Standardized CPUE abundance indices for adult yellowfin tuna caught in free-swimming school sets by the European purse-seine fleet in the Indian Ocean, 1991-2022

David M. Kaplan[1,2,✉], Giancarlo M. Correa[3], María Lourdes Ramos Alonso[5], Antoine Duparc[1,2], Jon Uranga[4], Josu Santiago[4], Laurent Floch[1,2], Jose Carlos Báez Barrionuevo[5], Vanessa Rojo Méndez[6], Pedro Pascual Alayón[6], and Gorka Merino[4]

[1] MARBEC, University of Montpellier, CNRS, Ifremer, IRD, Sète, France
[2] Institut de Recherche pour le Développement (IRD), Av. Jean Monnet, 34203 Sète, France
[3] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi ugartea z/g, 48395 Sukarrieta (Bizkaia), Spain
[4] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Herrera Kaia, Portualdea z/g, 20110 Pasaia (Gipuzkoa), Spain
[5] Instituto Español de Oceanografía, Centro Oceanográfico de Málaga, Puerto Pesquero, s/n Apdo., 29640, Fuengirola, Málaga, Spain
[6] Instituto Español de Oceanografía, Centro Oceanográfico de Canarias, Calle La Farola del Mar No. 22, Dársena Pesquera, 38180 Santa Cruz de Tenerife, Spain

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

SUMMARY

Indian Ocean EU purse seine free-swimming school catches of adult yellowfin tuna (YFT) per unit effort for 1991-2022 were standardized using a "Delta" modeling approaches consisting of three components. The first component modeled the detection rate of FSC per unit search time; the second component models the binomial probability that adult YFT are present in a set; and the third component models adult YFT biomass per set given presence assuming a log-normal distribution. Components were modeled using general additive mixed-effects models (GAMMs) including spatial, temporal, vessel, YFT quota and environmental explanatory variables. Predictions were made on a standard grid encompassing core fishing areas by quarter and robust estimate uncertainties were developed using prediction intervals. Results indicate adequate model fits. Estimated adult YFT abundance shows an upswing in estimated abundance since ~2019, consistent with a response to the implementation of a YFT quota in 2017.

KEYWORDS

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

1. Introduction

In this paper, we estimate standardized catch per unit effort (CPUE) time series for the period 1991 to 2022 for adult yellowfin tuna (YFT, *Thunnus albacares*). The indices are derived from free-swimming school (FSC) set catch-effort data of the European purse seine (PS) fleet operating in the Indian Ocean. These FSC adult YFT time series (i.e., including size categories 2 and 3, i.e., 10-30 kg and >30 kg, respectively) complement floating object (FOB) juvenile YFT time series developed elsewhere (Correa et al. 2024).

The methodology for developing these standardized CPUE indices is derived from that presented in Guéry et al. (2021) based on a Delta modeling approach (i.e., an approach splitting abundance into several different components), but with a number of notable modifications and/or improvements. First of all, we take into account non-linear impacts of predictors on dependent variables by using a general additive mixed-effect modeling

(GAMM) approach as opposed to the general linear mixed-effects modeling (GLMM) approach used previously. The model involves three components similar to those in Guéry et al. (2021): (1) a first component modeling the density of FSC schools based on the assumption that the number of FSC schools (including both positive and null sets) encountered per unit of vessel search time is proportional to the density, (2) a second component modeling the probability that adult YFT will be present in a given non-null FSC set, and (3) a third component modeling the biomass of adult YFT caught per non-null FSC set given that adult YFT were present in the set. The predictor variables used in each component have been improved to include factors accounting for the YFT quota in place since 2017 (IOTC Secretariat 2016), the catchability impacts of mixed-layer depth (MLD) and to use a more appropriate definition of search time in the model for Component 1 (see Section 2 for more details).

Predictions from the component models are combined to produce final standardized abundance indices. They are carried out on a standard spatio-temporal grid encompassing spatial areas accounting for the majority of fishing effort, and standardize for the effects of vessel capacity, vessel identifier, mixed layer depth (MLD) and YFT quota. Robust estimates of uncertainty are included based on prediction intervals, i.e., the uncertainty in new observations, as opposed to confidence intervals, i.e., the uncertainty in mean predictions, thereby providing a realistic representation of the likely uncertainty that would have resulted from fishing with a fixed spatio-temporal pattern of effort. Results from the three components are aggregated on annual and quarterly time scales, both individually and in combination, and compared with nominal (i.e., non-standardized) CPUE indices, thereby providing a detailed view of the different components that lead to final standardized abundance indices and how they differ from nominal CPUE indices.

Note that 2023 T3-corrected catch-effort data is expected to become available before the IOTC WPTT, and we hope to integrate this into our standardized CPUE time series for the final stock assessment.

2. Methods

2.1 Purse-seine catch-effort data

To derive European purse-seine standardized CPUEs for the YFT stock assessment, T3-processed logbook data from the French and Spanish purse seine fleets targeting tropical tunas in the indian Ocean from 1991 to 2022 analysed. Though fishery data exist prior to 1991, species composition information from the first years of the fishery are considered to be less reliable and mixed layer depth (MLD) environmental data was not available before 1993 (see Section 2.2 for details). Raw logbook data (Level0) 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) and species composition (based on port size sampling and the T3 methodology; see Pallarés & Petit 1998, Pianet et al. 2000) to generate the Level 1 logbook database used in this paper. Logbook data for the French and the Spanish fleets were provided by the Exploited Tropical Pelagic Ecosystems Observatory (IRD-Ob7) and the Spanish Institute of Oceanography (IEO), respectively. As port sampling data for the Spanish fleet from 2020 is not currently available, catch composition of 2020 logbook data was corrected using the T3 procedure based on port samples from 2021.

Our initial dataset was filtered to remove aberrant data and/or to ensure data homogeneity, limiting the data using the following set of filters applied in the order indicated:
- French and Spanish fishing vessels in the Indian Ocean for the period 1991-2022
- The minimum set of vessels representing at least 95% of the vessel-days in the dataset, all vessel activity types (e.g., fishing, searching) confounded. The objective of this condition was to remove vessels that were only active for a short period of time. After applying this condition, 67 purse-seine vessels remained in the dataset, each representing between 1,485 and 9,644 days of activity.
- Any vessel fishing days with unrealistic numbers of sets (> 5 sets) were removed. This condition removed 62 vessel-days out of a total of 265,504 vessel-days.
- Entire vessel fishing days with at least one problematic vessel activity type (e.g., equipment failures or indications that the vessel was not actively trying to fish), fishing sets of unknown school type or unknown vessel activity type were removed. The specific activity type codes removed were: 4 (route sans veille), 7 (avarie), 10 (en attente), 11 (transbordement en mer), 12 (transb. depuis un senneur), 13 (transb. vers un

senneur), 14 (chavire la poche), 15 (au port), 19 (transb. vers un canneur), 20 (transb. depuis un canneur). These conditions removed 23,747 vessel-days out of a total of 265,442 vessel-days (8.9% of vessel-days). Note that 0.4% of non-fishing French logbook data from the 1990's lack the vessel activity type field. In these cases, the data was treated as if the non-fishing vessel activities were "problematic" and therefore the corresponding vessel-days were removed from the data set.

- Entire vessel fishing days were removed if climatological MLD could not be estimated from the Copernicus Global Ocean Physics Reanalysis (see Section 2.2 for details) for any of the vessel activities on that day. Visual inspection of the positions of these activities for which MLD could not be determined indicated that they corresponded either to activities very close to the shoreline or a small number of position errors in the data. This condition removed 15,361 vessel-days out of a total of 241,695 vessel-days (6.4% of vessel-days).

The final dataset used for analyses consisted of 312,292 fishing activities corresponding to 240,653 vessel-days, 80,812 FSC sets and 139,939 FOB sets.

For a small number of fishing sets (146 sets out of a total of 220,751 sets), set duration was absent or indicated as zero, which was considered anomalous. In these cases, zeros were replaced by the average non-zero set duration for the given year. Given this corrected set duration data, search time for a given vessel-day was calculated as the number of daylight hours (i.e., sunset - sunrise) minus the total duration of all sets for that vessel-day. For 109 vessel-days, the calculated search time was ≤ 0 by no more than 0.29 hours. In these cases, search time was modified to be equal to the smallest non-zero value to permit use of search time as an offset in models. Additionally, for component 1 of the Delta model, a "FSC" search time was calculated as the number of daylight hours minus the total duration of FOB sets, again replacing any negative values with the lowest non-negative value (see Section 2.3 for details).

For a small number of fish trips (i.e., combination of vessel identifier and landing date), no fishing sets were carried out during the trip (109 trips out of a total 9,049 trips), prohibiting us from calculating the fraction of sets on FSC (considered for use as a measure of targeting in models; see Section 2.3). In these cases, the fraction of FSC sets was set to the fraction of FSC sets for the entire fleet (France or Spain) for the year of the fishing trip in question.

2.2 Environmental covariates

Changes in the depth of the mixed layer are known to affect the catchability of surface fisheries, including purse seine, as surface-dwelling tunas mostly gather above the thermocline (Green 1967, Cayré & Marsac 1993, Bertrand et al. 2002). A deep thermocline may therefore decrease the vulnerability of tuna schools to purse seining. As such, MLD was included as a potential covariate in CPUE standardization models. MLD variables were extracted from the Copernicus Global Ocean Physics Reanalysis (GLORYS) based on the current available real-time global forecasting CMEMS system, having a 0.083°x0.083° (1/12th of a degree) spatial resolution, daily timestep and covering the time period 1993-01-01 to 2022-12-31 (https://doi.org/10.48670/moi-00021, Accessed 2024-03-18). Daily data were downloaded from CMEMS for the tropical Indian Ocean. A daily climatology was created from these daily data as the average for each day of the year over the 30 years of data (the 366th day of the year being ignored for leap years). MLD "anomalies" were calculated as the difference between the daily MLD values and the day-of-the-year climatology (repeating the 365th day twice for leap years). Daily MLD data, day-of-the-year climatological MLD values and MLD anomalies were associated with catch-effort data based on the day of each fishing activity and the CMEMS 0.083°x0.083° grid cell encompassing the fishing activity location. As GLORYS MLD data are not available for 1991-1992, for these years, MLD was assumed to be equal to the climatological value and MLD anomaly was set to zero.

2.3 Modeling approaches

Our standardized CPUE abundance indices for adult YFT catch in purse seine FSC sets for the EU fleet from 1991 to 2022 are based on the combination of predictions from three component models following the general approach of Guéry et al. (2021):

- **Component 1:** The encounter rate of FSC schools in units of (null or positive) FSC sets per hour of "FSC" search time (see below for the definition of "FSC" search time)

- **Component 2:** The probability of a non-null (AKA positive) FSC set having a non-zero biomass of adult YFT. This binomial approach was warranted by the fact that most FSC sets either consisted almost entirely of adult YFT or zero YFT (for SKJ FSC sets)
- **Component 3:** The biomass of adult YFT per non-null FSC set given that adult YFT were present in the set (in units of tonnes per set)

These components are modeled using GAMMs with the following distribution families:
- **Component 1:** As the dependent variable is count data (number of sets per vessel-day), it could in principle be modeled assuming the data follows a Poisson distribution. Nevertheless, to avoid problems of overdispersion, we model it using assuming a Tweedie distribution.
- **Component 2:** Presence-absence of adult YFT per set is modeled as a binomial process.

- **Components 3:** Biomass per set is modeled as a log-normal process, as has been done elsewhere (e.g., Kaplan et al. 2023, Kaplan & Tolotti 2023).

An overview of the dependent and potential independent variables for standardization models is provided in Table 1 and Table 2, respectively, and detailed descriptions of each of the model components are provided below.

2.3.1 Component 1: FSC schools encountered per unit search time

For modeling the FSC school encounter rate, fishing and non-fishing vessel activities were grouped by vessel-day, and the number of FSC sets, the search time, the mean geographical position and the mean climatological and anomalous MLD were calculated. When counting the number of sets per vessel-day, both null and non-null sets were included as both were considered indicative of the presence of a fish school.

These vessel-day data were used to estimate the following model (in notation of the `mgcv` package of R):

```
gam(formula = num_sets_fsc ~
      te(lon, lat, by = quarter, k = 18) +
      te(year, month, k = c(20, 11), bs = c("cr", "cc")) +
      s(capacity, k = 13) +
      s(mld_anom, k = 13) +
      yft_quota * country +
      quarter +
      s(vessel, bs = "re") +
      offset(log(search_time_fsc)),
    family = "tw", data = D)
```

Note that vessel was modeled as a random effect (as indicated by `bs="re"`) and the amount of search time was included as an offset (log-transformed for a model with log link function). For search time, we initially considered using the standard definition of the number of daylight hours minus the total duration of all sets on a day, but this variable was found to be negatively correlated with the number of FSC sets for a given vessel-day (as both FSC and FOB sets reduced the search time), which is not appropriate for an offset and produced an anomalous number of large residuals in the model fit. Instead, we developed an alternative "FSC" search time measure (`search_time_fsc`) calculated as the number of daylight hours minus the total duration of FOB sets.

One concern with modeling the number of FSC sets per vessel day including null sets is that this may double count some FSC schools. If multiple attempts to fish on the same FSC are carried out (e.g., a null set is followed shortly afterward by a successful non-null set on the same school), then this would lead to over-estimation of the rate of encounter of FSC schools. To assess the magnitude of this effect, first we observed that only 6.2% of all vessel-days had multiple FSC sets, at least one of which was null (representing 30.0% of vessel-days with FSC sets). Next, focusing on data since 2000 for which logbook entries generally represent a single fishing set, the distances between pairs of successive FSC sets on a given vessel-day were calculated, separating pairs into null/null, null/non-null, non-null/null and non-null/non-null pairs. If a vessel attempts to fish twice on the same set, the first time being a null set and the second time being a non-null set, then we would expect there to be an excess of short distances for null-non-null pairs (and potentially null-null pairs). However, no significant differences were observed in the

distributions of distances for the four classes of pairs. Therefore it was considered that multiple attempts to fish the same FSC are a relatively rare occurrence in the data. As such, no specific procedure was used to eliminate multiple attempts to fish the same FSC.

Since 2017, a quota has been implemented for YFT in the Indian Ocean, potentially leading fishers to change behavior when encountering YFT-dominated FSC. To account for this potential effect in the standardization process, a logical predictor was developed that was true for all fishing activities on or after 2017-01-01. This predictor was included in all component models individually and in interaction with the predictor `country` (in case French and Spanish fleets responded differently to the quota; note that the model formula term `yft_quota * country` expands to include both direct and interaction terms). We initially hoped to also include this predictor variable in interaction with `month` to account for an reduction in targeting of YFT-dominated FSC at the end of the year when vessels approach their quota, however, it was found that this made models unstable due to the redundancy between this effect and the year-month tensor product. As such, this interaction between the quota and month was not included, but we explore model residuals as a function of month before and after the quota to assess the potential for this effect.

2.3.2 Component 2: Probability of presence of adult YFT in positive FSC sets

When modeling the probability of encountering adult YFT in a FSC set, only positive sets were considered as null sets would typically indicate failure to capture the fish school (as opposed to a true absence of adult YFT). Adult YFT was considered present if there was any non-zero biomass of adult YFT for the set, though in practice that vast majority of FSC sets were either almost entirely adult YFT (e.g., >90% adult YFT) or 0% adult YFT (for SKJ-dominated FSC).

The following model was used for component 2:

```
gam(formula = yft_adult_present ~
      te(lon, lat, by = quarter, k = 18) +
      te(year, month, k = c(20, 11), bs = c("cr", "cc")) +
      s(capacity, k = 13) +
      s(mld_anom, k = 13) +
      yft_quota * country +
      quarter +
      s(vessel, bs = "re"),
    family = "binomial", data = D, weights = num_pos_sets)
```

Note that vessel was modeled as a random effect (as indicated by `bs="re"`) and the number of positive sets that correspond to a given logbook line was used as a weight in the binomial model. Other effects were as in component 1.

For logbook entries corresponding to multiple positive sets (principally found in data from the 1990's), in principle the response variable, proportion of presences, could be different from zero or one (e.g., for a logbook entry corresponding to two sets, the proportion could be 0, 0.5 or 1 depending on whether adult YFT was present in 0, 1 or 2 of the sets, respectively). Nevertheless, we only considered presence and absence (i.e., 1 and 0) because upon examining the fraction of catch biomass that was adult YFT, we found that the vast majority of fractions were either very close to 100% or very close to 0%, with few in between as would occur if logbooks recorded multiple sets, some of which were negative for adult YFT (but positive for something else, typically SKJ). We hypothesize that this is because captains rarely recorded two sets with very different species compositions in a single logbook entry.

2.3.3 Component 3: Catch of adult YFT per positive FSC set

When modeling catch biomass of adult YFT per positive FSC set, only positive sets for which the biomass of adult YFT was >0 were considered as null sets would typically indicate failure to capture the fish school and this model was to be combined with component 2 indicating presence of adult YFT (i.e., this model was for abundance given presence). Adult YFT biomass was estimated as the sum of catches for YFT categories 2 and 3.

The following model was used for component 3:

```
gam(formula = log(yft_adult) ~
    te(lon, lat, by = quarter, k = 18) +
    te(year, month, k = c(20, 11), bs = c("cr", "cc")) +
    s(capacity, k = 13) +
    s(mld_anom, k = 13) +
    yft_quota * country +
    quarter +
    s(vessel, bs = "re") +
    offset(log(num_pos_sets)),
  family = "gaussian", data = filter(D, yft_adult_present))
```

Note that vessel was modeled as a random effect (as indicated by `bs="re"`) and the number of sets that correspond to a given logbook line was included as an offset (log transformed as this is a log-normal model). Other effects were as in component 1.

2.4 Prediction grid

When predicting the final standardized CPUE abundance indices, a prediction grid was developed crossing the following different predictor variable levels:

- The years 1991 to 2022
- The months 1 to 12
- The minimum set of $1° \times 1°$ spatial grid cells present in the core fishing area for each quarter of the year for all 8 year time periods of the data set. Specifically, the data set was divided into 8 year time periods, and for each quarter of the year the minimum set of spatial grid cells representing at least 99% of the total search time (calculated based on the standard definition of daylight hours minus total set duration) was identified, producing four such sets for each time period. The length of the time periods was chosen so as to be long enough to have an approximately stable distribution of effort, but less than a decade and also evenly dividing the 32 year extent of the data set. For each quarter, the intersection of the sets for the time periods was calculated as the final prediction area so that only grid cells extensively used throughout the study time period were included in predictions. The outline of the areas used for each quarter, each representing between 82.0% and 88.5% of the total search time in a quarter, are shown in Figure 1. When making predictions the center of each grid cell was used for the longitude and latitude, and predictions were weighted by the inverse of the total number of spatial grid cells used in a quarter so that quarterly weights sum to one.
- Countries France and Spain, weighting predictions for each by the inverse of the fraction of total search time (calculated based on the standard definition of daylight hours minus total set duration) carried out by each country over the entire time series.
- The remaining predictive variables were fixed to single values as follows: (i) capacity was set to the weighted mean capacity over the entire dataset, weighting by search time (=1,759 m^3); (ii) predictor MLD anomaly (i.e., `mld_anom`) was set to zero; (iii) the offsets, "FSC" search time (`search_time_fsc`) and the number of positive FSC sets (`nom_pos_sets`), were each set to one.
- The random effect of the unique vessel identifier was zeroed out when making predictions, essentially predicting a mean over all vessels and ignoring the variability among vessels when estimating uncertainty.

