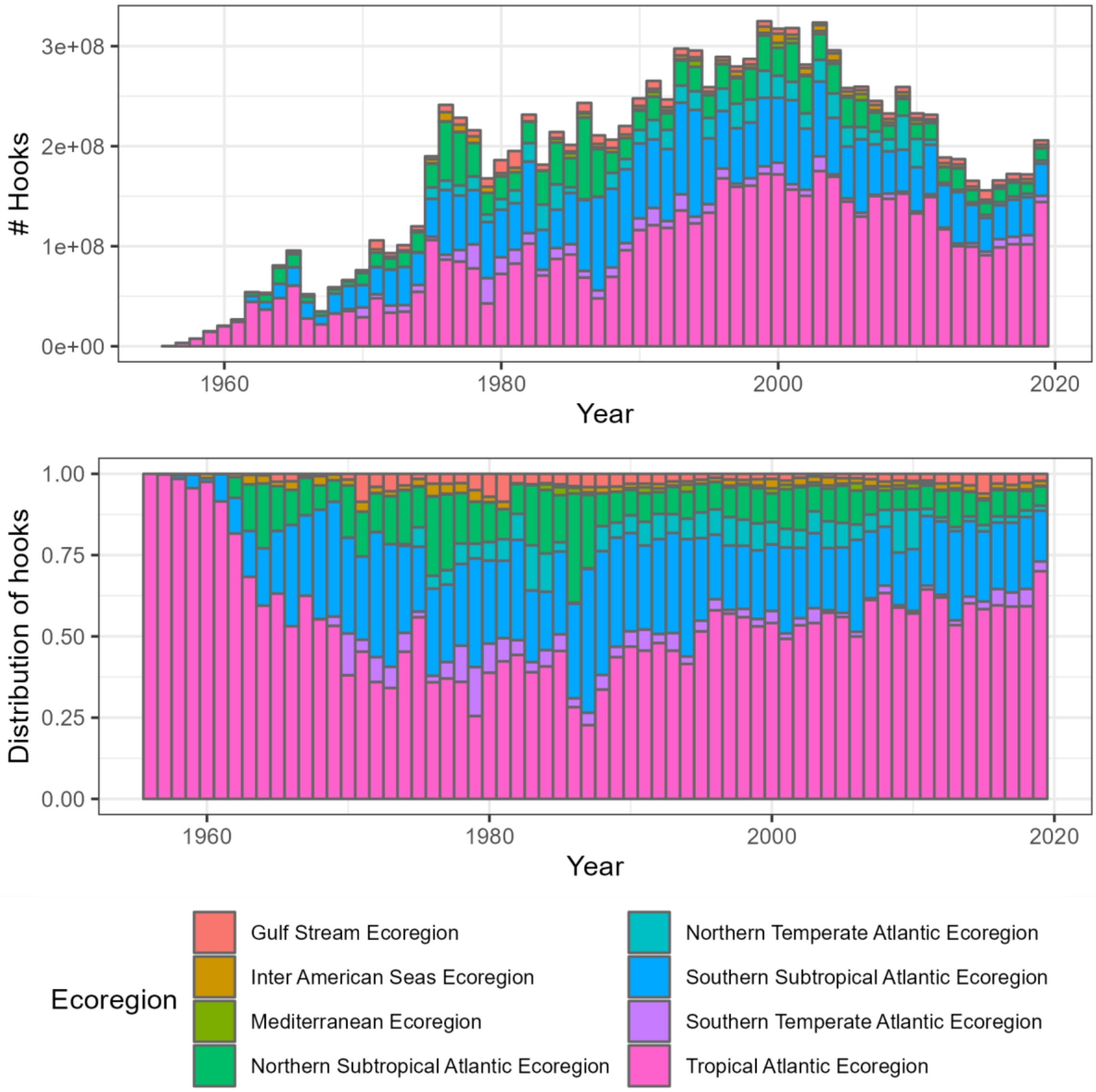

Figure 3. Total number of hooks (top panel) and distribution of hooks (proportion) in the longline gear by the proposed ecoregion () in the Task 2 database.

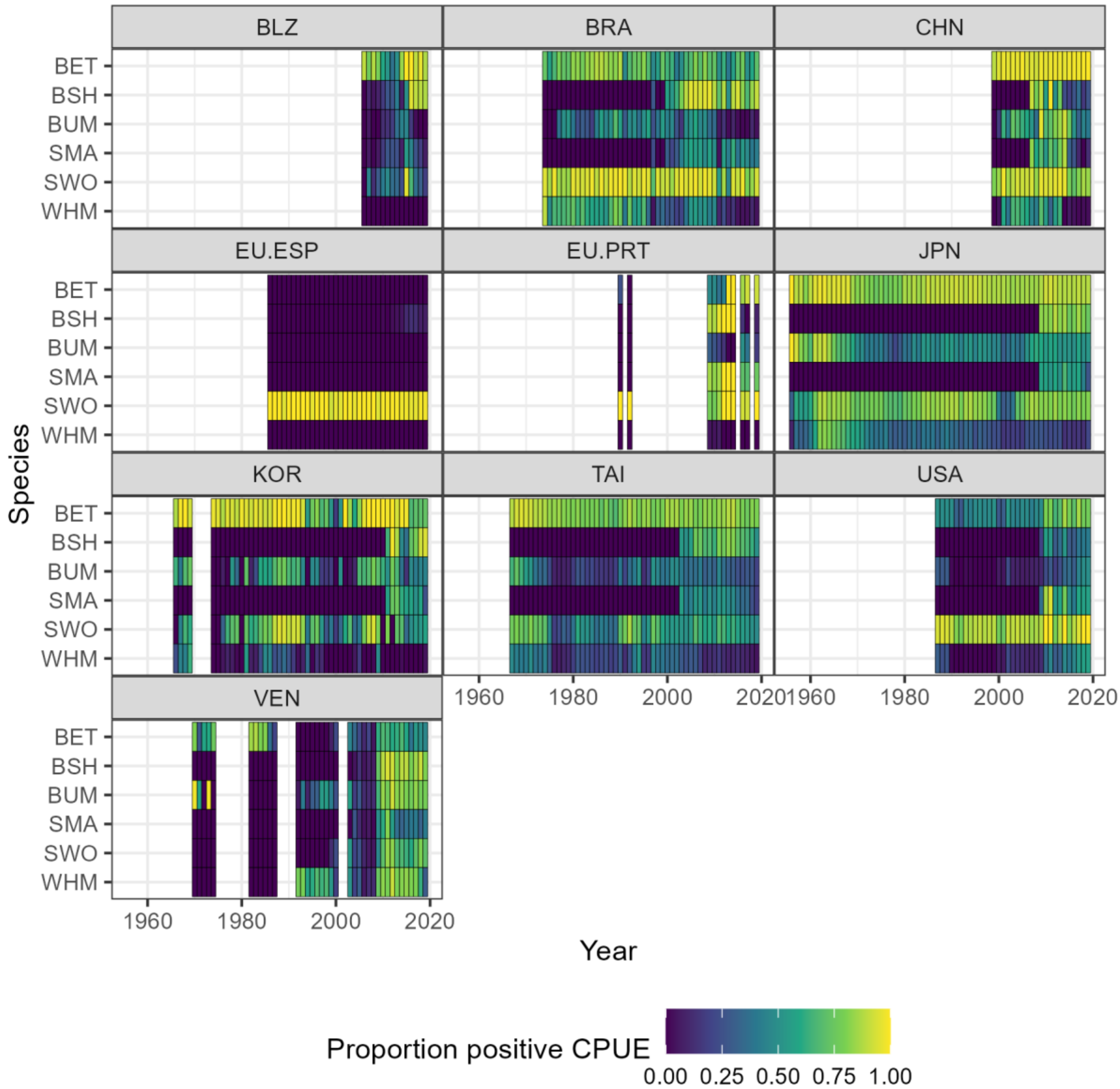

Figure 4. Proportion of positive catch in the Task 2 database for six species and ten flags with the highest number of hooks deployed since 2010. Periods of zero proportions should indicate no reporting for a species-flag combination.

