SCIENTIFIC COMMITTEE
FOURTEENTH REGULAR SESSION

Busan, Republic of Korea
8-16 August 2018

Background Analysis for the 2018 stock assessment of South Pacific albacore tuna

WCPFC-SC14-2018/ SA-IP-07

Tremblay-Boyer L[1], S. McKechnie[1] and G. Pilling[1]

[1] Oceanic Fisheries Programme, The Pacific Community (SPC)

Contents

1 Executive summary

This information paper provides details of the key significant supplementary analyses that supported the 2018 assessment of south Pacific albacore. These include:

- Changes made to the regional structures of the assessment. Compared to the 2015 stock assessment, the regional structure was simplified by combining regions of the same latitude from 140° E to 150° W, resulting in 5 regions in total.
- Changes to the regional structure had consequences for fisheries definitions. The previously 'all flags' longline fleets were subdivided into 'DWFN', 'PICT' and 'AU/NZ' fleets, dependent upon the region, to better model differences in CPUE and length frequencies (selectivities) among different fishing fleets now covering greater areas than in the previous assessment. In turn, 'Index' fisheries were developed which receive CPUE-reweighted length frequencies as well as all-fleet-combined standardized CPUE indices. This new configuration resulted in 21 fisheries (16 capture fisheries + 5 'Index' fisheries).
- Fishery-specific summaries for the catch, effort and length frequency data that are utilised in the assessment, and approaches to processing the length frequency data and construction of the input tagging files, are described.
- Approaches to the standardisation of the CPUE indices and their associated regional weights are described. In particular, the geostatistical approach to CPUE standardisation, and the 'traditional' approach used in previous assessments are detailed, and the results compared. In addition to these methodological changes, the two key differences with the 2015 CPUE analyses were that Japanese longline operational data were included, and that targeting clusters were not used as a filtering variable but as a covariate in the CPUE standardization.
- Finally, the approach to develop the maturity-at-length function for input into MULTIFAN-CL is described.

2 Background information

Assessments of pelagic fish stocks in the western and central Pacific Ocean undertaken by the Pacific Community (SPC) typically require extensive background analyses, for example, investigating certain aspects of the biological or fisheries systems, or investigation of the best methods for including the input data in the stock assessment. Often, these analyses will be the result of methodological advances, the provision of new data sources, or simply the continual progression of research to improve the incorporation of information into the assessments. If these analyses are significant then they cannot be presented within the stock assessment report itself and must become stand-alone or grouped together in one or more ancillary papers.

Another source of important background information which is essential for interpreting a stock assessment is an outline of the fisheries definitions and their data summaries, particularly if there have been changes to the definitions subsequent to the previous stock assessment. These changes might be the consequence of weaknesses observed in the last assessment or subsequent analyses of fisheries data. For example, differences in size frequency data among different flagged vessels might suggest splitting the data into separate fisheries to allow flag-specific selectivity. The second common cause of modified fisheries definitions is changing boundaries of stock assessment regions.

All the major stock assessments of tuna in the WCPO have undergone changes to regional structures in recent years (Rice et al., 2014; McKechnie et al., 2017a; Tremblay-Boyer et al., 2017a), and the pre-assessment workshop held in April 2018 (PAW; Pilling and Brouwer, 2018) recommended the use of a new spatial structure for south Pacific albacore, to simplify the model while also retaining biological realism. This led to significant changes to the fisheries definitions used in the assessment.

This paper is intended to be read in conjunction with the main south Pacific albacore stock assessment report (Tremblay-Boyer et al., 2018). Firstly, it outlines the changes made to the regional structures of the assessment and the consequences for fisheries definitions. It provides fisheries-specific summaries for the catch, effort and length frequency data that are utilised in the assessment. It also presents other analyses of input data in more detail than can be addressed in Tremblay-Boyer et al. (2018), particularly the processing of the length frequency data, standardisation of the CPUE indices and their associated regional weights, and construction of the input tagging files.

3 Regional structure

3.1 Modifications to assessment boundaries

The 2015 stock assessment marked the progression from the single region model used in 2012 to a multi-region (8 regions) assessment model which allowed movement among the sub-populations. The latitudinal boundaries in 2015 were established on the basis of biological hypotheses of seasonal movement, spatial structuring of the population by age, and patterns of fishing activity. The longitudinal boundary at 170° E was established to support regional bioeconomic models that, in-turn, were influenced by the assessment models of the suite of tropical tuna species whose southern boundaries extended into the southern WCPFC convention area.