2.5 Standardized CPUE indices

Standardized CPUE indices were developed by estimating mean predictions and 95% prediction intervals for each element of the prediction grid described in the previous section, multiplying together the three components of the model as appropriate, and then combining predictions as needed on a quarterly or annual basis using the weighting described in the previous section to develop quarterly and annual standardized CPUEs. The final results include the primary abundance index that represents the average amount of adult YFT biomass encountered in FSC per hour of "FSC" search time, as well as indices for adult YFT catch per set, adult YFT catch per set for which adult YFT was present, the encounter rate of FSC schools per unit search time and the probability of adult YFT being present in a

set. Prediction intervals are like confidence intervals, but whereas confidence intervals express certainty in the mean, prediction intervals express certainty in new observations. The methodology for estimating GAM prediction intervals is described by Andersen (2022) and Dumont et al. (2024). In practice, using prediction intervals produces larger uncertainties for the standardized CPUE indices, giving a better idea of the uncertainty in results that would occur if fishing had actually followed a standardized protocol with constant effort over space and time.

As the prediction areas differ for each quarter of the year, abundance indices varied quite extensively between quarters. As such, we also calculated "annual" quarterly indices as the running average of four consecutive quarterly values, with the date associated with each value being that of the first of the four values (e.g., January 1, 1991 corresponds to the average of the four quarterly values for the year 1991). For quarterly stock assessment models, these are the preferred indices.

3. Results

3.1 Basic logbook statistics

Basic graphs of the logbook data showing, per year, the number of vessel, the number of vessel-days, the number of sets, average FSC set duration, total search time and fraction of FSC sets per fishing trip are presented in Figure 2. Notable trends are the long term decrease in the number of active vessels with a particularly large temporary decrease 2009-2013 associated with the piracy period (Figure 2 a), and the long term decrease in the fraction of FSC sets for both the French and Spanish fleets, particularly since the onset of the YFT quota in 2017 (Figure 2 f).

3.2 Cross-correlations among predictors

Cross-correlation analysis among potential predictors for component 1 (Figure 3) and components 2 and 3 (Figure 4) revealed strong relationships between the year, the year a vessel entered into service, the YFT quota boolean variable and the vessel capacity, as well as between capacity and country and between the two definitions of search time (i.e., search_time and search_time_fsc). More moderate, but still important, relationships were identified between latitude and month, between FSC search time and the fraction of sets per trip on FSC, between longitude and MLD climatology, between longitude and latitude, and between month and year with the fraction of sets during a fishing trip that are on FSC (i.e., frac_fsc). Based on these results, we decided to remove the year a vessel entered into service (year_service), MLD climatology (mld_clim) and the fraction of FSC sets during a trip (frac_fsc) from models, the last of these partly being removed because interpretation of a decrease in this variable as indicative of a reduction in targeting of FSC schools, as opposed to a reduction in the presence of FSC schools, is unclear. We kept country and capacity as these variables were considered to be key vessel effects over which to standardize, and we kept longitude, latitude, year and month as these variables are essential to model structure and are modeled conjointly in pairs in GAMMs. However, we did not include both a year-month interaction and a month-YFT quota interaction in models given the obvious relationship between year and YFT quota.

3.3 Area of prediction

As shown in Figure 1, our scheme for developing quarterly prediction areas produced four generally contiguous prediction zones predominantly in the western Indian Ocean. The four zones each consist of between 220 and 336 $1° \times 1°$ grid cells representing between 82.0% and 88.5% of the total search time in a quarter.

3.4 Nominal CPUEs

Focusing on FSC sets (the right-hand panels of Figure 5), the nominal number of FSC sets per unit of (standard, i.e., not "FSC") search time has varied extensively over the data set time series, showing a marked increase during the "golden years" 2003-2005 and then a long term decline after that period until 2018, after which time values begin to recover somewhat (top-right panel of Figure 5). Total catch per FSC set also shows a major peak during the golden

years, largely driven by adult YFT catch, but catch per set has an increasing pattern from 2007-2022 (middle-right panel of Figure 5). However, there is a strong decline in nominal catch of adult YFT per set after the imposition of the YFT quota in 2017 (middle-right and bottom-right panels of Figure 5).

is approximately constant over the data set time period, with slight decrease 2020-2022, commensurate with the COVID pandemic, though this analysis does not permit identifying a causal link between COVID and this decrease. Total catch per FSC set has a peak in 2006, with lower, but similar and more or less constant, levels before and after this year (middle-right panel of Figure 5). 2006 is characterized by a large drop in the number of active vessels in our dataset (Figure 2 a), perhaps driven by the movement of vessels to the Indian Ocean in response to the Indian Ocean golden years (~2003-2005), immediately followed by the onset of Somali Piracy, which drove a return of vessels to the indian Ocean. The fraction of FSC catch that is adult YFT has a gradual increasing trend from the 1990's until ~2008, but is approximately stable after this time period at ~70%-80% (bottom-right of Figure 5). This drop is primarily due to a decrease in the nominal probability of presence of adult YFT in FSC sets (Figure 6 a).

3.5 GAMM results and diagnostics

3.5.1 Component 1

Summary output from the component 1 GAMM are:

```
Family: Tweedie(p=1.01)
Link function: log

Formula:
num_sets_fsc ~ te(lon, lat, by = quarter, k = k.sp) + te(year,
    month, k = c(k.y, k.m), bs = c("cr", "cc")) + s(capacity,
    k = k.ot) + s(mld_anom, k = k.ot) + yft_quota * country +
    quarter + s(vessel, bs = "re") + offset(log(search_time_fsc))

Parametric coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              -3.75542    0.09306 -40.355  < 2e-16 ***
yft_quotaTRUE            0.40662    0.08988   4.524 6.08e-06 ***
countrySpain            -0.05842    0.09388  -0.622  0.53374
quarter2                -0.54636    0.13515  -4.043 5.29e-05 ***
quarter3                -0.99086    0.16488  -6.010 1.86e-09 ***
quarter4                -1.26995    0.47912  -2.651  0.00804 **
yft_quotaTRUE:countrySpain 0.04386  0.02923   1.500  0.13350
---
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 178.371 198.884 17.590 < 2e-16 ***
te(lon,lat):quarter2 184.962 207.878 25.442 < 2e-16 ***
te(lon,lat):quarter3 143.651 162.675 29.554 < 2e-16 ***
te(lon,lat):quarter4 166.648 187.807 28.526 < 2e-16 ***
te(year,month)       189.706 196.922 65.241 < 2e-16 ***
s(capacity)            1.016   1.016  8.069 0.00424 **
s(mld_anom)           10.119  10.956 88.942 < 2e-16 ***
s(vessel)             63.162  64.000 48.892 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) =  0.206    Deviance explained = 25.3%
-REML =  99064   Scale est. = 1.0162     n = 240653
```

All explanatory variables and smooths included in component 1 had significant relationships to the FSC encounter rate except the parametric country and country-YFT quota interaction terms (Table 3), and overall the model explained 25.3% of the deviance. The QQ-plot indicated a reasonable fit of the model to the data, though there is some remaining discrepancy at higher theoretical quantiles (though much less so than when `search_time` instead of `search_time_fsc` was used as an offset; top-left of Figure 7). Residuals plots as a function of year (Figure 8) or month before and after the YFT quota (Figure 9) did not show any strong temporal trend, indicating that temporal effects, including any by-month effects of the YFT quota, are captured by model parametric and smooth effects.

Checking to see if the basis dimensions chosen for the smooth effects (i.e., the `k` parameters) are sufficient for component 1 with the `gam.check` function of the `mgcv` package indicated that this is not the case.

Marginal effects plots for the GAMM of FSC schools encountered per unit search time indicate a slightly higher encounter rate in the first half of the year than in the second half of the year (Figure 10). The year-month pattern shows increases in FSC encounter rate during the golden years (~2003-2005) and in 2022 (Figure 11). Any depression in the FSC encounter rate during the last quarter of the year after the imposition of the YFT quota (2017) appears to be at best a weak effect (top-right of Figure 11). Vessel capacity increases the FSC encounter rate (Figure 12 a), and typical MLD anomaly values have a decreasing effect on FSC encounter rate (part of Figure 12 b immediately above red horizontal bar). The vessel random effect is largely consistent with a normally-distributed random effect (Figure 12 c).

3.5.2 Component 2

Summary output from the component 2 GAMM are:

```
Family: binomial
Link function: logit

Formula:
yft_adult_present ~ te(lon, lat, by = quarter, k = k.sp) + te(year,
    month, k = c(k.y, k.m), bs = c("cr", "cc")) + s(capacity,
    k = k.ot) + s(mld_anom, k = k.ot) + yft_quota * country +
    quarter + s(vessel, bs = "re")

Parametric coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                1.72451    0.51580   3.343 0.000828 ***
yft_quotaTRUE              1.43312    0.45268   3.166 0.001546 **
countrySpain             -0.48286    0.10571  -4.568 4.93e-06 ***
quarter2                130.30266  707.34674   0.184 0.853846
quarter3                 -0.35019    0.92559  -0.378 0.705178
quarter4                749.49986 1347.45215   0.556 0.578050
yft_quotaTRUE:countrySpain -0.01164   0.11637  -0.100 0.920294
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                        edf  Ref.df  Chi.sq p-value
te(lon,lat):quarter1 140.459 156.772 1442.93  <2e-16 ***
te(lon,lat):quarter2 191.194 193.486 1457.28  <2e-16 ***
te(lon,lat):quarter3  59.288  73.073  600.95  <2e-16 ***
te(lon,lat):quarter4 178.247 181.317  747.95  <2e-16 ***
te(year,month)       191.979 197.357 2423.41  <2e-16 ***
```

```
s(capacity)              5.139   5.377    7.05   0.225
s(mld_anom)              3.284   4.259   38.97   <2e-16 ***
s(vessel)               48.735  64.000  216.93   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.419   Deviance explained = 37.9%
UBRE = -0.26802  Scale est. = 1         n = 43880
```

All explanatory variables and smooths included in component 2 had significant relationships to the probability of a positive FSC set including adult YFT except the smooth for capacity and the parametric effects for quarter and country-YFY quota interaction (Table 4). Overall the model explained 37.9% of the deviance. The model QQ-plot indicated a good fit of the model to the data (top-left of Figure 13). Residuals plots as a function of year (Figure 14) or month before and after the YFT quota (Figure 15) did not show any strong temporal trend, indicating that temporal effects, including any by-month effects of the YFT quota, are captured by model parametric and smooth effects. Nevertheless, there is notably larger variability in residuals for the years 2018-2020 and 2022, perhaps driven by within-country differences among vessels in targeting of SKJ- versus YFT-dominated FSC schools after the implementation of the YFT quota combined with the low levels of FSC sets for those years (Figure 2 f).

Checking to see if the basis dimensions chosen for the smooth effects (i.e., the k parameters) are sufficient for component 2 with the gam.check function of the mgcv package indicated that this is generally the case. There are indications that the dimensionality of the lon,lat and year,month tensor product interactions could be increased somewhat, but it was decided not to do so as the effective degrees of freedom are generally considerably below the maximum theoretical degrees of freedom and as doing so would increase model run times.

Marginal effects plots for the GAMM of adult YFT presence indicate relatively small seasonal variability among the spatial patterns of probability of adult YFT presence with instability of results in areas outside of core fishing zones where sampling effort is small (Figure 16). The year-month pattern that is characterized by a positive impact of June-July on probability of presence and a peak in probability of presence ~2011-2015, followed by a decrease in probability approximately after the imposition of the YFT quota in 2017 (Figure 17). This latter effect is perhaps not captured by the yft_quota parametric effect because it is changing over years. Vessel capacity has an unstable and small effect on probability of presence (Figure 18 a), whereas MLD anomaly has a negative impact on probability of presence for typical values of MLD anomaly (Figure 18 b). The vessel random effect is largely consistent with a normally-distributed random effect (Figure 18 c).

3.5.3 Component 3

Summary output from the component 3 GAMM are:

```
Family: gaussian
Link function: identity

Formula:
log(yft_adult) ~ te(lon, lat, by = quarter, k = k.sp) + te(year,
    month, k = c(k.y, k.m), bs = c("cr", "cc")) + s(capacity,
    k = k.ot) + s(mld_anom, k = k.ot) + yft_quota * country +
    quarter + s(vessel, bs = "re") + offset(log(num_pos_sets))

Parametric coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.13597    0.07479  41.932   <2e-16 ***
yft_quotaTRUE       0.11798    0.15673   0.753   0.4516
countrySpain       -0.07540    0.03342  -2.256   0.0241 *
quarter2           -0.10208    0.15978  -0.639   0.5229
quarter3           -0.58188    0.26014  -2.237   0.0253 *
```

```
quarter4                       -0.14845     0.14277  -1.040   0.2985
yft_quotaTRUE:countrySpain  0.02357     0.04736   0.498   0.6187
---
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 118.409 139.14   5.137  < 2e-16 ***
te(lon,lat):quarter2 109.247 128.12   6.349  < 2e-16 ***
te(lon,lat):quarter3  49.683  61.61   3.162  < 2e-16 ***
te(lon,lat):quarter4  64.721  82.28   4.827  < 2e-16 ***
te(year,month)       171.039 187.65   8.367  < 2e-16 ***
s(capacity)            1.000   1.00  15.908 6.69e-05 ***
s(mld_anom)            3.194   4.15   5.340 0.000234 ***
s(vessel)             48.409  64.00   2.839  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.166   Deviance explained = 17.4%
GCV = 1.0468  Scale est. = 1.029      n = 33599
```

All explanatory variables and smooths included in component 3 except YFT quota and the interaction of YFT quota with country had significant relationships to the biomass of adult YFT caught in positive FSC sets for which adult YFT was present (Table 5), and overall the model explained 17.4% of the deviance. The model QQ-plot indicated a reasonable fit of the model to the data (top-left of Figure 19).

Checking to see if the basis dimensions chosen for the smooth effects (i.e., the k parameters) are sufficient for component 3 with the gam.check function of the mgcv package indicated that this is generally the case. There is a marginally significant indication that the dimensionality of the lon,lat tensor product interaction could be increased somewhat, but it was decided not to do so as the effective degrees of freedom are generally considerably below the maximum theoretical degrees of freedom and as doing so would increase model run times.

Marginal effects plots for the GAMM of biomass of adult YFT per set for which adult YFT is present indicate seasonal patterns that are variable and difficult to interpret (Figure 22). The year-month pattern is characterized by higher biomass August-November and somewhat less biomass per set in recent years, potentially linked to the imposition of the YFT quota (Figure 23). Vessel capacity increases the adult YFT biomass (Figure 24 a), and typical MLD anomaly values have a small, marginally increasing impact on adult YFT biomass (Figure 24 b). The vessel random effect is largely consistent with a normally-distributed random effect (Figure 24 c).

3.6 Standardized CPUE indices

3.6.1 Annual indices

Standardized indices of each of the three components (Figure 25) generally reflected overall trends in the nominal CPUE indices, but there were some notable differences. The standardized FSC encounter rate (component 1 of Figure 25) has a somewhat smaller peak during the golden years (2003-2005) and remains somewhat below the nominal FSC encounter rate for most of the period 2005-2022. Standardized probability of adult YFT presence (component 2 of Figure 25) is somewhat below the nominal CPUE over most of the time series and shows increased uncertainty and a somewhat larger discrepancy since the start of the YFT quota in 2017. Catch of adult YFT per positive FSC set for which adult YFT were present (component 3 of Figure 25) follows quite closely the nominal index until ~2013, after which time and in particular after 2017 it is consistently below the nominal index and subject to increased uncertainty.

Final combined abundances indices of biomass of adult YFT encountered per unit search time (bottom panel of Figure 26) or caught per set (top panel of Figure 26) are again broadly consistent with nominal CPUEs, but the peak

during the golden years (2003-2005) is less pronounced in the standardized indices than in the nominal indices and standardized adult YFT catch per set (top panel of Figure 26) and, to a lesser degree, standardized adult YFT encountered per unit search time (bottom panel of Figure 26) are below nominal catch per set since the implementation of the YFT quota. Both indices, and in particular the biomass of adult YFT encountered per unit search time, show an increasing trend since ~2019, though neither has attained levels observed ~10 years ago.

3.6.2 Quarterly indices

We also present quarterly indices in Figure 27 and Figure 28 and running-average-smoothed quarterly indices in Figure 29 and Figure 30. The unsmoothed quarterly indices show considerable variability between quarters, though generally less so than the standardized indices.

3.6.3 Tabular data with indices

Annual, quarterly and running-average-smoothed quarterly indices for the Indian Ocean in tabular format (with semi-colon as the delimiter), along with a brief metadata file, can be downloaded at the following web address: https://drive.ird.fr/s/StDkw9KJNawBpT2.

4. Discussion

Overall, the modeling approach used provides a reasonable, comprehensive and flexible methodology to developing standardized abundance indices from PS FSC catch-effort data. Results indicate an increase in abundance since ~2019, commensurate with the imposition of the YFT quota in 2017 (bottom panel of Figure 26). This increase is primarily associated with an increase in the FSC encounter rate, though the other two components, probability of adult YFT presence in positive FSC sets and catch of adult YFT per positive set for which adult YFT are present, have also increased somewhat in recent years (Figure 25).

One important question is whether or not our standardization approach accurately corrects for the impacts of the YFT quota on targeting and catch of FSC schools. One would naively expect that the YFT quota would lead fishers to avoid setting on YFT-dominated FSC schools, either by fishing in areas or time periods not characterized by the such schools or by purposely not setting on such schools if they encounter them, particularly in the final months of the calendar year when vessels approach their YFT quota. Though we have included a binary YFT quota parametric effect in our standardization models, the somewhat negative trends in the year-month tensor product marginal effects in recent years (particularly, Figure 17 and Figure 23) are consistent with naive expectations regarding the impact of the quota and with the parametric `yft_quota` terms not fully accounting for the impacts of the YFT quota on catch patterns. Fully accounting for these effects in the standardization model is difficult due to the covariability between the quota and the year effects. Nevertheless, we can predict how fully accounting for these YFT quota effects would impact our standardized CPUE indices. If our interpretation that the observed negative trends in component indices in recent years are due to the quota and not real changes in abundance, then correcting for these effects would lead to a larger increase in estimated abundance in recent years (i.e., since ~2019) than what is actually observed in our standardized CPUE indices. In this sense, our indices are likely conservative regarding the positive impact of the YFT quota on YFT abundance, the true stock status being somewhat more positive than our estimates.