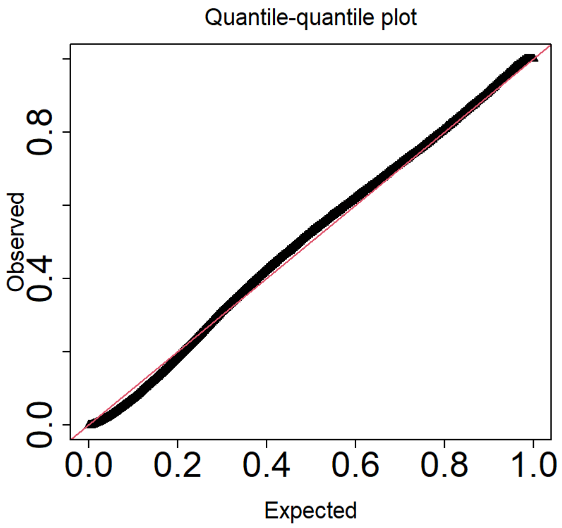

Figure 5. QQ-plot of residuals in the 6-factor SFA model. Residuals were generated by the DHARMa package by sampling the covariance matrix of fixed and random effects conditional on the maximum likelihood estimates.

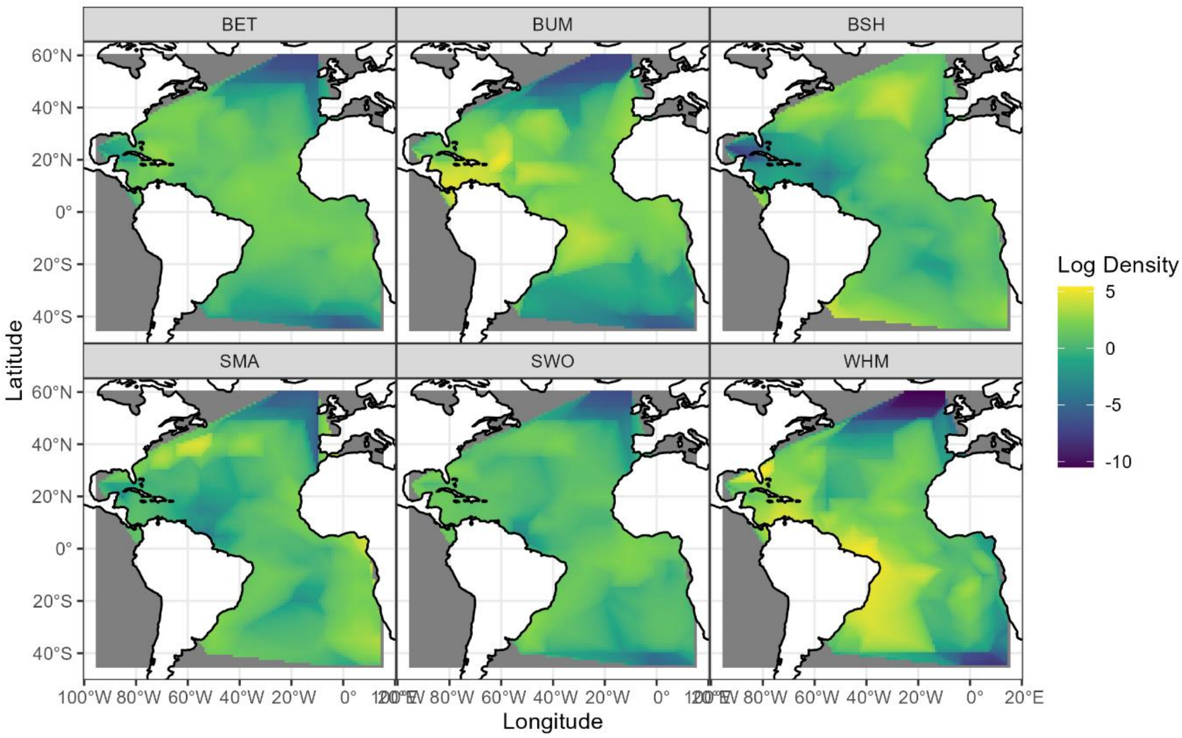

Figure 6. Spatial distribution of the log density of six species in the 6-factor SFA model. Density was predicted in the VAST model from a spatial grid of 321 knots. This figure was generated by interpolation across the Atlantic using the `interp` R package.

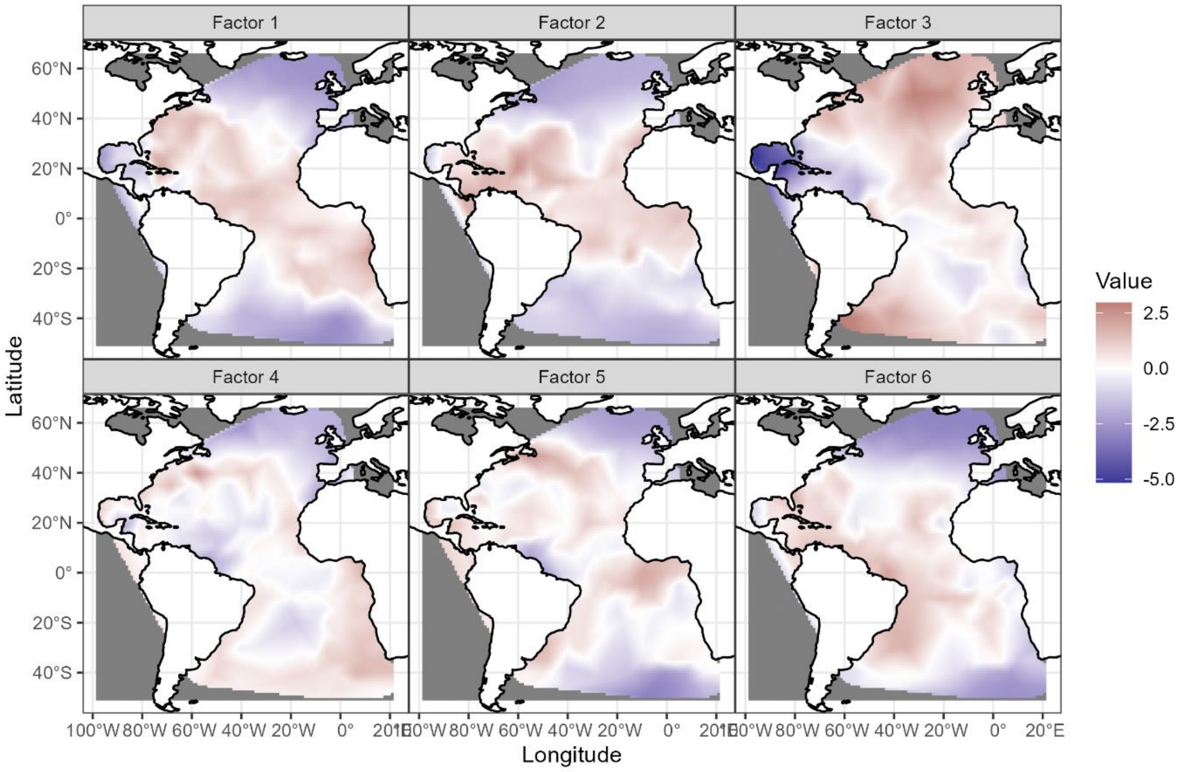

Figure 7. Estimated spatial factors in the 6-factor SFA model. The $\boldsymbol{\Omega}$ matrix was estimated from a spatial grid of 321 knots. This figure was generated by interpolation across the Atlantic using the `interp` R package.