Overall, our indices are consistent with a somewhat decreasing trend in abundance 2012-2017 that led to the imposition of the YFT quota, followed by an initial recovery in population status in recent years. We hope to integrate 2023 catch-effort data into our models when they become available to help confirm the durability of these trends.

References

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

Bertrand A, Josse E, Bach P, Gros P, Dagorn L (2002) Hydrological and trophic characteristics of tuna habitat: Consequences on tuna distribution and longline catchability. *Canadian Journal of Fisheries and Aquatic Sciences* **59**:1002–1013. doi:10.1139/f02-073

Cayré P, Marsac F (1993) Modelling the yellowfin tuna (Thunnus albacares) vertical distribution using sonic tagging results and local environmental parameters. *Aquatic Living Resources* **6**:1–14. doi:10.1051/alr:1993001

Correa GM, Kaplan DM, Grande M, Uranga J, Ramos Alonso ML, Alyón PP, Rojo V, Merino G, Santiago J (2024) Standardized catch per unit effort of yellowfin tuna in the Indian Ocean for the European purse seine fleet operating on floating objects. IOTC 26th Working Party on Tropcial Tunas, Data Preparatory Meeting

Dumont A, Duparc A, Sabarros PS, Kaplan DM (2024) Modeling bycatch abundance in tropical tuna purse seine fisheries on floating objects using the Δ method. *ICES Journal of Marine Science*:fsae043. doi:10.1093/icesjms/fsae043

Green RE (1967) Relationship of the Thermocline to Success of Purse Seining for Tuna. *Transactions of the American Fisheries Society* **96**:126–130. doi:10.1577/1548-8659(1967)96[126:ROTTTS]2.0.CO;2

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)

IOTC Secretariat (2016) On an interim plan for rebuilding the Indian Ocean yellowfin tuna stock in the IOTC area of competence. Resolution 16/01. Indian Ocean Tuna Commission

Kaplan DM, Tolotti MT (2023) Silky shark abundance index based on CPUE standardization of French Indian Ocean tropical tuna purse seine observer bycatch data. IOTC-2023-WPEB19-34_Rev1. IOTC 19th Working Party on Ecosystems and Bycatch, Réunion, Fance

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

Pallarés P, Petit C (1998) Tropical tunas: New sampling and data processing strategy for estimating the composition of catches by species and sizes. *ICCAT Collective Volume of Scientific Papers* **48**:230–246

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

Tables

Table 1: Dependent variables for the three CPUE standardization component models. Note that when running component 3, `adult_yft` is restricted to just the sets for which `adult_yft_present` is true.

Component	Variable	Variable type	Description
Comp. 1	num_sets_fsc	Count	Number of (positive or null) FSC sets for a given vessel-day
Comp. 2	adult_yft_present	Boolean	Presence-absence of adult YFT in each positive (>=1 tonne) FSC set
Comp. 3	adult_yft	Continuous, >=0	Tonnes of adult YFT caught per positive (>=1 tonne) FSC set

Table 2: Candidate explanatory variables for CPUE standardization models.

Variable	Variable type	Description
country	Categorical	France or Spain
vessel	Categorical	Unique vessel identifier
year_service	Continuous	The year that the vessel was put into service
capacity	Continuous, >0	Vessel storage capacity in cubic meters
num_pos_sets	Count	Number of non-null sets for a given logbook entry
year	Categorical or continuous	Year of the fishing set
quarter	Categorical	Quarter of the fishing set
month	Circular	Month of the fishing set
yft_quota	Logical	True if date >= 2017-01-01; False otherwise
lon,lat	Continuous	Geographic coordinates of fishing set or centroid of activities
mld_clim	Continuous	30-year climatology of mixed layer depth in meters
mld_anom	Continuous	Mixed layer depth anomaly relative to climatology
search_time	Continuous, >0	Number of daylight hours minus the total duration of all fishing sets for a given vessel-day
search_time_fsc	Continuous, >0	Number of daylight hours minus the total duration of all FOB fishing sets for a given vessel-day
fsc_frac	Proportion	Proportion of sets that were FSC during entire fishing trip

(a) Parametric terms

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.755	0.093	-40.355	0.000	***
yft_quotaTRUE	0.407	0.090	4.524	0.000	***
countrySpain	-0.058	0.094	-0.622	0.534	
quarter2	-0.546	0.135	-4.043	0.000	***
quarter3	-0.991	0.165	-6.010	0.000	***
quarter4	-1.270	0.479	-2.651	0.008	**
yft_quotaTRUE:countrySpain	0.044	0.029	1.500	0.134	

(b) Smoothed terms

	edf	Ref.df	F	p-value	
te(lon,lat):quarter1	178.371	198.884	17.590	0.000	***
te(lon,lat):quarter2	184.962	207.878	25.442	0.000	***
te(lon,lat):quarter3	143.651	162.675	29.554	0.000	***
te(lon,lat):quarter4	166.648	187.807	28.526	0.000	***
te(year,month)	189.706	196.922	65.241	0.000	***
s(capacity)	1.016	1.016	8.069	0.004	**
s(mld_anom)	10.119	10.956	88.942	0.000	***
s(vessel)	63.162	64.000	48.892	0.000	***

Table 3: Summary statistics and p-values for fixed and smooth terms included in the component 1 GAMM.

(a) Parametric terms

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.725	0.516	3.343	0.001	***
yft_quotaTRUE	1.433	0.453	3.166	0.002	**
countrySpain	-0.483	0.106	-4.568	0.000	***
quarter2	130.303	707.347	0.184	0.854	
quarter3	-0.350	0.926	-0.378	0.705	
quarter4	749.500	1347.452	0.556	0.578	
yft_quotaTRUE:countrySpain	-0.012	0.116	-0.100	0.920	

(b) Smoothed terms

	edf	Ref.df	Chi.sq	p-value	
te(lon,lat):quarter1	140.459	156.772	1442.927	0.000	***
te(lon,lat):quarter2	191.194	193.486	1457.283	0.000	***
te(lon,lat):quarter3	59.288	73.073	600.948	0.000	***
te(lon,lat):quarter4	178.247	181.317	747.949	0.000	***
te(year,month)	191.979	197.357	2423.414	0.000	***
s(capacity)	5.139	5.377	7.050	0.225	
s(mld_anom)	3.284	4.259	38.973	0.000	***
s(vessel)	48.735	64.000	216.933	0.000	***

Table 4: Summary statistics and p-values for fixed and smooth terms included in the component 2 GAMM.

(a) Parametric terms

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.136	0.075	41.932	0.000	***
yft_quotaTRUE	0.118	0.157	0.753	0.452	
countrySpain	-0.075	0.033	-2.256	0.024	*
quarter2	-0.102	0.160	-0.639	0.523	
quarter3	-0.582	0.260	-2.237	0.025	*
quarter4	-0.148	0.143	-1.040	0.298	
yft_quotaTRUE:countrySpain	0.024	0.047	0.498	0.619	

(b) Smoothed terms

	edf	Ref.df	F	p-value	
te(lon,lat):quarter1	118.409	139.137	5.137	0	***
te(lon,lat):quarter2	109.247	128.125	6.349	0	***
te(lon,lat):quarter3	49.683	61.614	3.162	0	***
te(lon,lat):quarter4	64.721	82.276	4.827	0	***
te(year,month)	171.039	187.653	8.367	0	***
s(capacity)	1.000	1.000	15.908	0	***
s(mld_anom)	3.194	4.150	5.340	0	***
s(vessel)	48.409	64.000	2.839	0	***

Table 5: Summary statistics and p-values for fixed and smooth terms included in the component 3 GAMM.

Figures
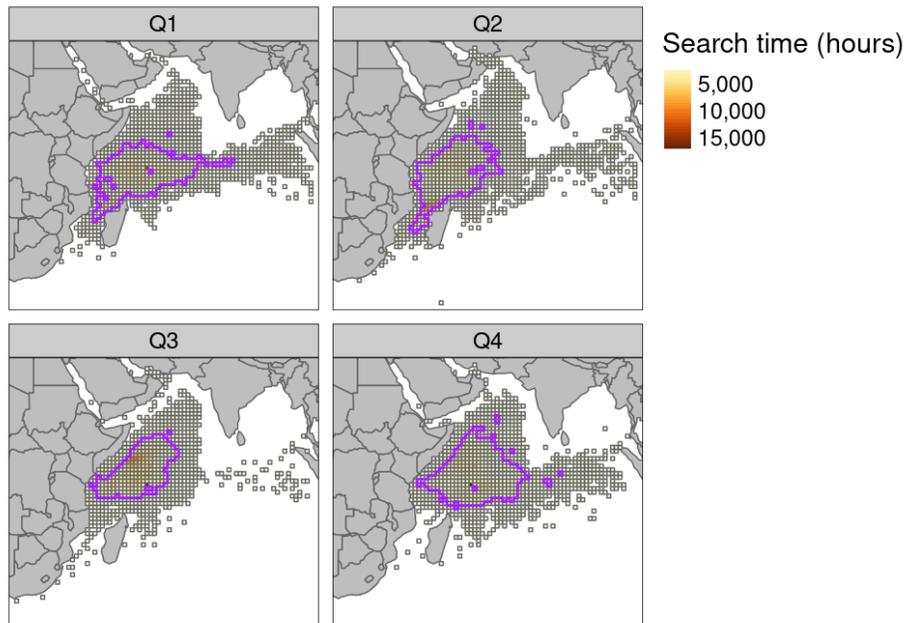
Figure 1: Total search time in the dataset for each 1° × 1° grid cell explored by the fishery for each quarter of the year. The purple curve indicates the border of the area used for model prediction determined as the intersection of the smallest set of grid cells for each 8 year period of the dataset representing at least 99% of the search effort for that 8 year period.
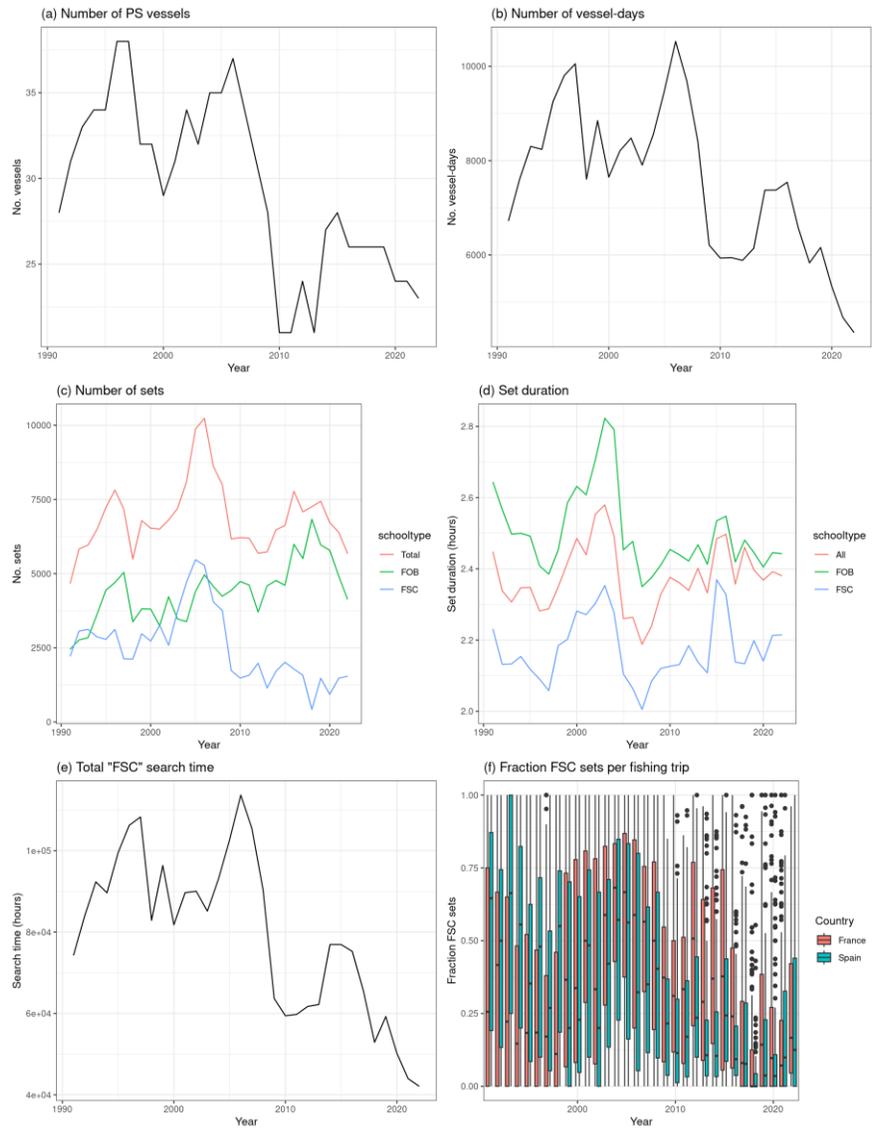
Figure 2: The number of vessels (a), the number of vessel-days (b), the number of sets (c), average FSC set duration (d), total "FSC" search time (e), and fraction of FSC sets per fishing trip (f) as a function of year for the final dataset used for standardizing CPUEs.
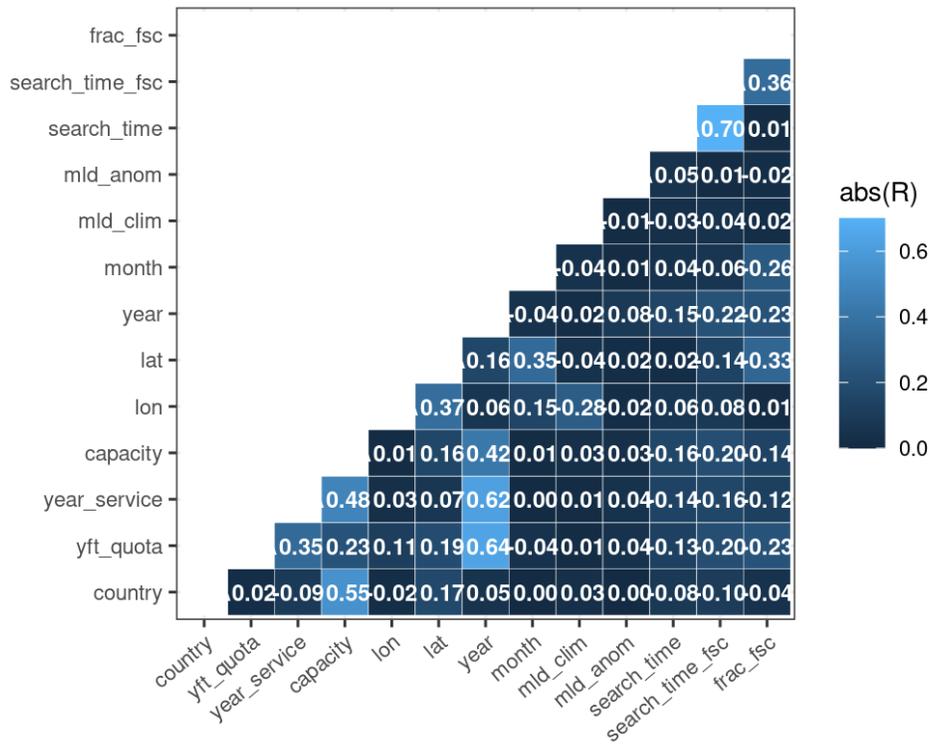
Figure 3: Cross-correlation matrix among potential predictors to be used in component 1 of the standardization models.
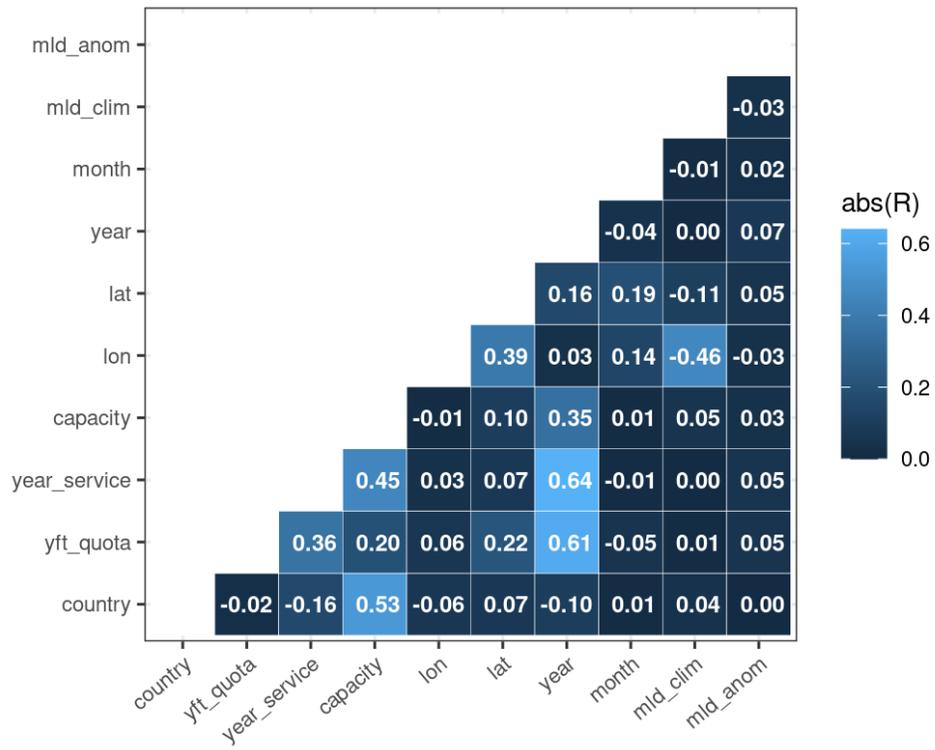
Figure 4: Cross-correlation matrix among potential predictors to be used in components 2 and 3 of the standardization models.
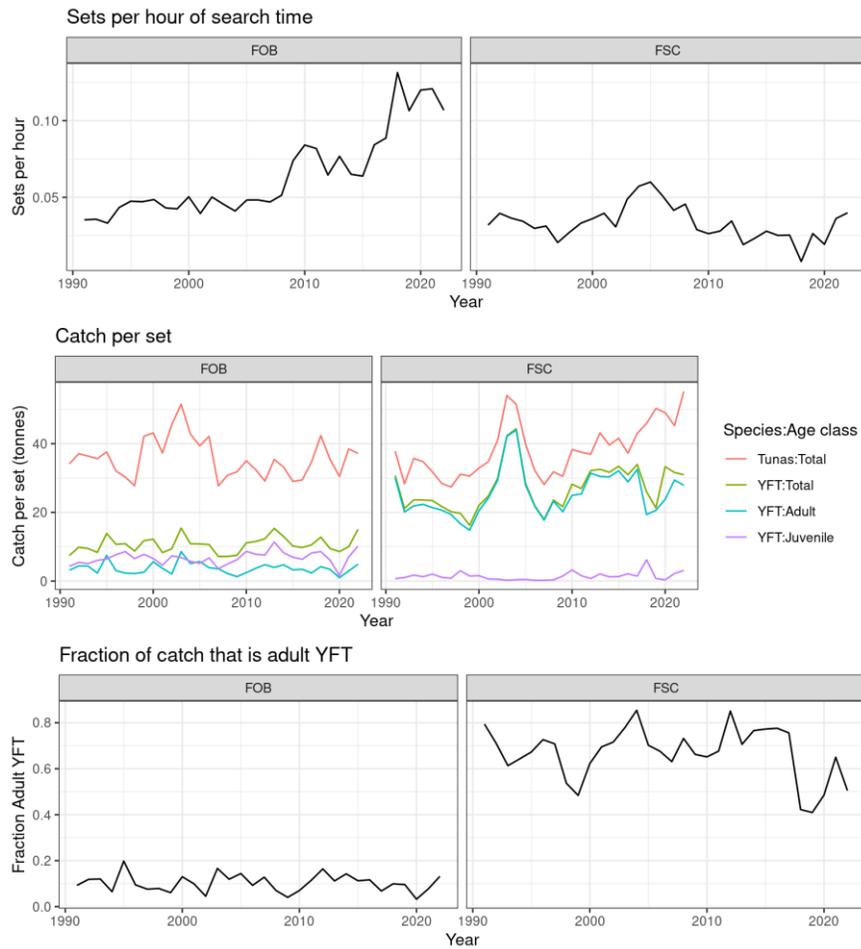
Figure 5: Nominal CPUE indices by school type as a function of year for various components of abundance (fish school encounter rate, size of fish schools and fraction of fish school biomass that is adult YFT). The top two panels indicate the average number of fishing sets carried out per unit of search time (*note:* this is standard search time, not "FSC" search time used in the GAMM for compnent 1), the middle two panels indicate the catch per set for different classes of fish (all tunas versus just YFT; all, adult and juvenile) and the bottom two panels indicate the fraction of catch biomass that is adult YFT.
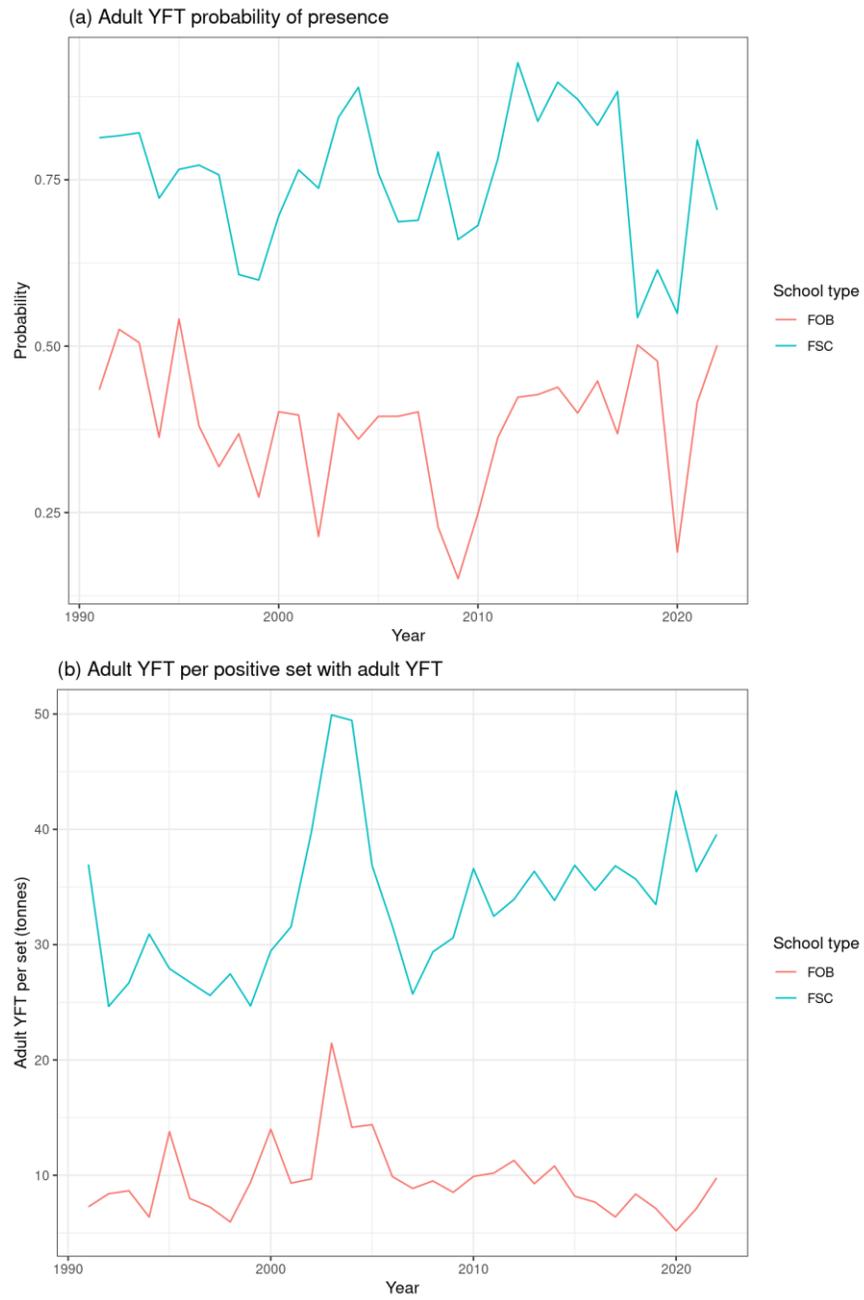