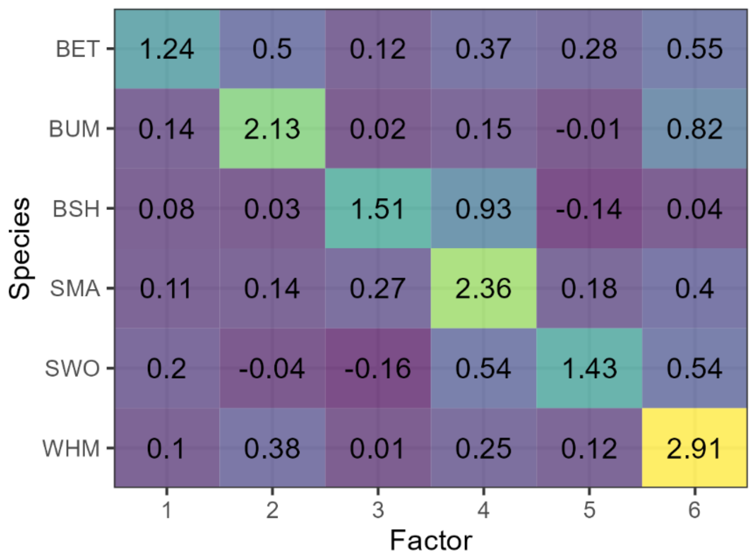

Figure 8. The loadings $\boldsymbol{\psi}$ matrix which relates the factors to species density. Factors shown here were rotated by varimax rotation to maximize contrast among species.

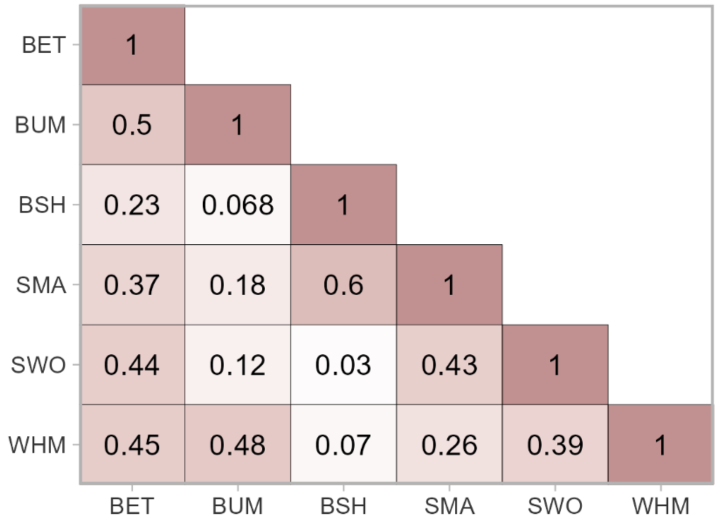

Figure 9. Correlation in species overlap, calculated from $\boldsymbol{\psi}^T\boldsymbol{\psi}$ from the 6-factor SFA.

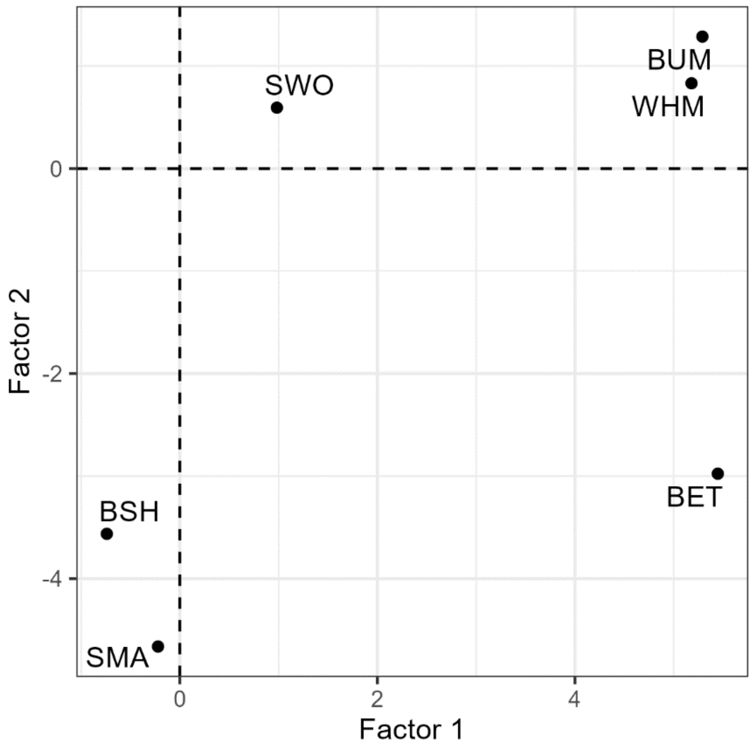

Figure 10. The loading matrix $\boldsymbol{\psi}$ from the 2-factor SFA shown as scatterplot matrix. SFA can be used as an ordination technique to group similar species together, e.g., for species complexes. Here, it is demonstrated that BSH and SMA are similar in distribution as well as with BUM and WHM.

7. Appendix

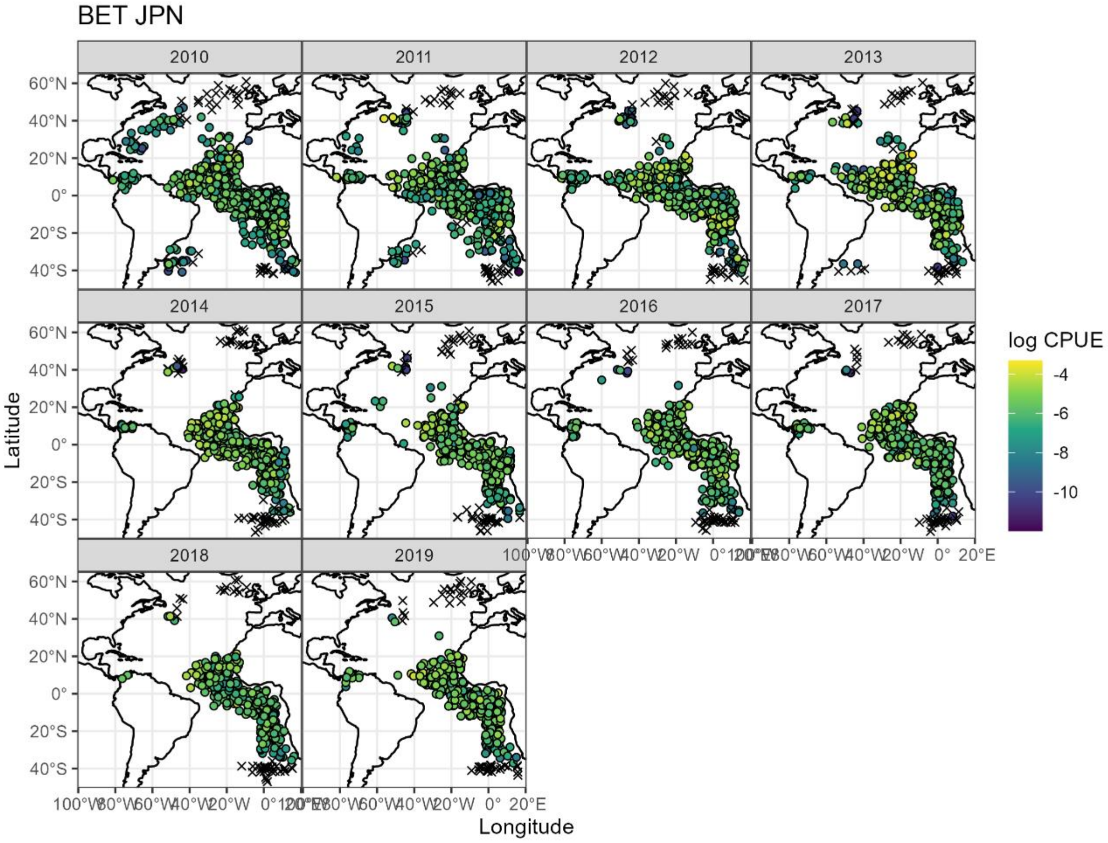

Figure A1. BET CPUE from the Japanese longline fleet. “X” corresponds to zero catch.