Figure 6: Nominal CPUE indices by school type as a function of year for (a) probability of presence of adult YFT in a positive set and (b) biomass of adult YFT in a positive set given that adult YFT was present in the set.
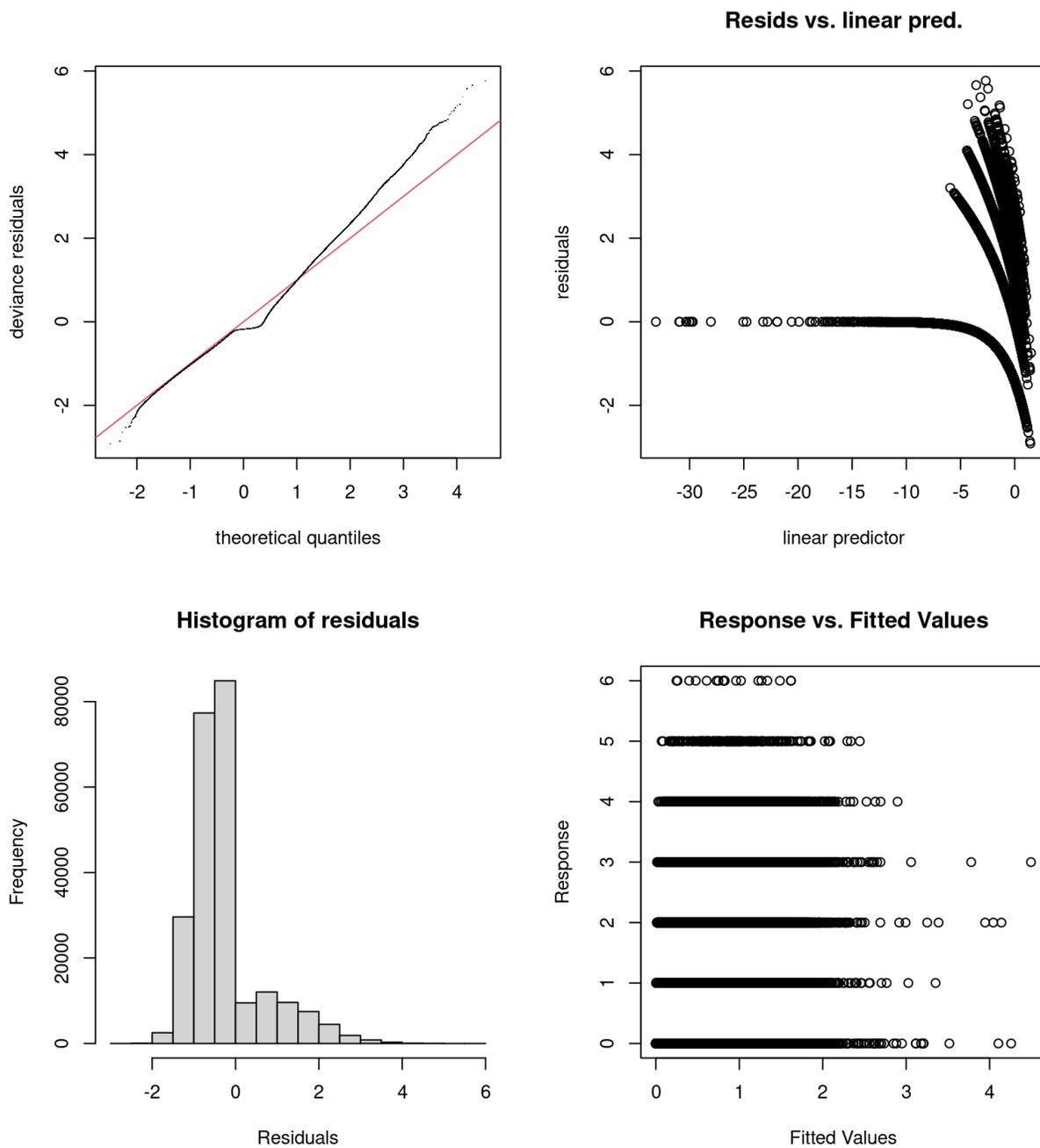
Figure 7: Standard diagnostic plots for the component 1 GAMM. *top-left:* QQ-plot; *top-right:* deviance residuals versus the linear predictor; *bottom-left:* histogram of deviance residuals; and *bottom-right:* Response variables as a function of fitted values.
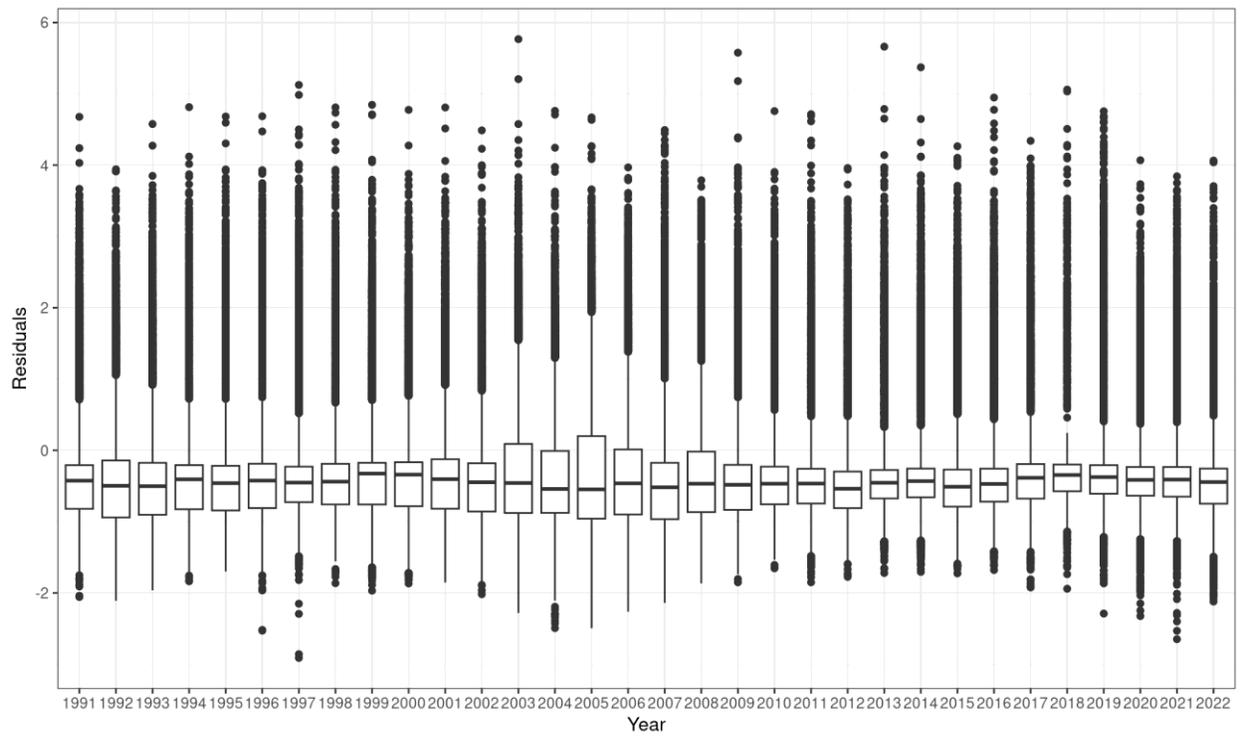
Figure 8: Boxplot of residuals of component 1 model as a function of year.
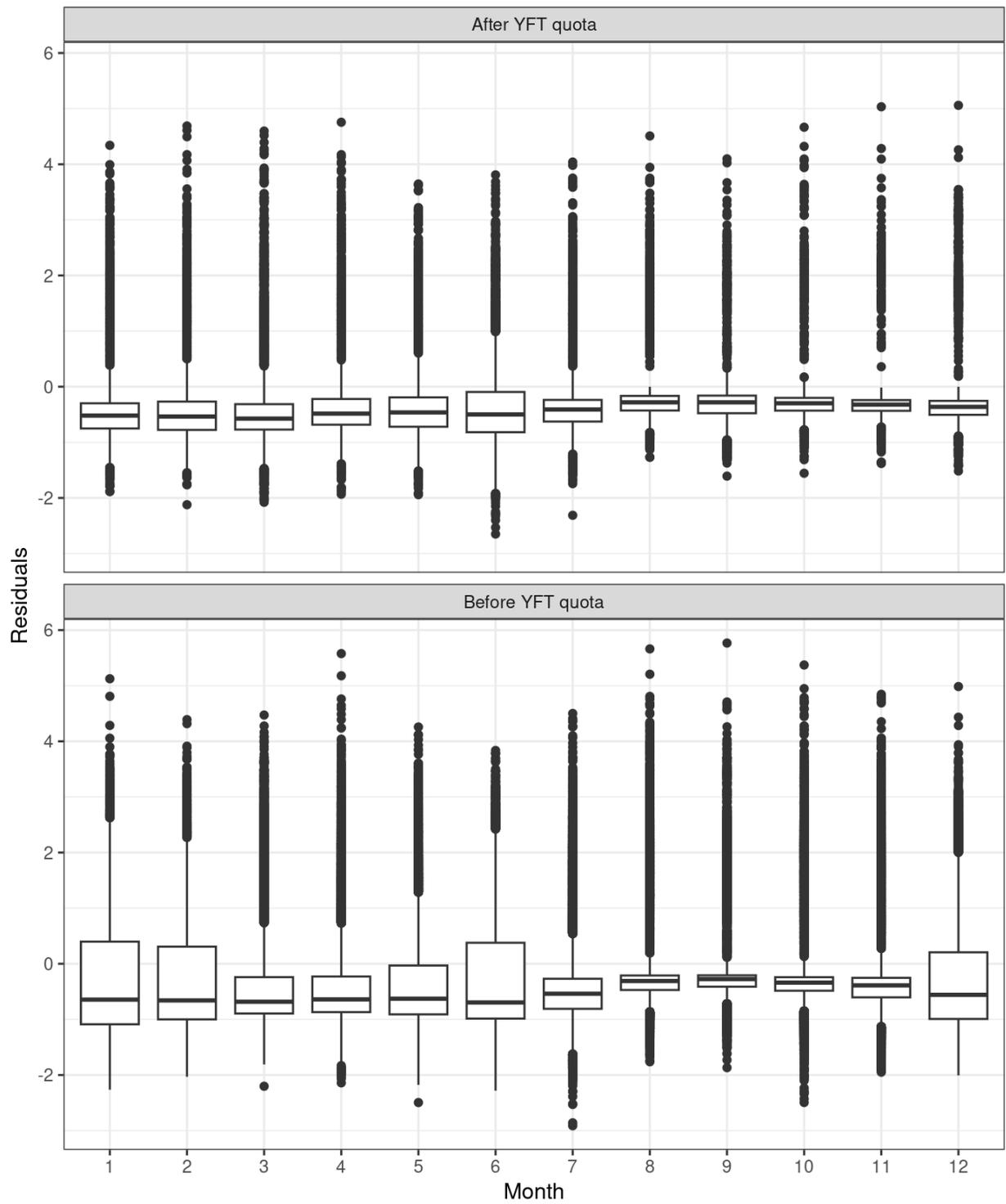
Figure 9: Boxplot of residuals of component 1 model as a function of month before (bottom) and after (top) the implementation of the YFT quota.
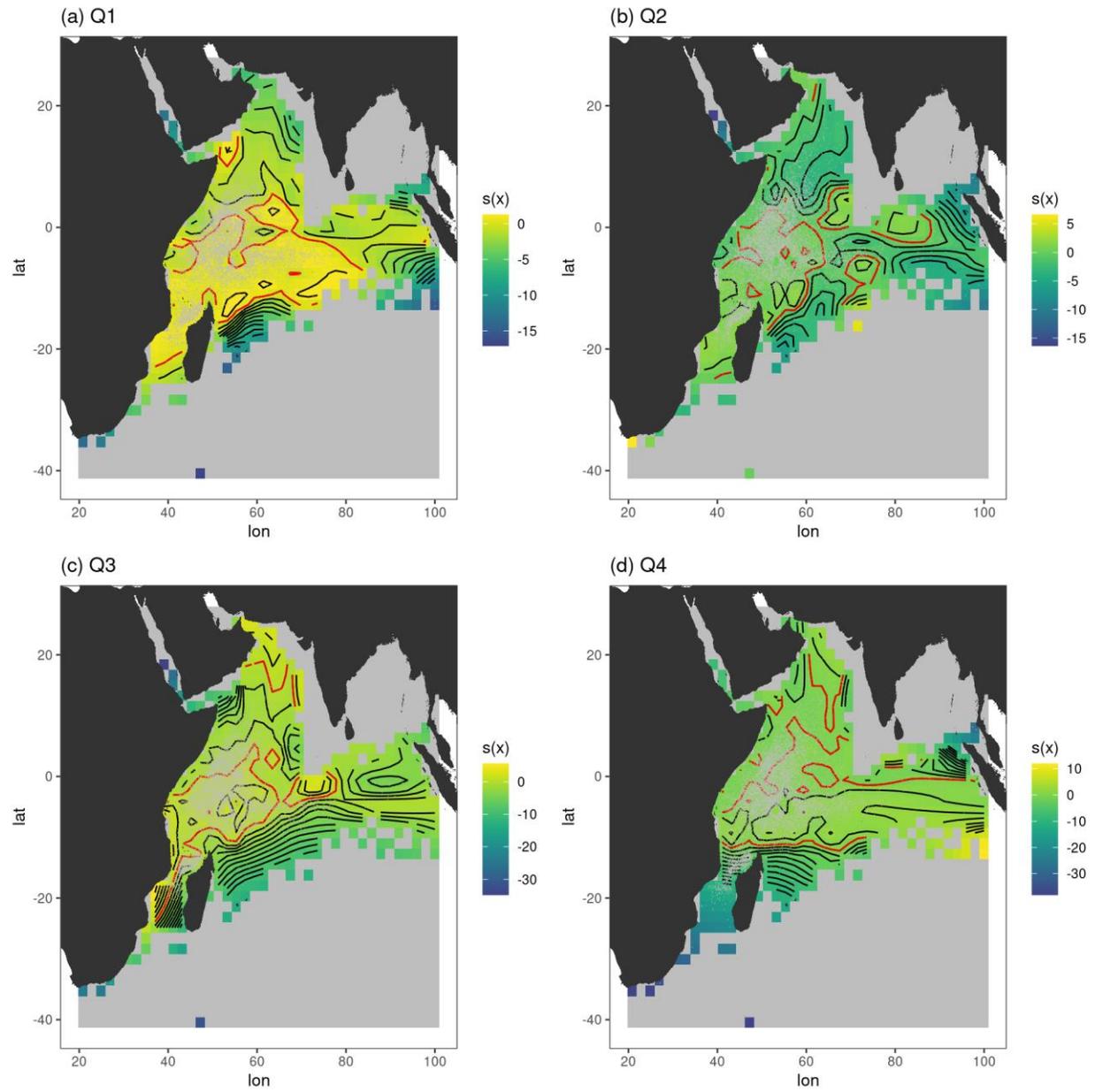
Figure 10: Marginal effect of lon,lat on FSC schools encountered per hour of search time (component 1) for each of the four quarters. Contour lines go from a smooth effect of -10 to an effect of 10 stepping by 1, with the zero contour line shown in red.
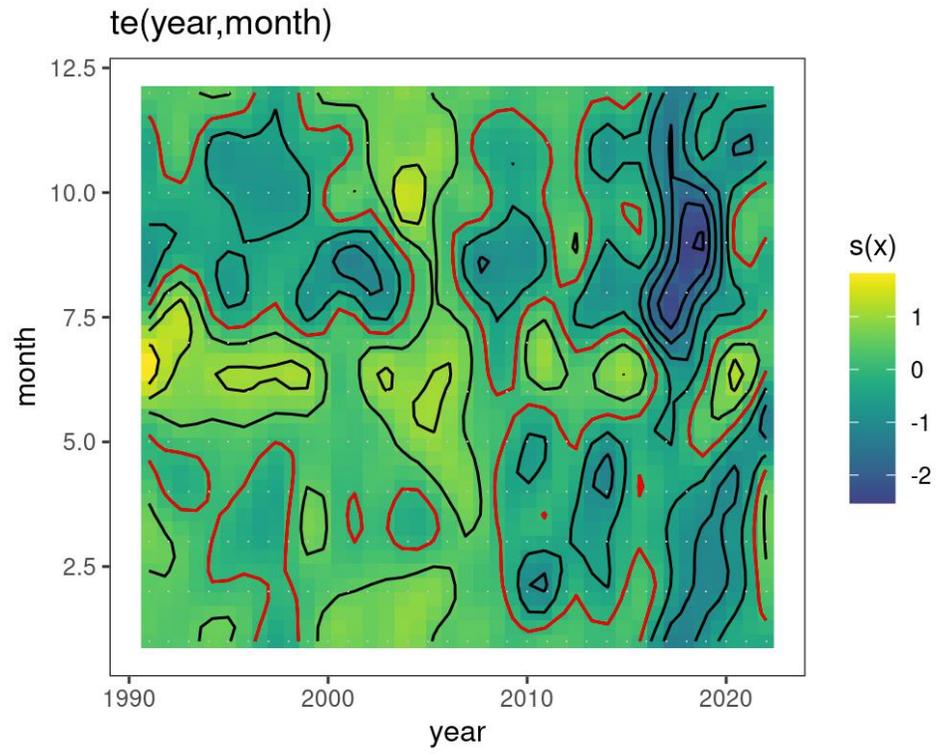
Figure 11: Marginal effect of year,month tensor product smooth on FSC schools encountered per hour of search time (component 1). The zero effect contour line is shown in red.
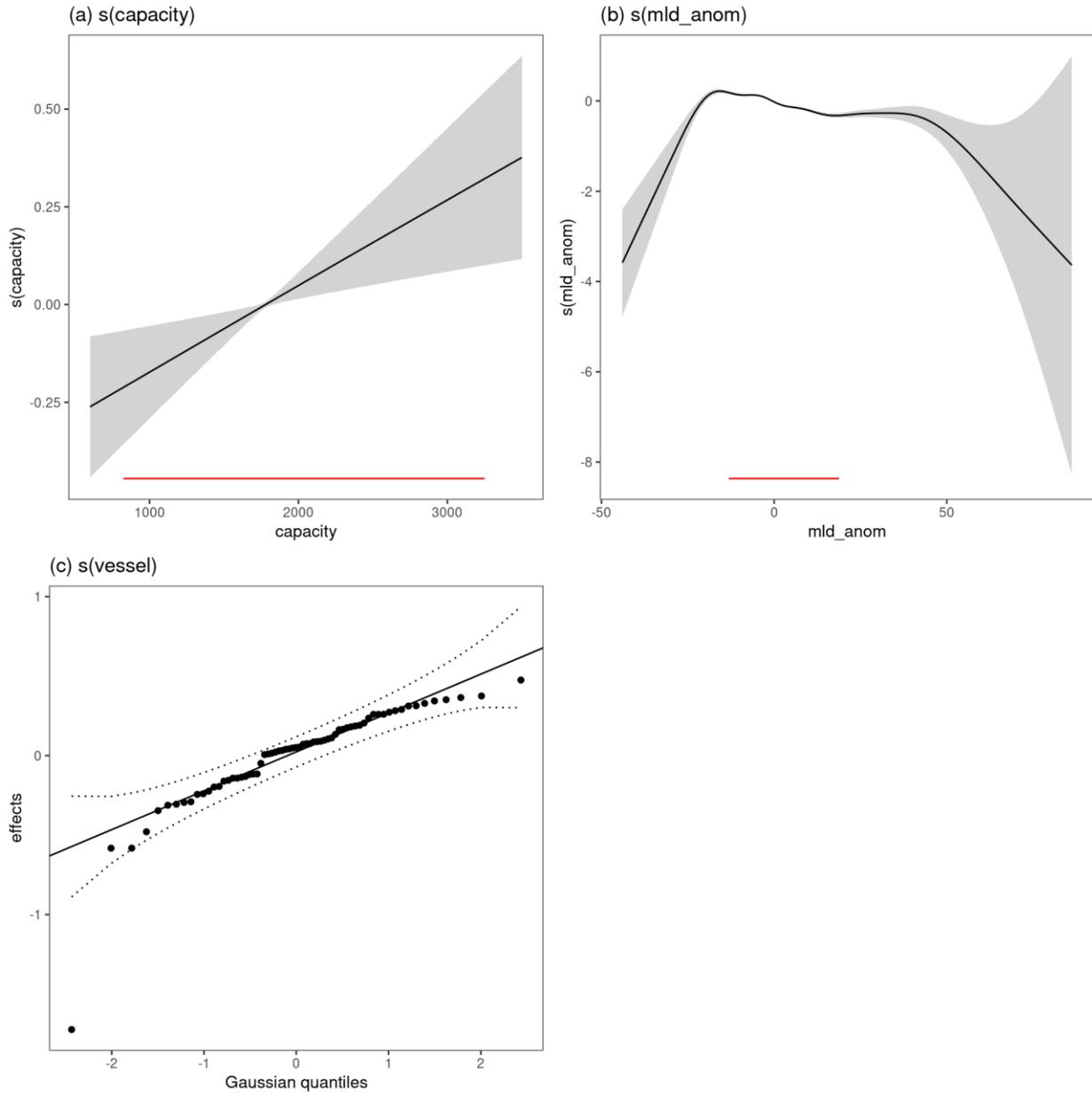
Figure 12: Marginal effects of year and fishing-efficiency-related individual smooths on FSC schools encountered per hour of search time (component 1). The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.
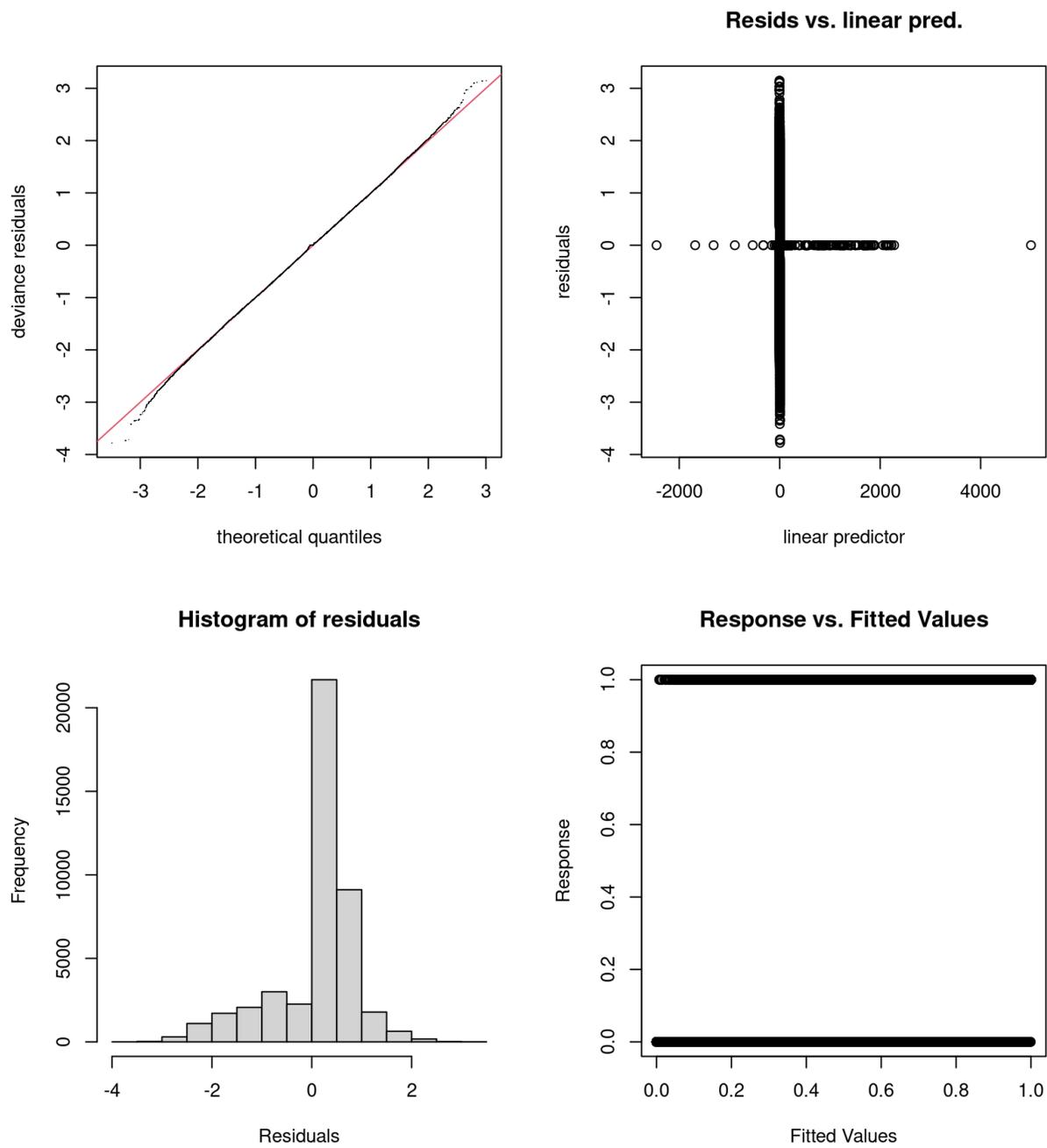
Figure 13: Standard diagnostic plots for the component 2 GAMM. *top-left:* QQ-plot; *top-right:* deviance residuals versus the linear predictor; *bottom-left:* histogram of deviance residuals; and *bottom-right:* Response variables as a function of fitted values.
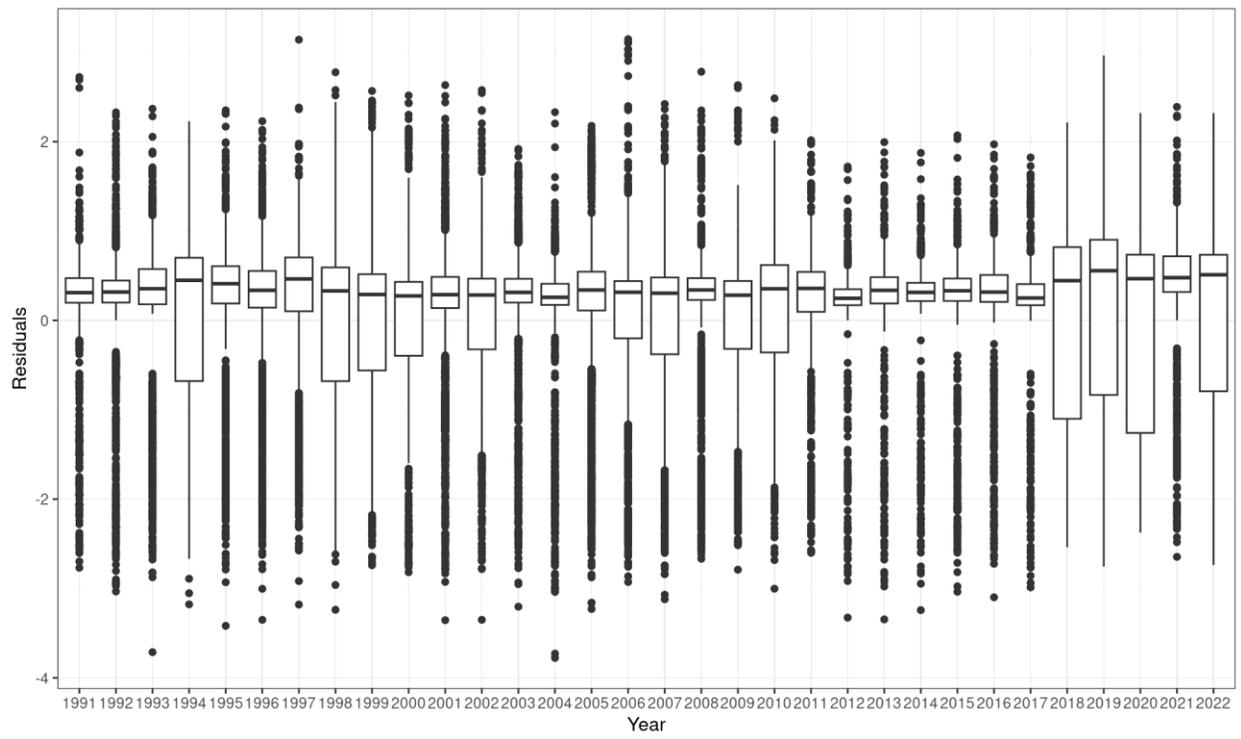
Figure 14: Boxplot of residuals of component 2 model as a function of year.
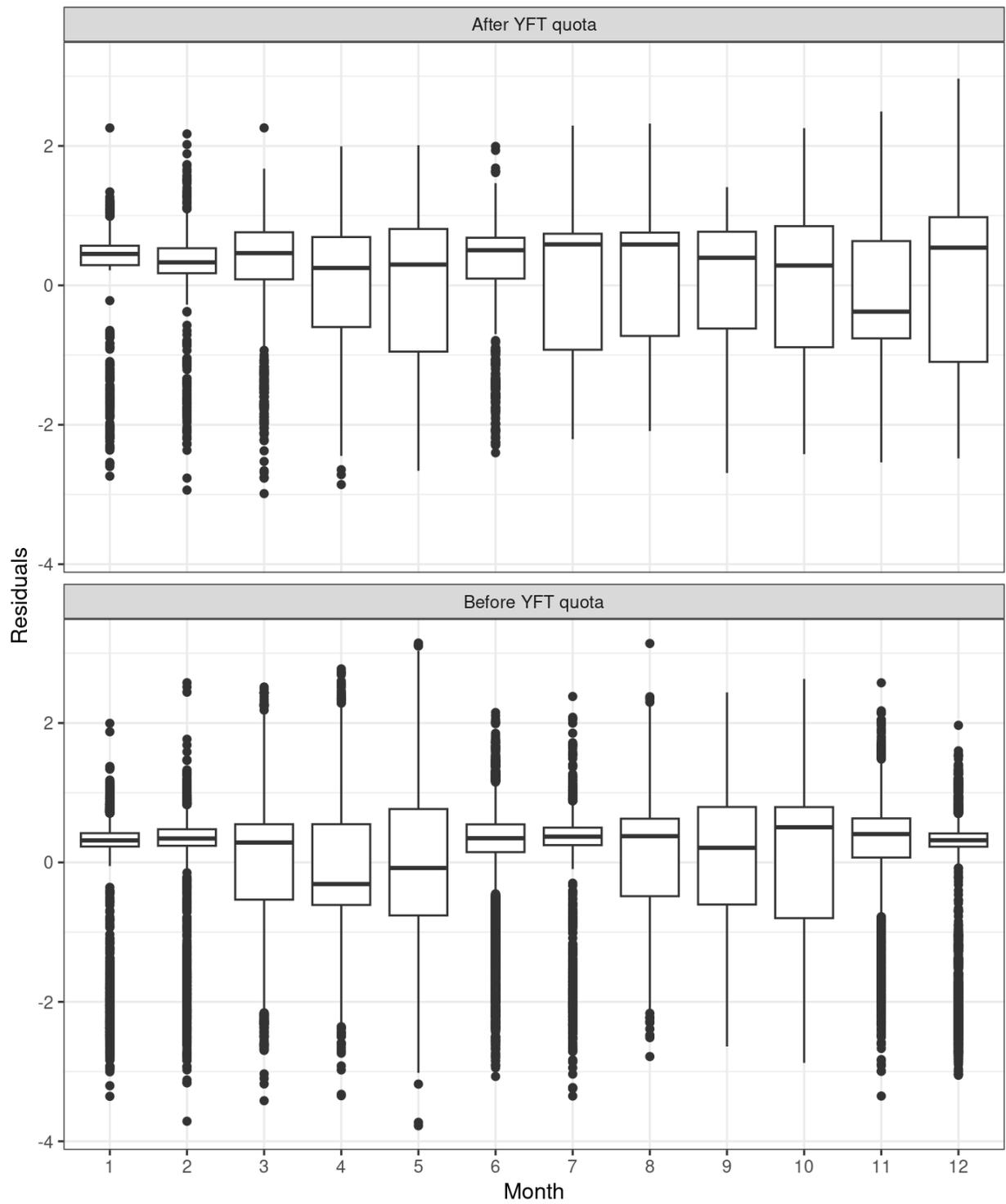
Figure 15: Boxplot of residuals of component 2 model as a function of month before (bottom) and after (top) the implementation of the YFT quota.
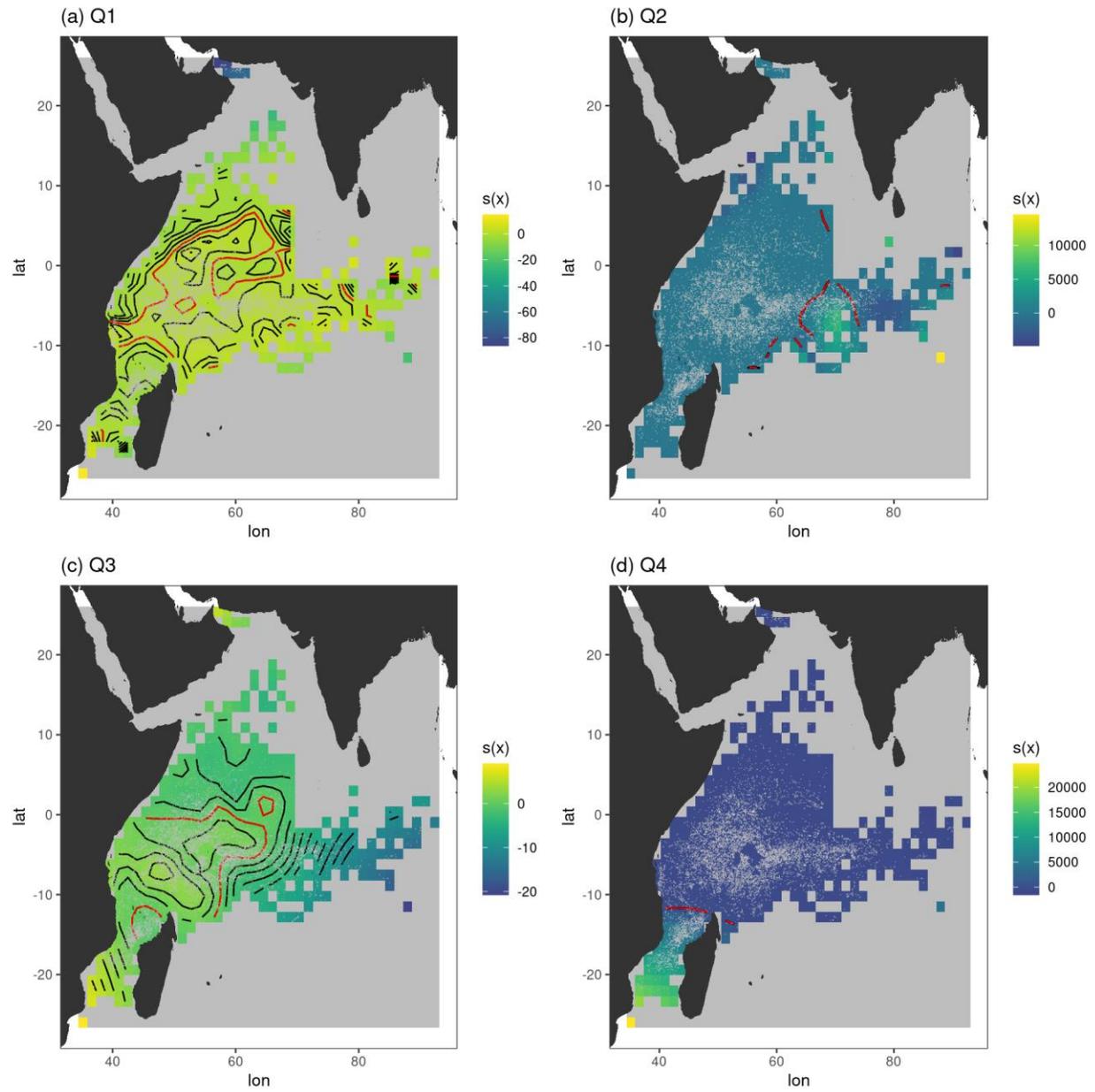
Figure 16: Marginal effect of lon,lat on probability of adult YFT presence in a positive FSC set (component 2) for each of the four quarters. Contour lines go from a smooth effect of -10 to an effect of 10 stepping by 1, with the zero contour line shown in red.
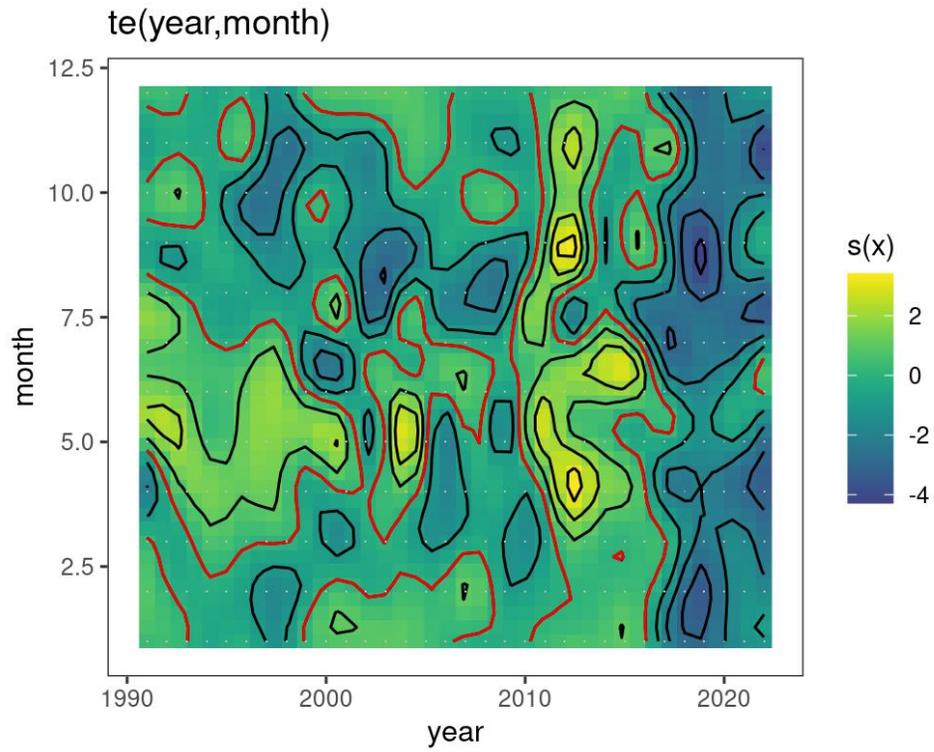
Figure 17: Marginal effect of year,month tensor product smooth on probability of adult YFT presence in a positive FSC set (component 2). The zero effect contour line is shown in red.
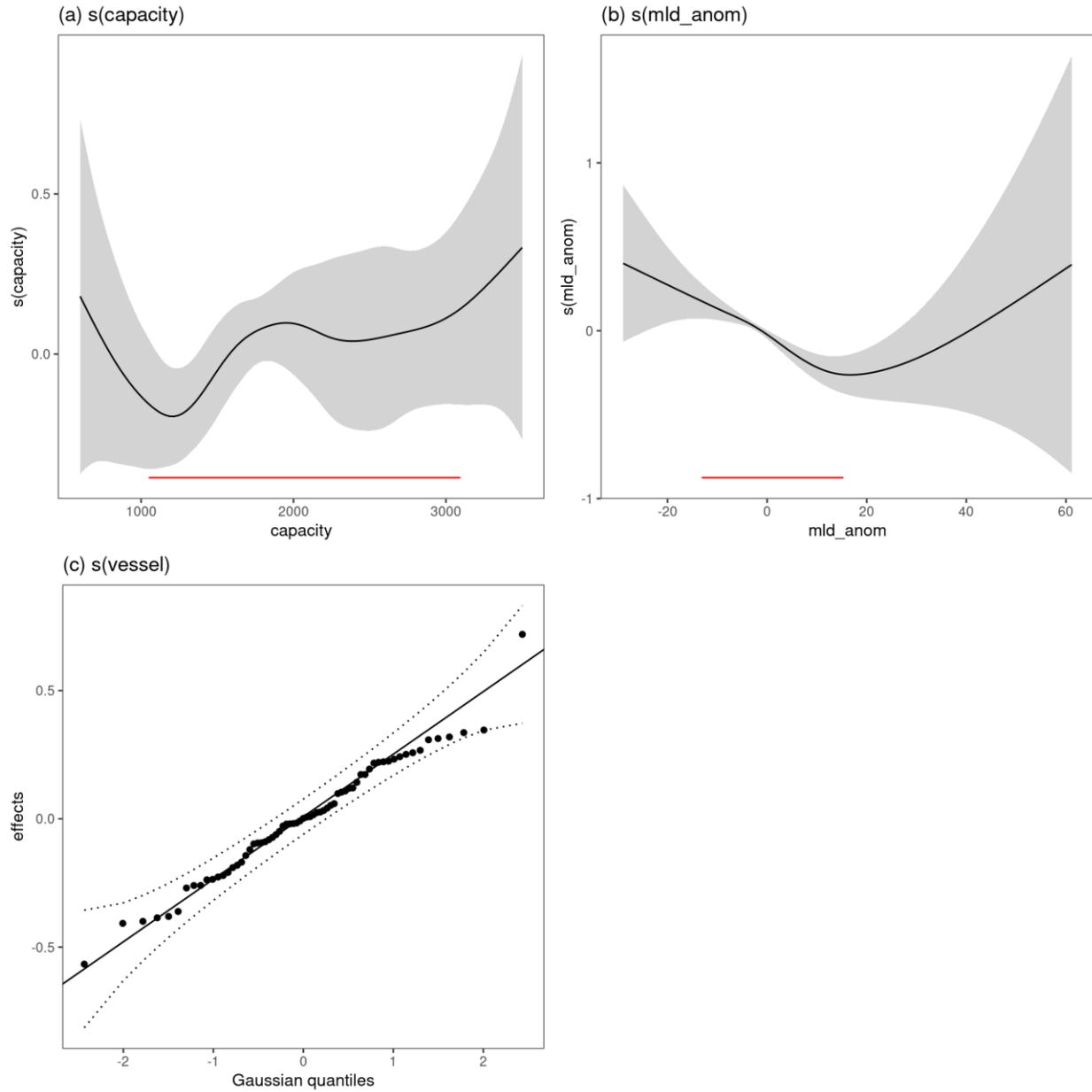
Figure 18: Marginal effects of year and fishing-efficiency-related individual smooths on probability of adult YFT presence in a positive FSC set (component 2). The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.
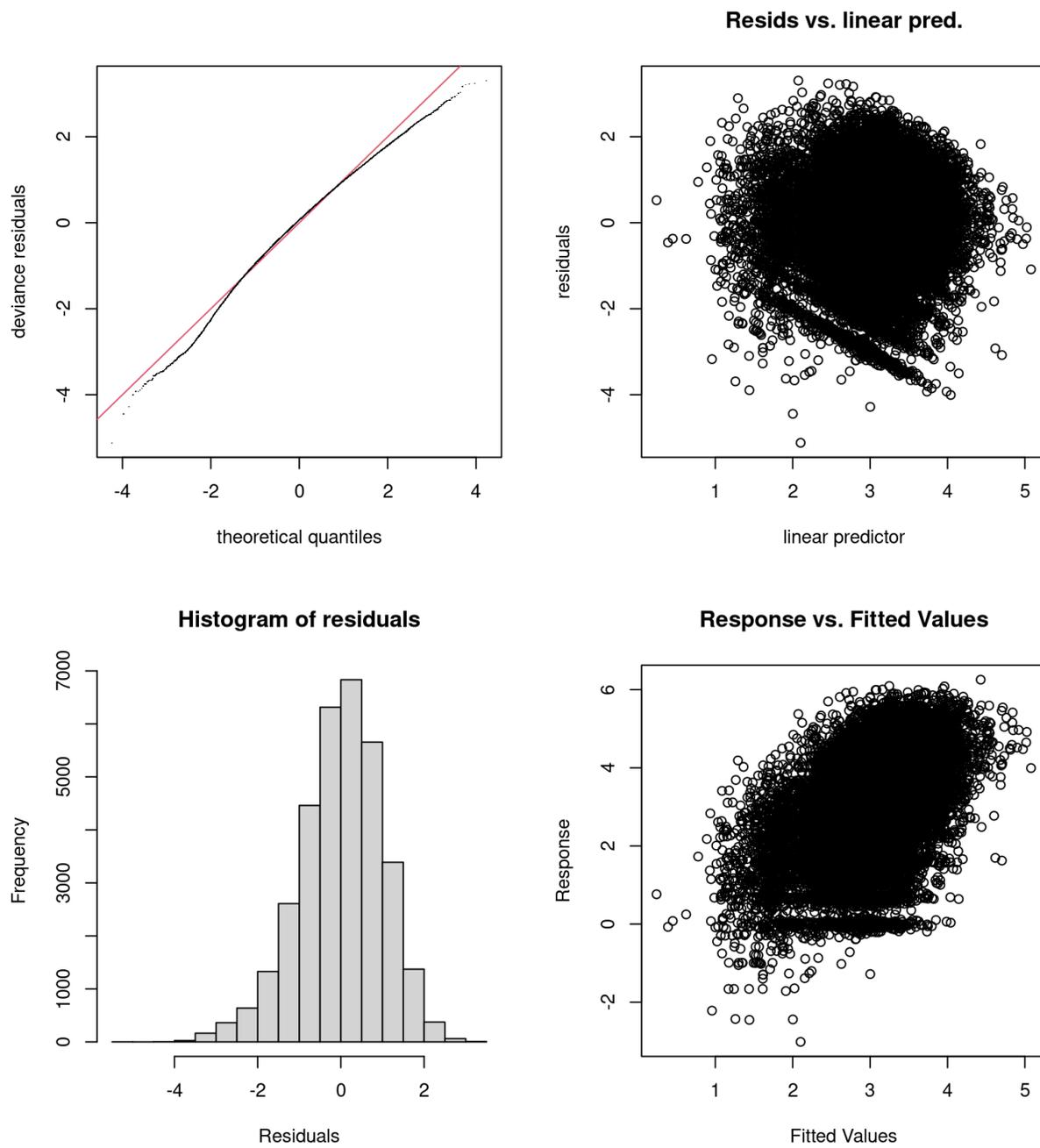
Figure 19: Standard diagnostic plots for the component 3 GAMM. *top-left:* QQ-plot; *top-right:* deviance residuals versus the linear predictor; *bottom-left:* histogram of deviance residuals; and *bottom-right:* Response variables as a function of fitted values.
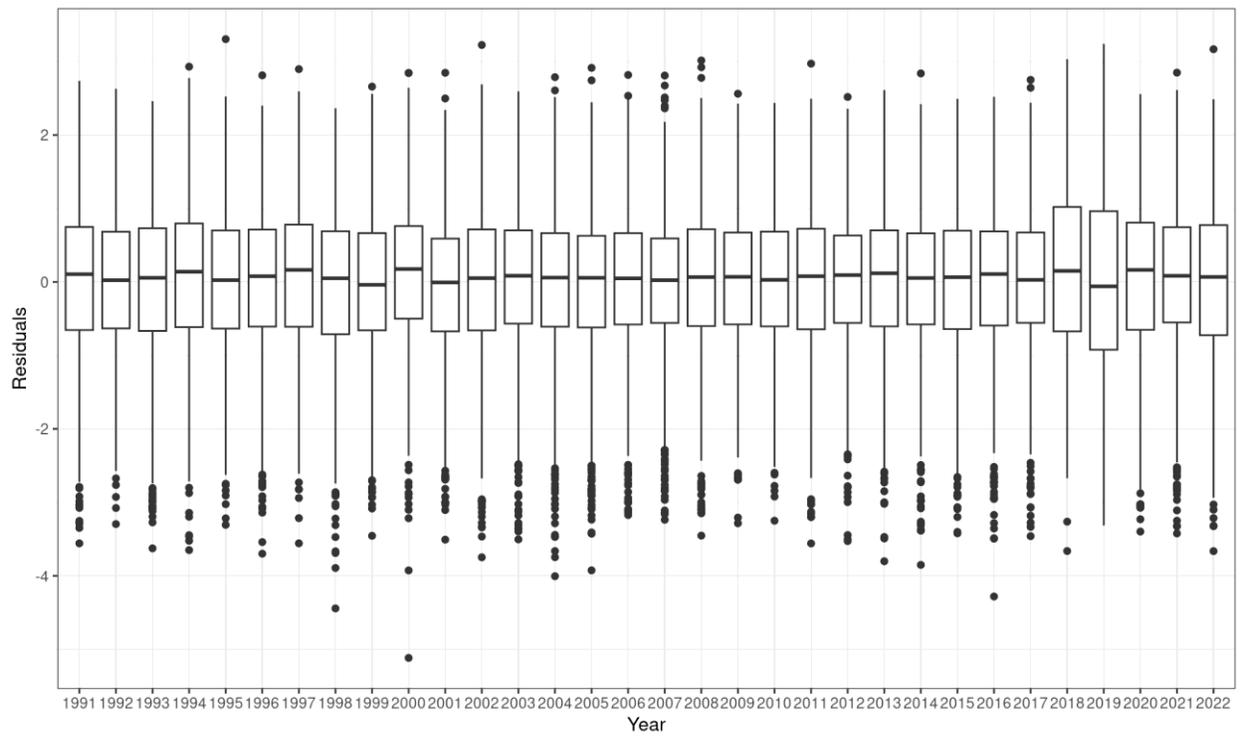
Figure 20: Boxplot of residuals of component 3 model as a function of year.
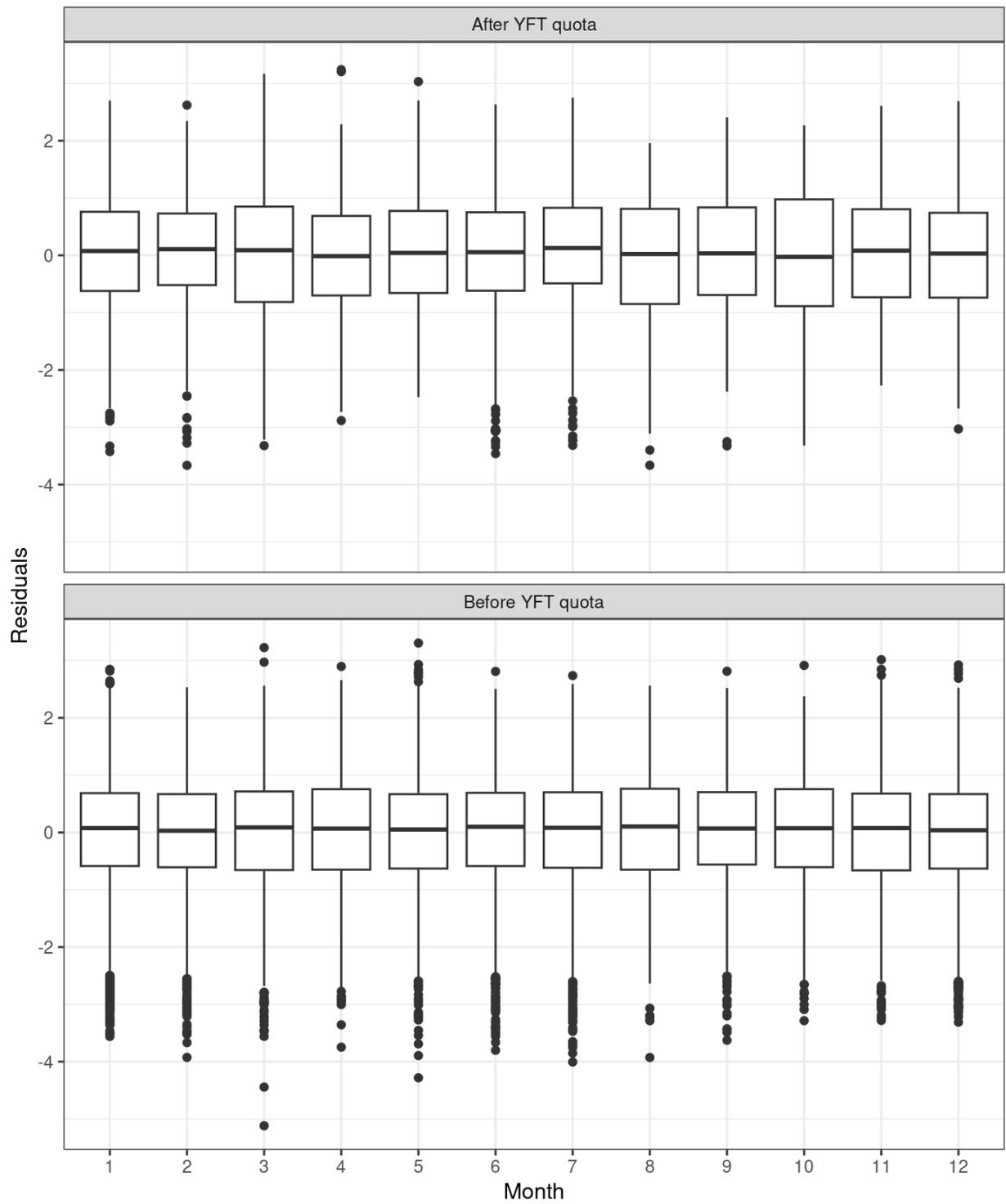
Figure 21: Boxplot of residuals of component 3 model as a function of month before (bottom) and after (top) the implementation of the YFT quota.
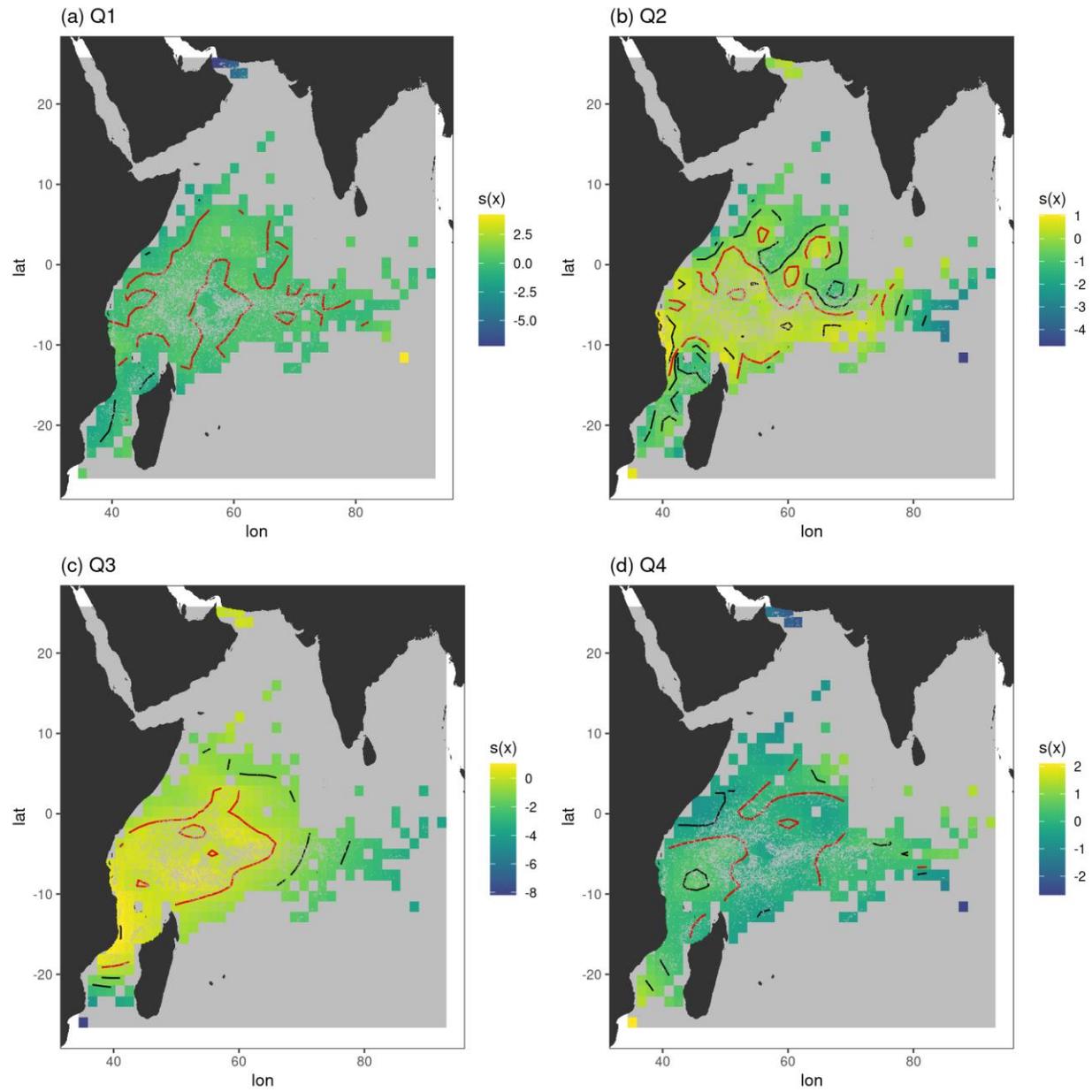