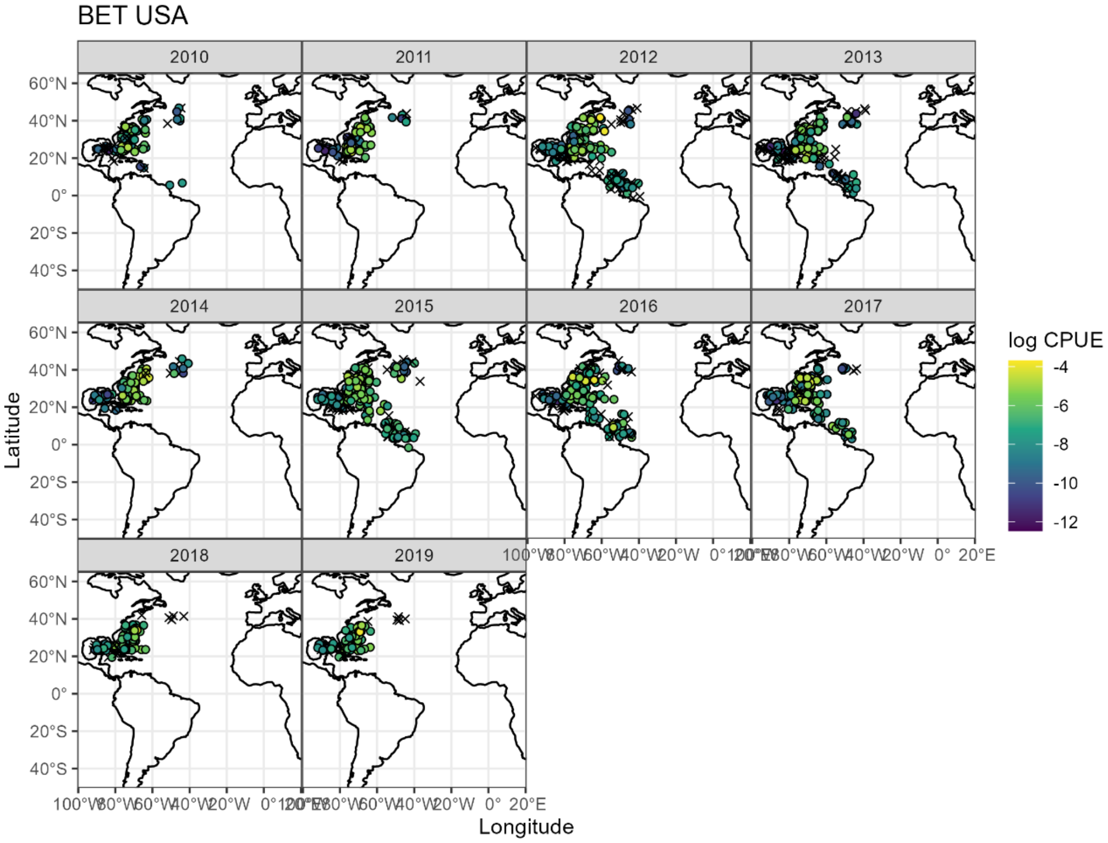

Figure A2. BET CPUE from the USA longline fleet. “X” corresponds to zero catch.

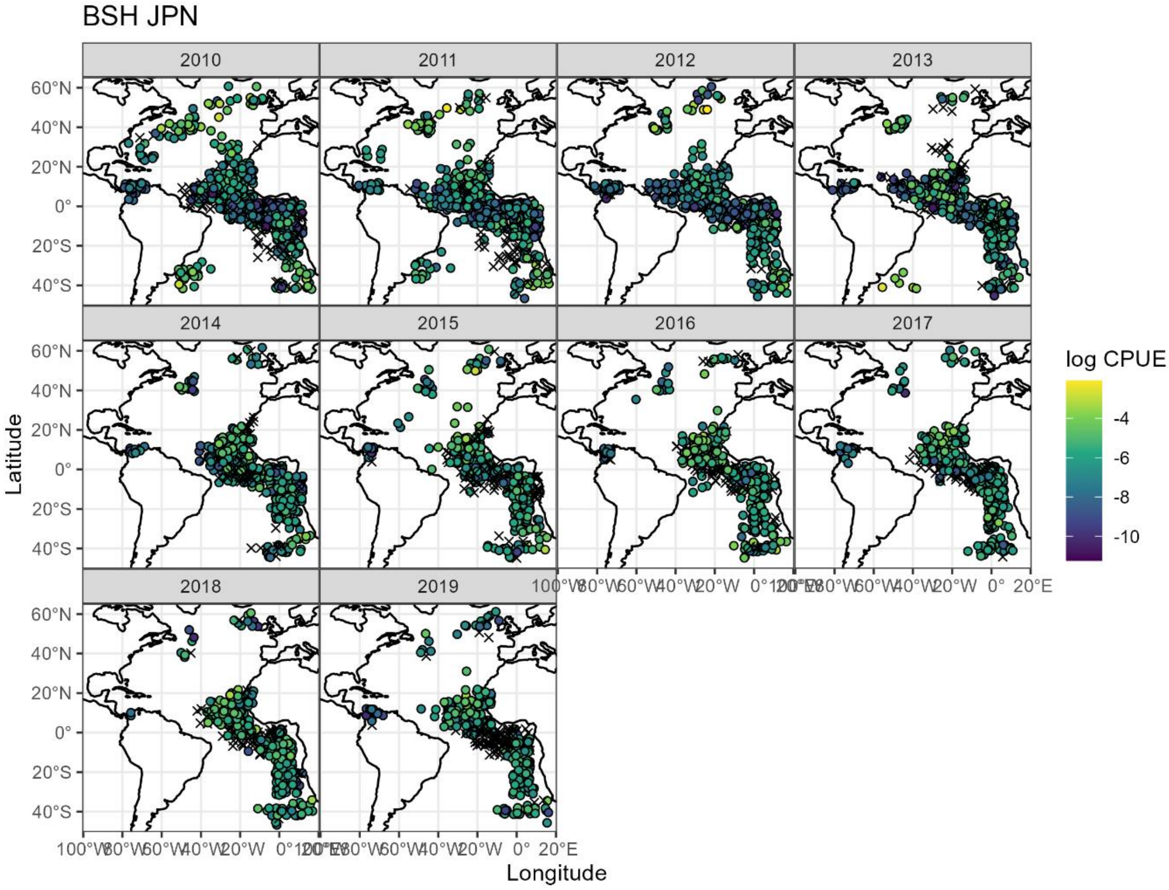

Figure A3. BSH CPUE from the Japanese longline fleet. "X" corresponds to zero catch.