Figure 22: Marginal effect of lon,lat on biomass of adult YFT caught in positive FSC sets for which adult YFT was present (component 3) for each of the four quarters.
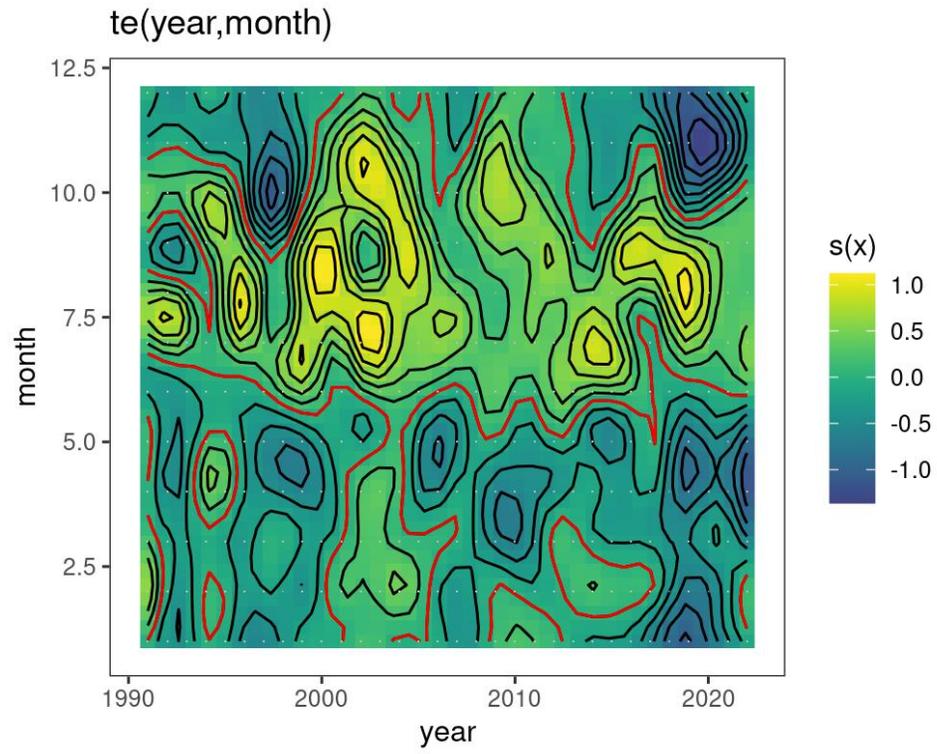
Figure 23: Marginal effect of year,month tensor product smooth on biomass of adult YFT caught in positive FSC sets for which adult YFT was present (component 3).
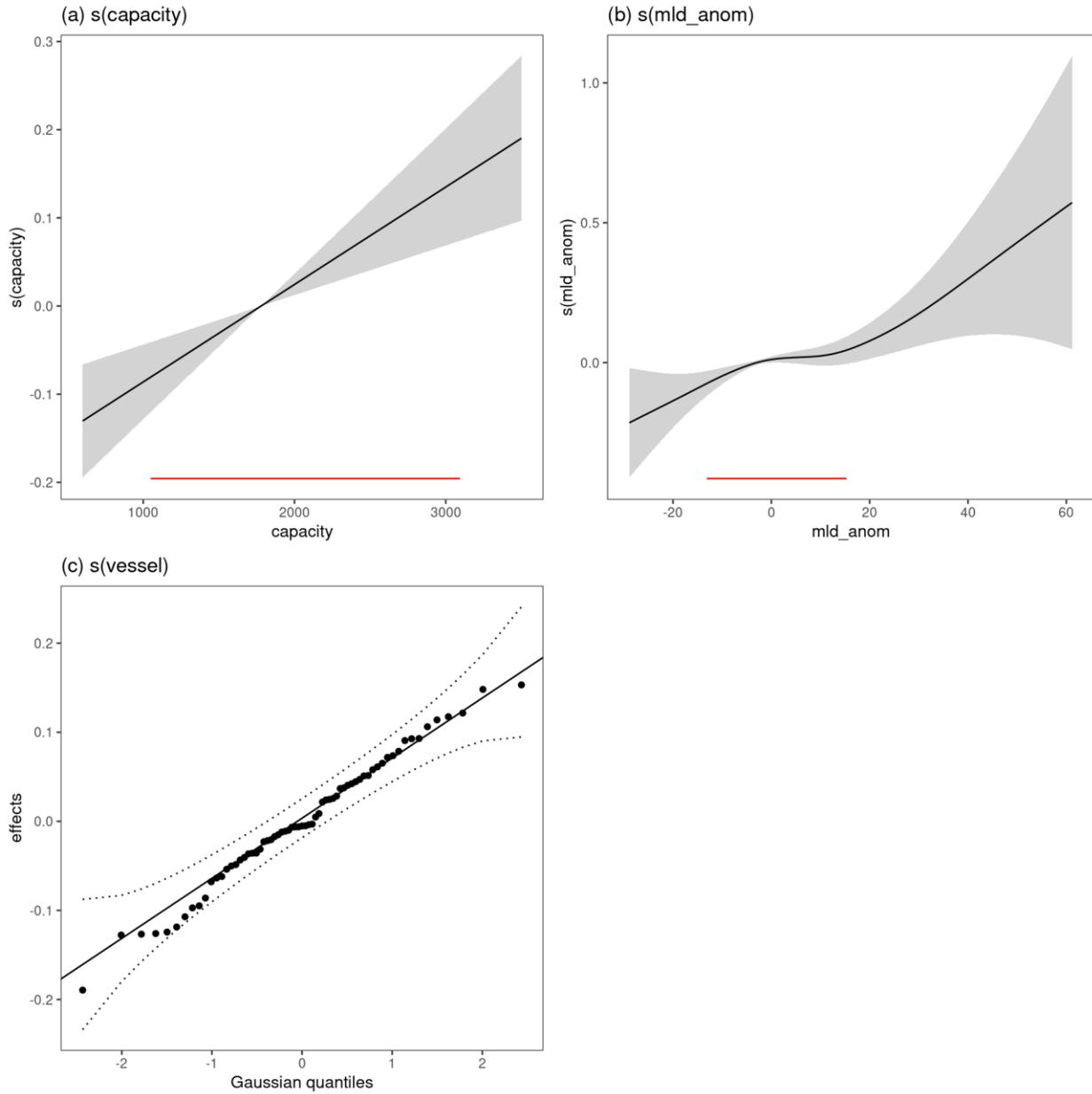
Figure 24: Marginal effects of year and fishing-efficiency-related individual smooths on biomass of adult YFT caught in positive FSC sets for which adult YFT was present (component 3a). The red horizontal bars on the panels indicate the central 95% of the data of the corresponding predictor variable in the model training data set.
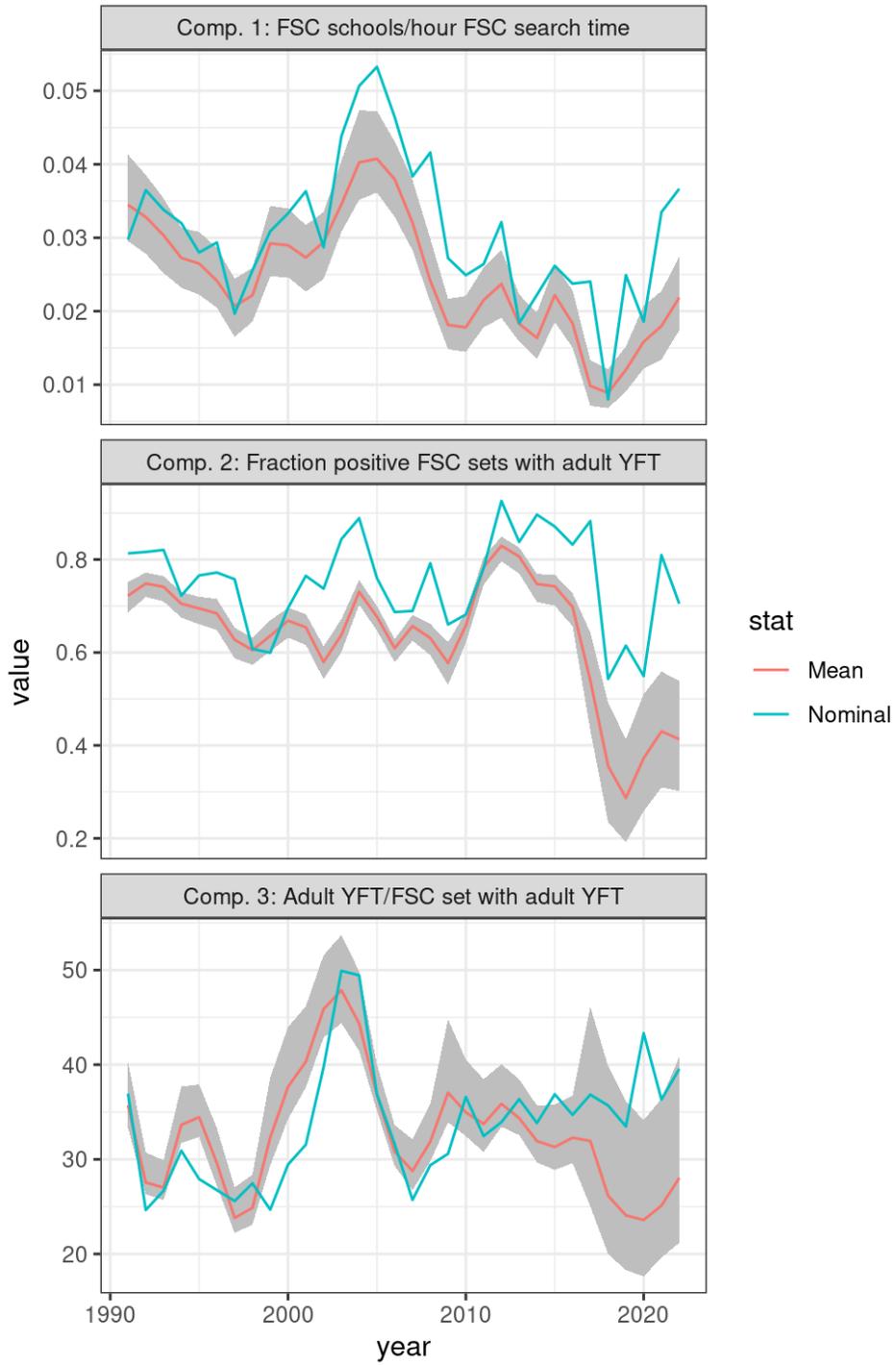
Figure 25: Standardized predictions for each of the three components of our CPUE standardization approach.
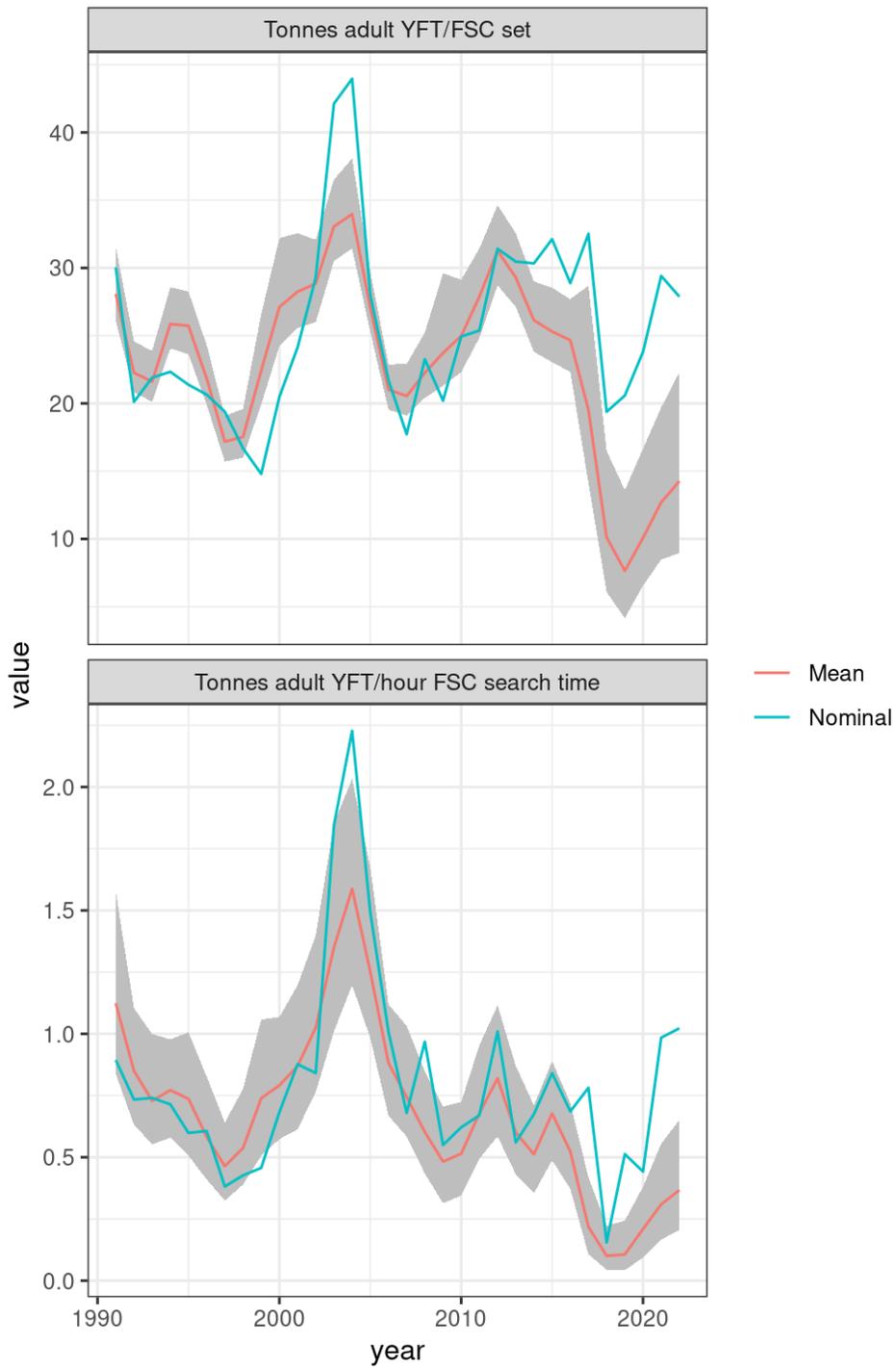
Figure 26: Standardized abundance indices based on combining components 1, 2 and 3 (bottom panel; encounter rate of adult YFT in FSC sets per hour of "FSC" search time), and combining components 2 and 3 (top panel; tonnes of adult YFT caught per non-null set), respectively. Note that for comparison sake the nominal rate of encounter of adult YFT biomass in FSC schools per unit of search time (blue curve in bottom panel) has been corrected for the fraction of null sets in the data by dividing the nominal adult YFT biomass per unit "FSC" search time by the faction of FSC sets that are positive for a given year.
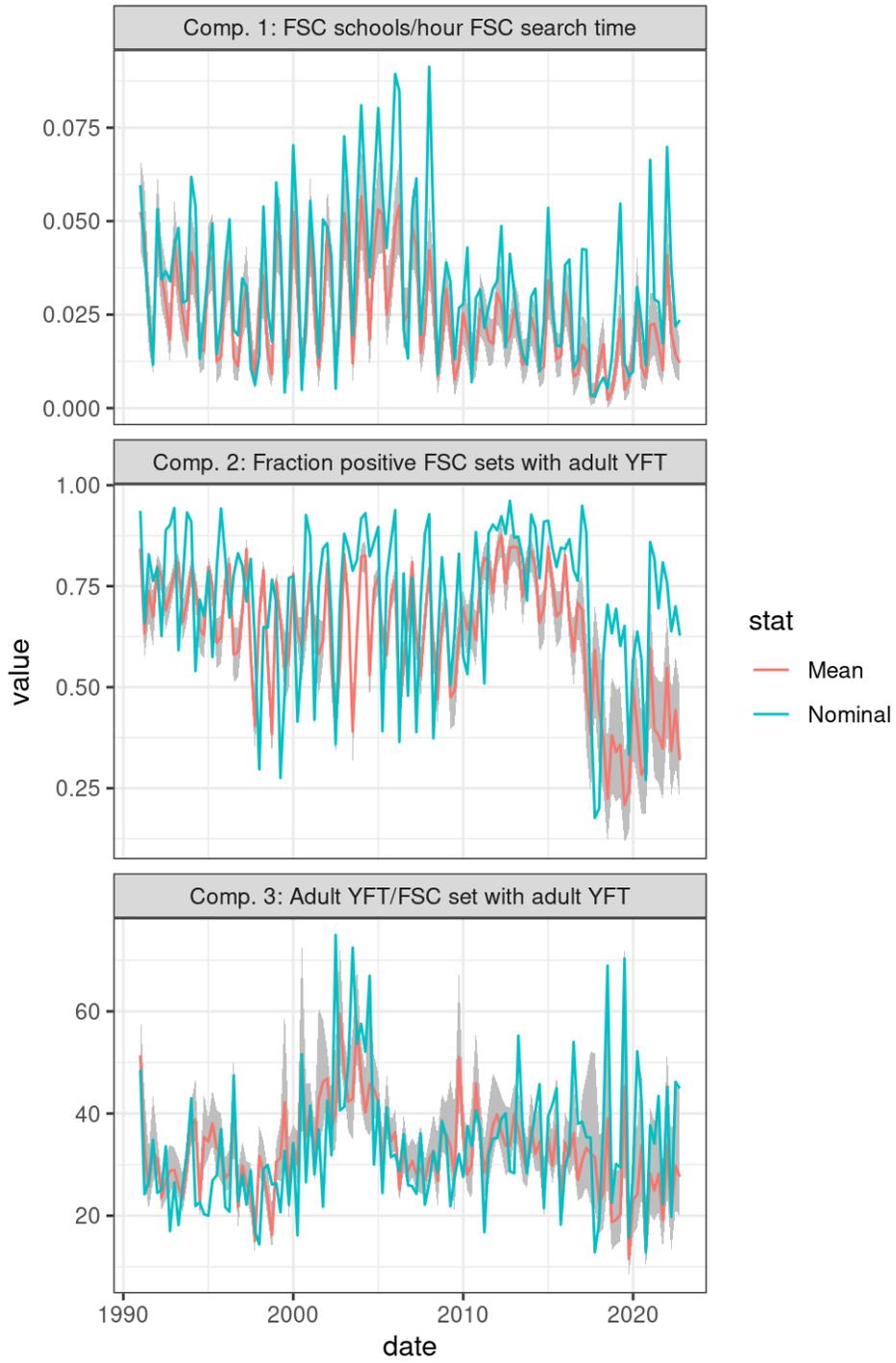
Figure 27: Standardized quarterly predictions for each of the three components of our CPUE standardization approach.
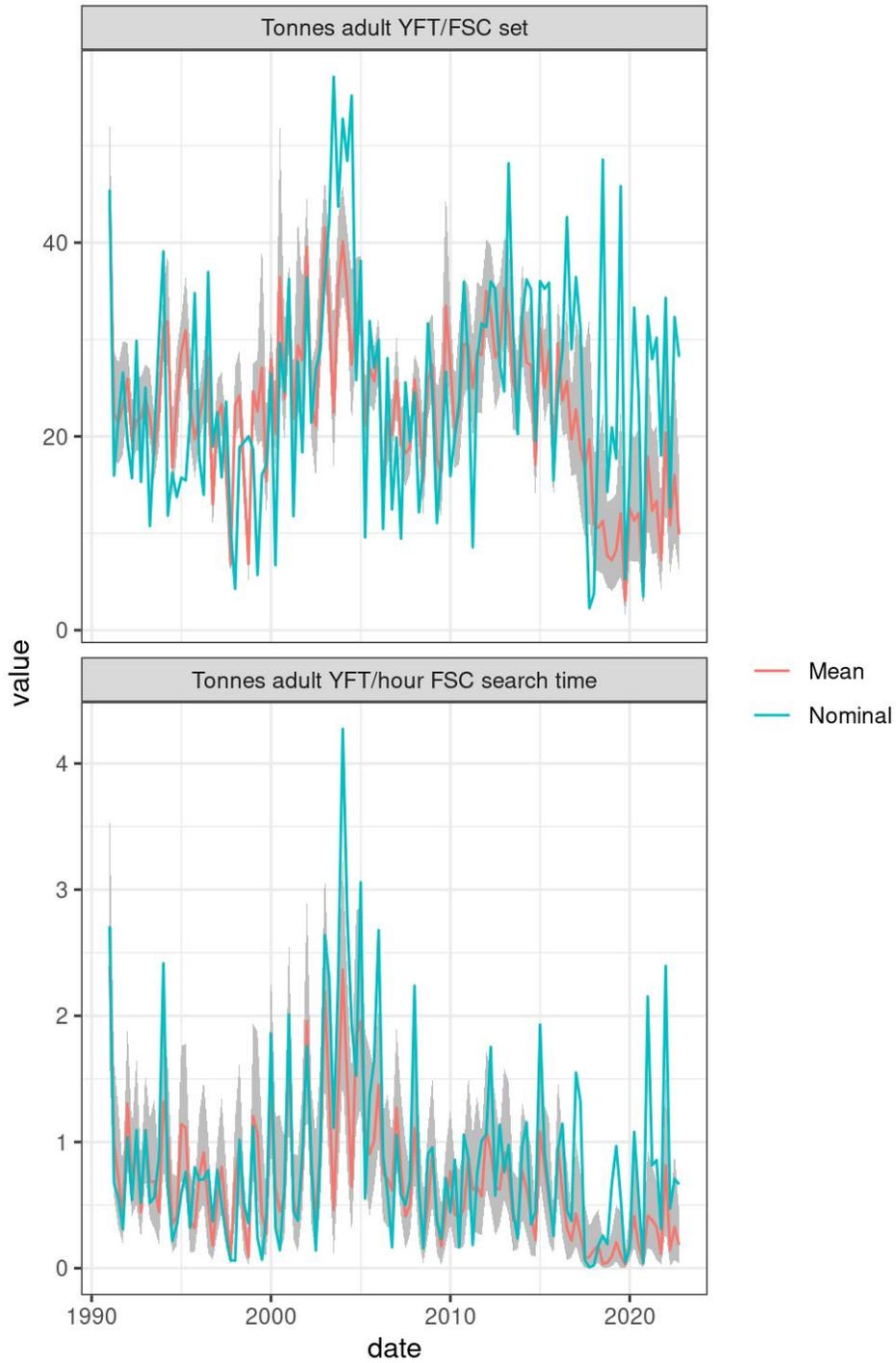
Figure 28: Standardized quarterly abundance indices based on combining components 1, 2 and 3 (bottom panel; encounter rate of adult YFT in FSC sets per hour of "FSC" search time), and combining components 2 and 3 (top panel; tonnes of adult YFT caught per non-null set), respectively. Note that for comparison sake the nominal rate of encounter of adult YFT biomass in FSC schools per unit of search time (blue curve in bottom panel) has been corrected for the fraction of null sets in the data by dividing the nominal adult YFT biomass per unit "FSC" search time by the faction of FSC sets that are positive for a given year-quarter combination.
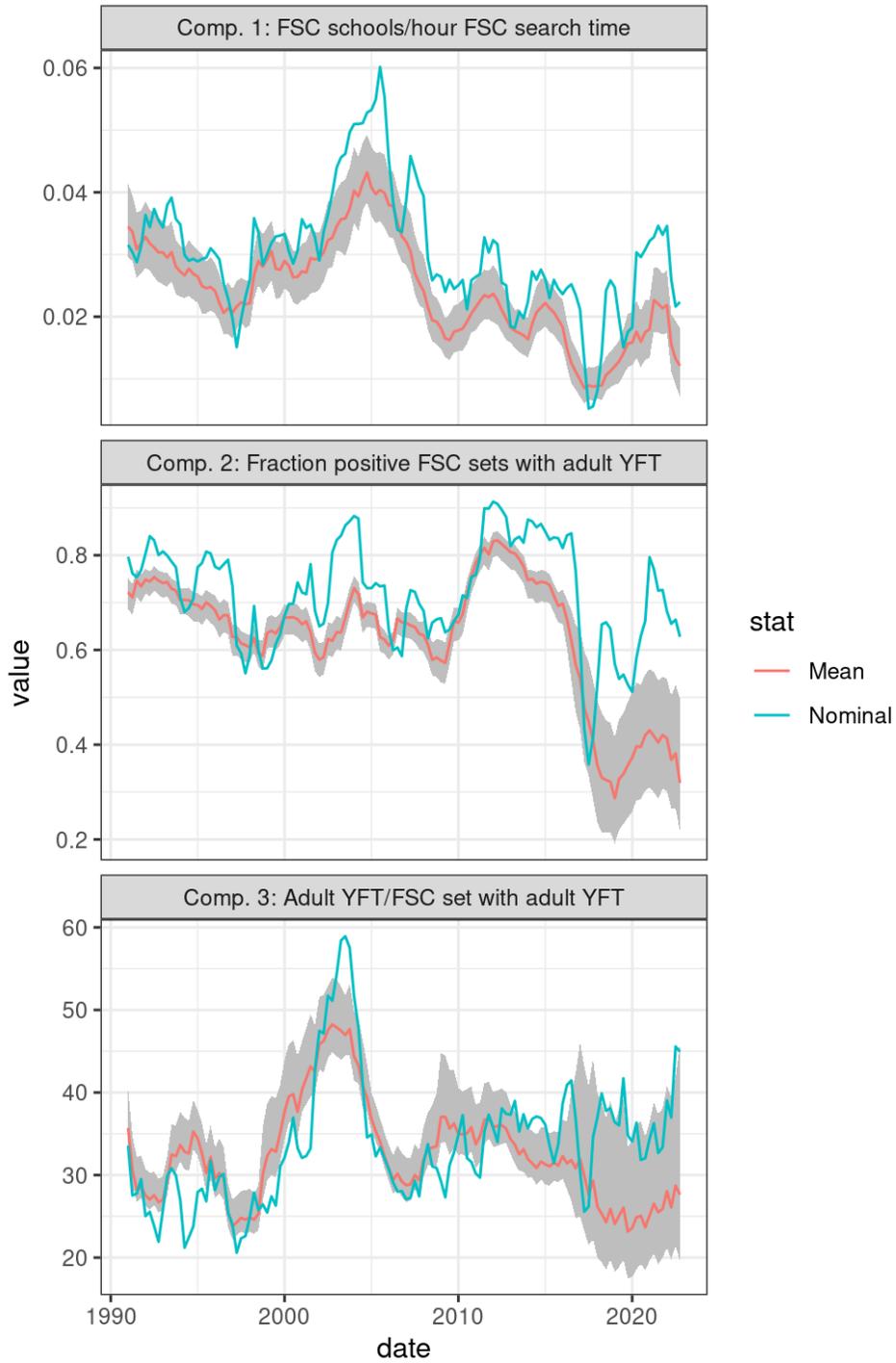
Figure 29: Standardized quarterly predictions for each of the three components of our CPUE standardization approach. A running average has been applied to these indices so that each represents the average over a year (i.e., four quarters) *starting* at the date indicated on the x-axis.
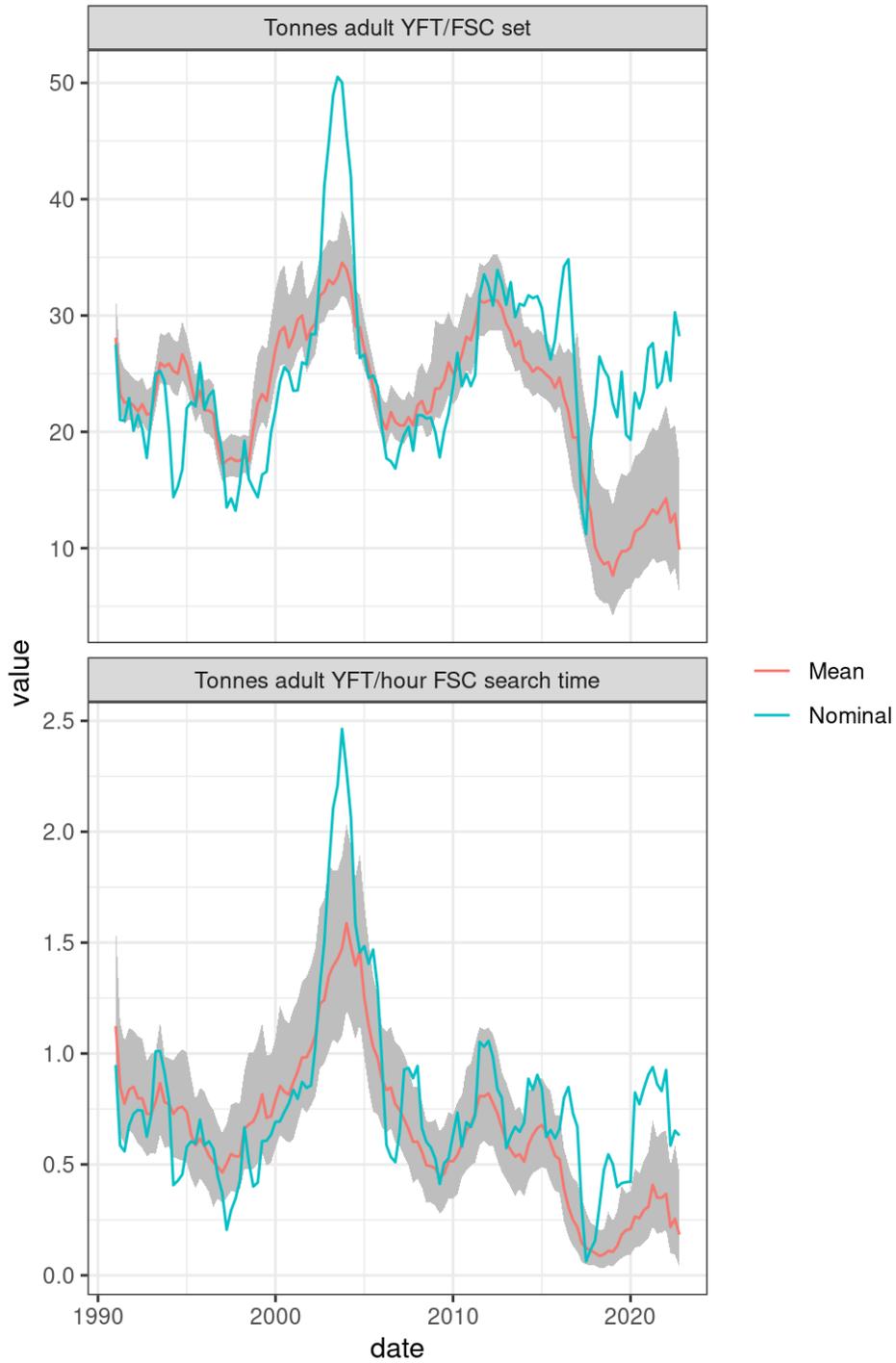
Figure 30: Standardized quarterly abundance indices based on combining components 1, 2 and 3 (bottom panel; encounter rate of adult YFT in FSC sets per hour of search time), and combining components 2 and 3 (top panel; tonnes of adult YFT caught per non-null set), respectively. A running average have been applied to these indices so that each represents the average over a year (i.e., four quarters) *starting* at the date indicated on the x-axis. Note that for comparison sake the nominal rate of encounter of adult YFT biomass in FSC schools per unit of search time (blue curves in bottom panel) has been corrected for the fraction of null sets in the data by dividing the nominal adult YFT biomass per unit search time by the faction of FSC sets that are positive for a given year-quarter

combination.