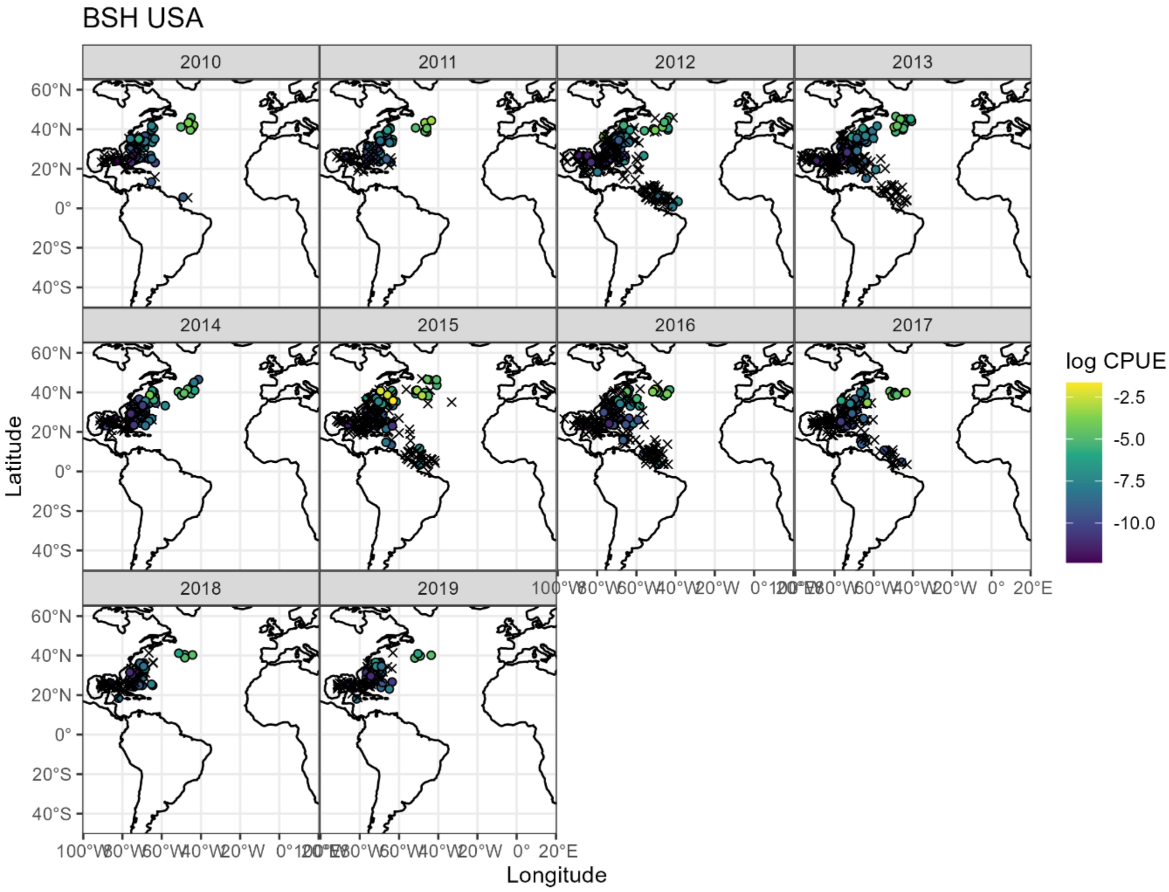

Figure A4. BSH CPUE from the USA longline fleet. "X" corresponds to zero catch.

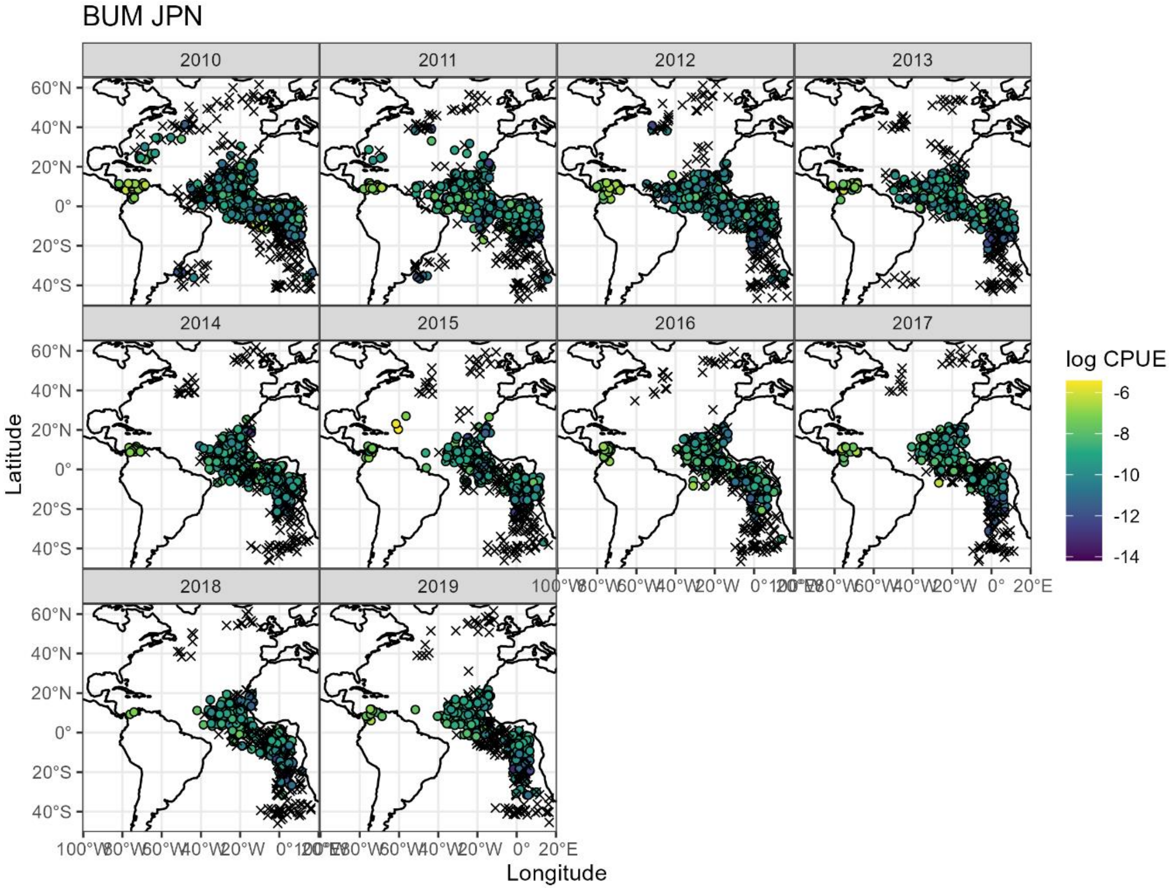

Figure A5. BUM CPUE from the Japanese longline fleet. "X" corresponds to zero catch.

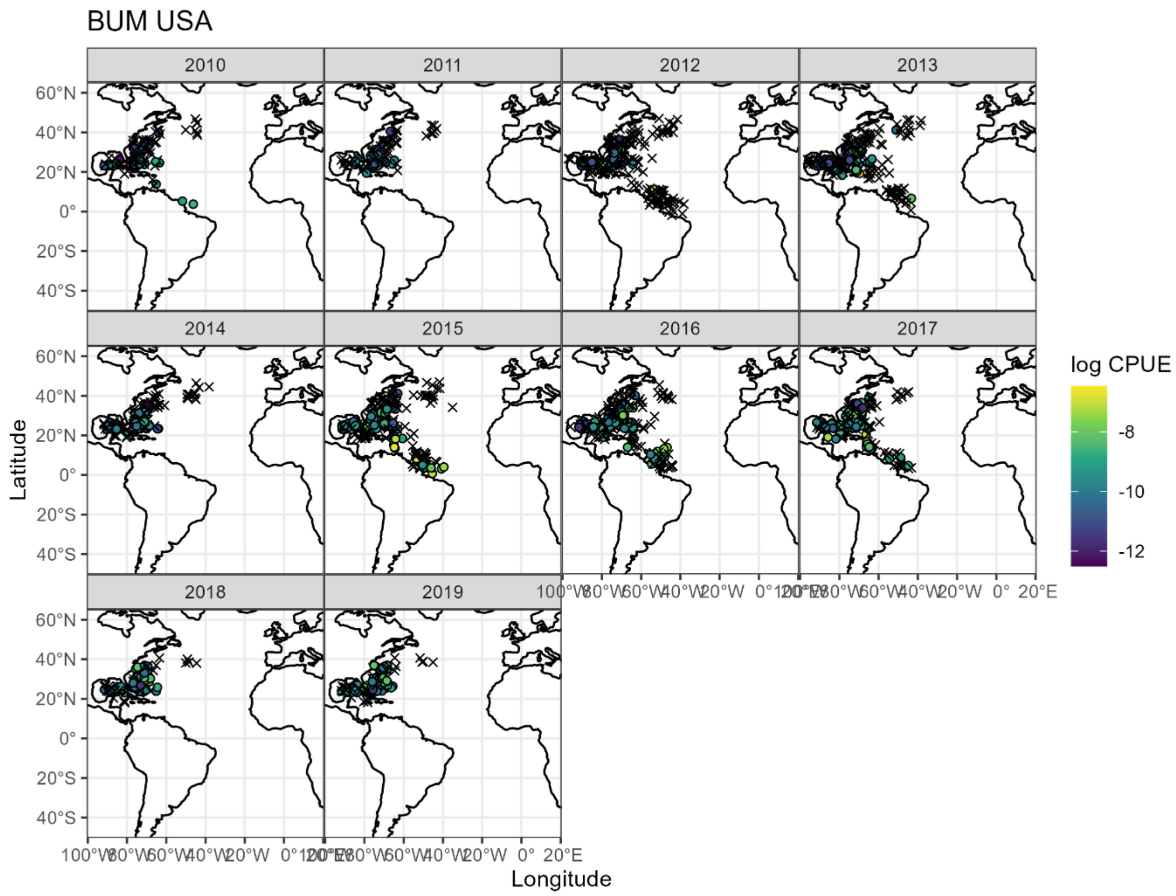

Figure A6. BUM CPUE from the USA longline fleet. “X” corresponds to zero catch.

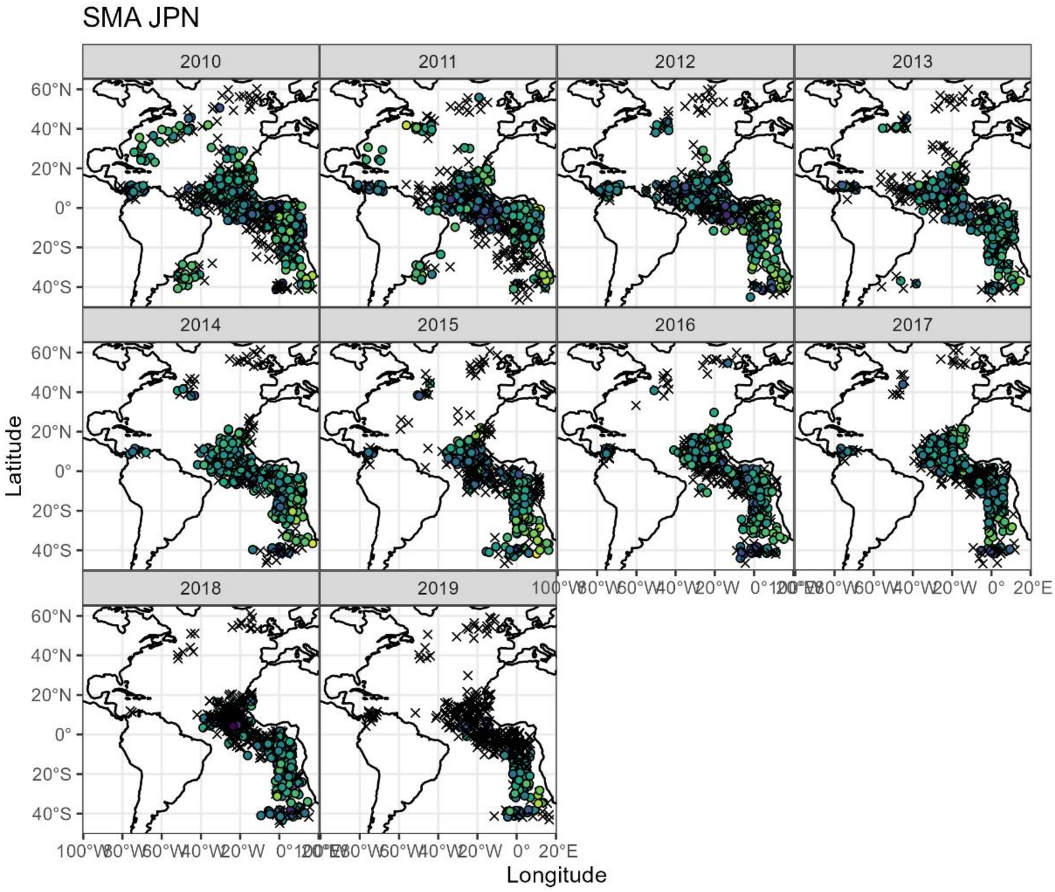

Figure A7. SMA CPUE from the Japanese longline fleet. "X" corresponds to zero catch.

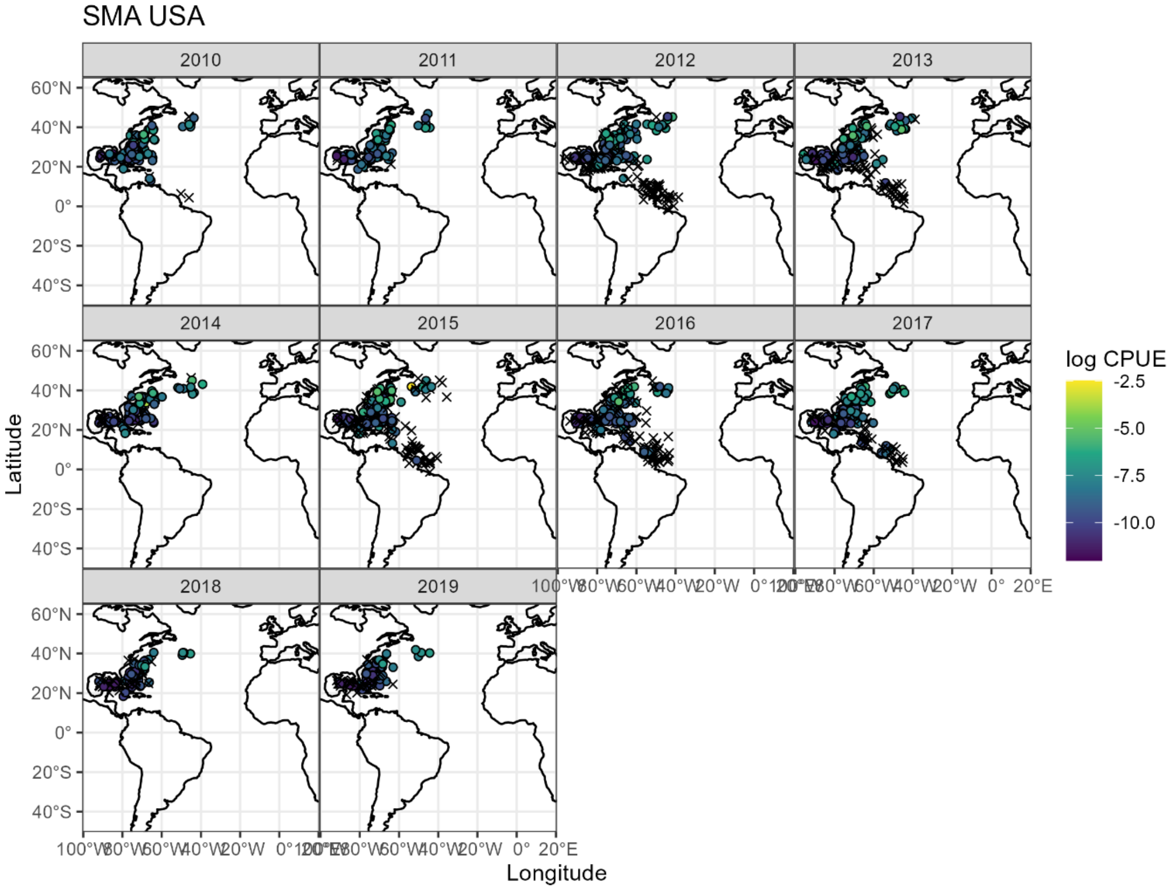

Figure A8. SMA CPUE from the USA longline fleet. "X" corresponds to zero catch.

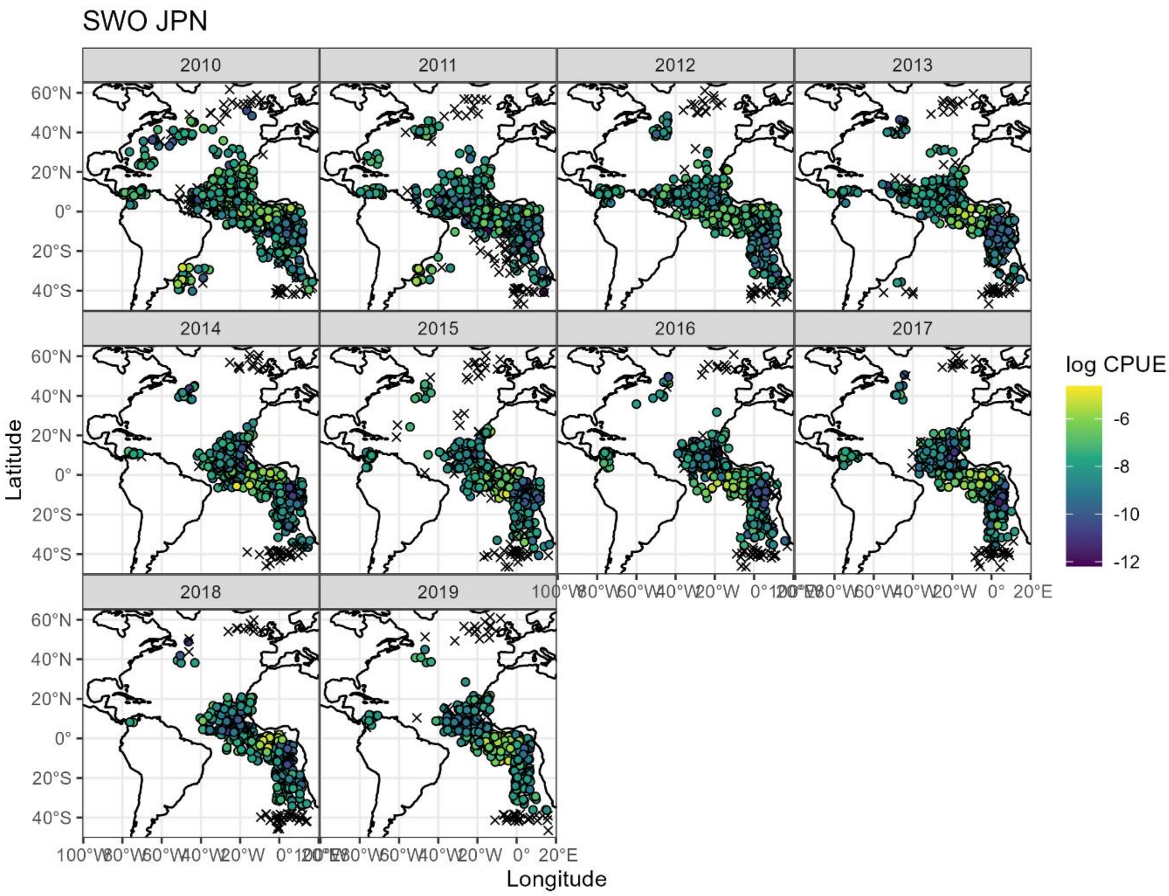

Figure A9. SWO CPUE from the Japanese longline fleet. "X" corresponds to zero catch.

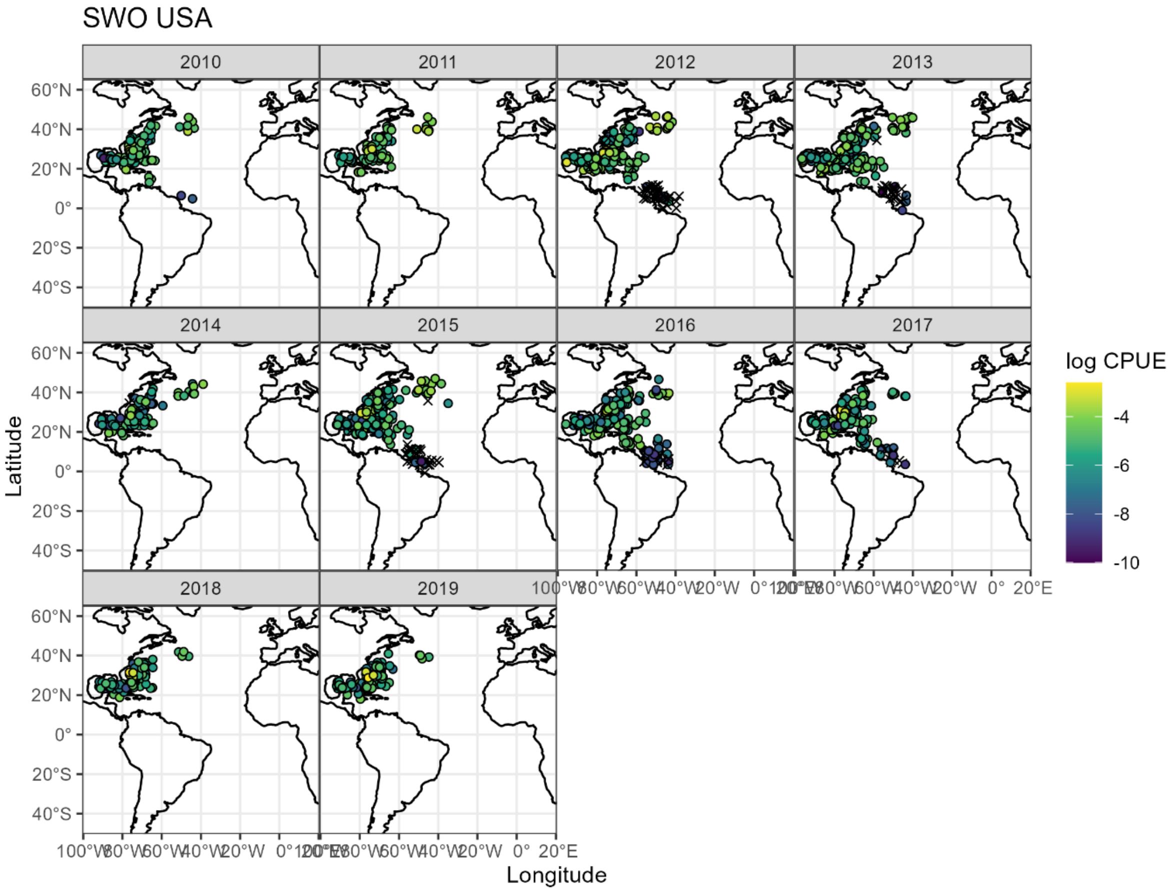

Figure A10. SWO CPUE from the USA longline fleet. “X” corresponds to zero catch.

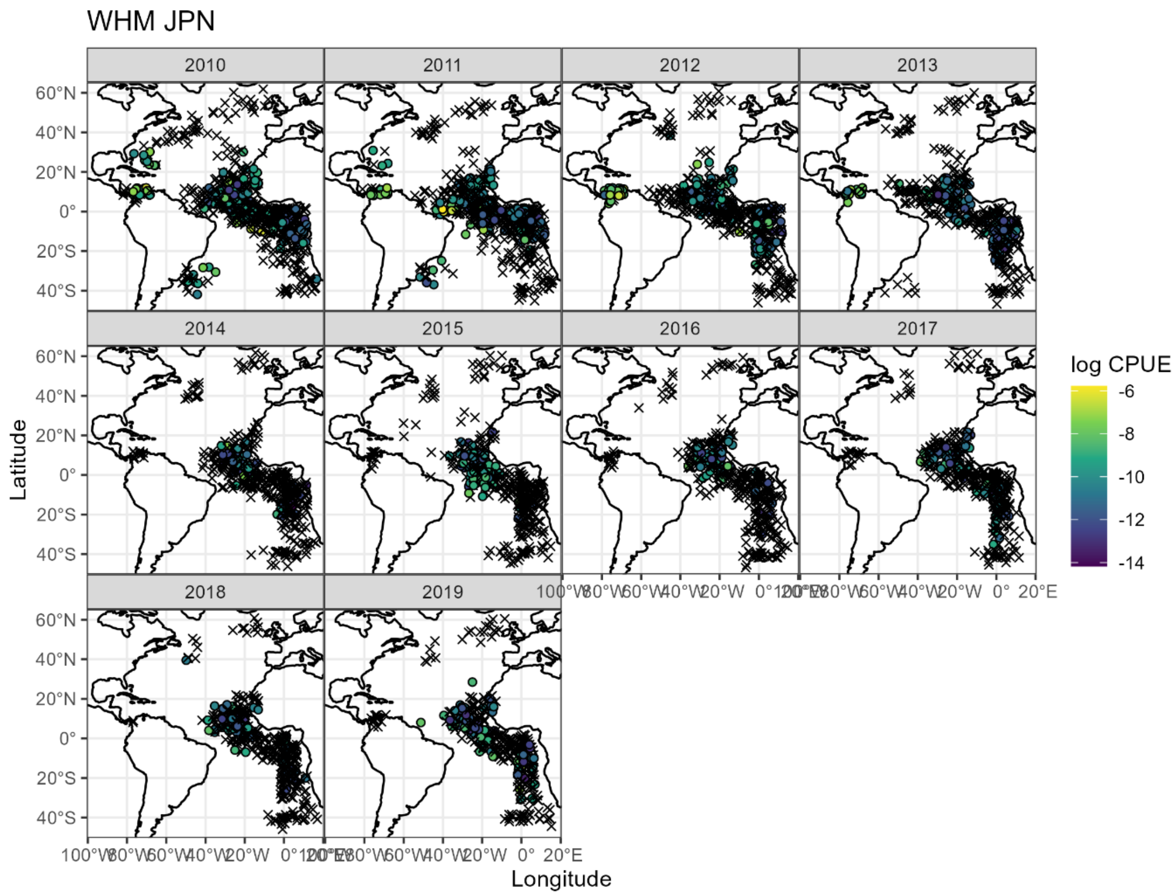

Figure A11. WHM CPUE from the Japanese longline fleet. “X” corresponds to zero catch.

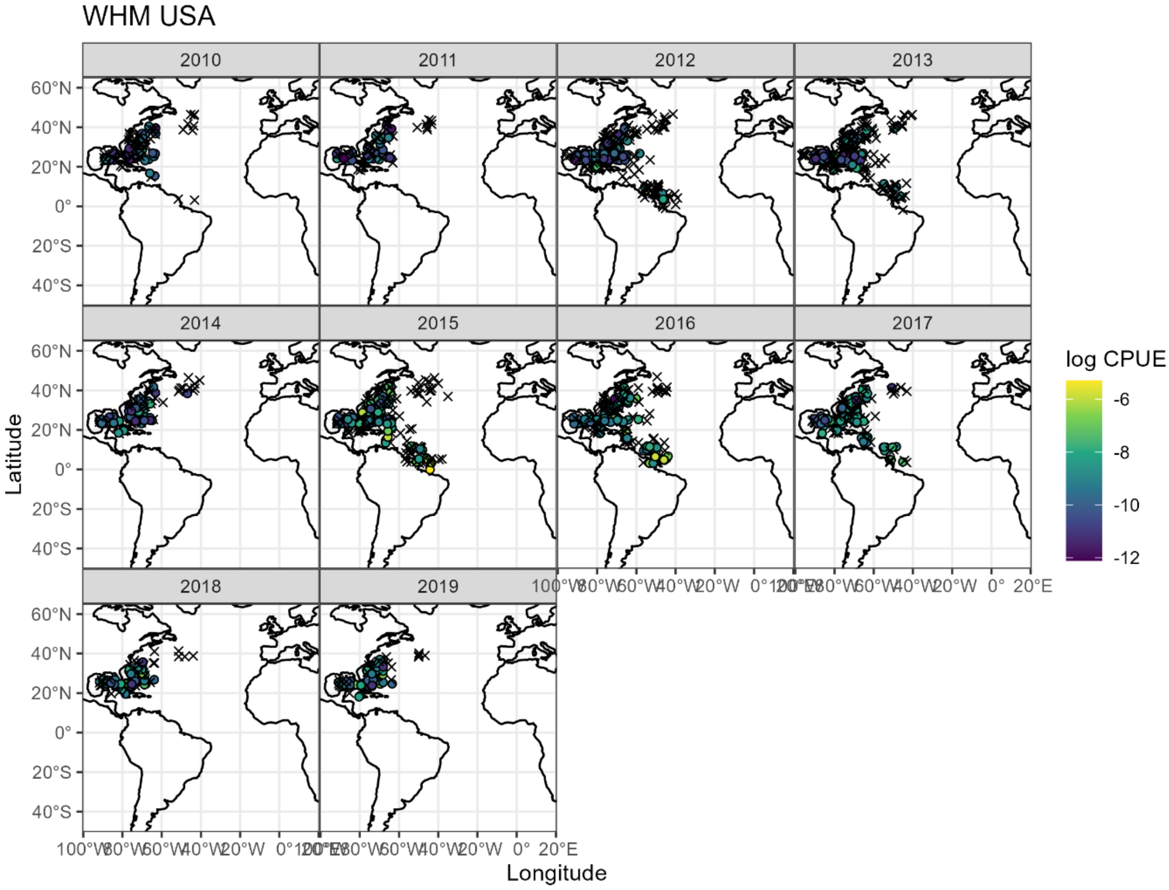

Figure A12. SWHM CPUE from the USA longline fleet. "X" corresponds to zero catch.