The PAW recommended the switch to a new spatial structure to simplify the model while also retaining biological realism. This recommendation was based on the fact that there is little to inform longitudinal movement in either the tagging or the size-frequency data. This new structure, denoted the "2018 regional structure" resulted from removing all region boundaries at 170° E and thus effectively pairing regions 1–6 from the 2015 assessment ("2015 regional structure"; Figure 1) into three regions extending from 140° E – 150° W (Figure 2). The 2018 regional structure is thus fully nested within the 2015 structure and hence there are no boundary shifts per se, unlike the bigeye and yellowfin structure changes investigated last year (2017).

The 210° E boundary remains for consistency with other assessments which do not include the convention overlap area, but it should be noted that there are also quite a few tag releases that occur near this boundary so that information on movement between the eastern and central parts of the assessment regions also remains limited.

A consequence of these modifications was that both 2015 and 2018 regional structures were utilised at various stages of the development of the 2018 stock assessment (Tremblay-Boyer et al., 2018). The earlier steps of the progression from the 2015 to 2018 diagnostic models retained the 2015 structure up until model '2018R trad CPUE + JP, all clusters' (see Figures 14 and 15 from Tremblay-Boyer et al., 2018). The diagnostic case as well as all grid outputs were produced with the 2018 region structure.

3.2 Consequences for fisheries definitions

Modification of regional boundaries generally alters the fisheries definitions of an assessment model. Here we address the changes necessitated by the reduced number of regions, while also outlining other changes made to the definitions in an attempt to improve the modelling of the range of input data available for the assessment.

Fisheries definitions for the 2015 assessment were very simple, with one, "all flags" fishery (a fishery that includes data from all vessels fishing the relevant gear) for each gear type in each region where that gear was active (Tremblay-Boyer et al., 2015a). This resulted in one longline fishery in each region (eight in total) and a gillnet and troll fishery in each of the three southern most regions (3, 6, 8; resulting in three fisheries for each gear), giving a total of 14 fisheries. By effectively joining three sets of neighboring regions (1 and 4, 2 and 5, 3 and 6) it is a simple matter of aggregating the affected fisheries within a gear type. Therefore, we are left with only two gillnet (fisheries 12 and 13 aggregated) and troll (fisheries 9 and 10 aggregated) fisheries under the 2018 structure (Table 1).

The definitions for the longline fisheries under the 2018 structure become more complicated owing to dissagregation of flags within regions. This stemmed from a desire to better model differences in CPUE and length frequencies (selectivities) among different sets of vessels from fishing fleets now covering greater areas. This involved splitting the fisheries into three groups depending on if the vessels were flagged to an a) distant water fishing nation (DWFN), b) Pacific Island Countries and Territories (PICT), or c) New Zealand or Australia (NZAU). Instead of one 'all flags' fishery in each region, we established the following longline fisheries:

- Region 1; One DWFN and one PICT fishery,
- Region 2; One DWFN, one PICT and one NZAU fishery,
- Region 3; One DWFN, one PICT and one NZAU fishery,
- Region 4; One DWFN and one PICT fishery,
- Region 5; One DWFN and one PICT fishery

which are detailed in Table 1.

4 Addition of "Index" fisheries

In the 2018 albacore assessment a new approach was developed that utilised "Index fisheries" that are explained in Tremblay-Boyer et al. (2018), which effectively allowed the size data to be used to remove fish correctly but to also use the sizes to inform about fluctuations in population abundance and sizes. To achieve this we established an additional longline fishery in each region. This is considered an all flags fishery and it receives length frequencies spatially reweighted using CPUE (Section 5) as well as all-fleet-combined standardized CPUE. The final number of fisheries under the 2018 regional structure is therefore 21.

5 Scaling of size-frequency data

An ongoing issue with the WCPO stock assessments conducted using MFCL has been the difficulties in balancing competing objectives of modelling the size frequency data. This data is used to firstly remove fish from the population at the correct size, but secondly to provide information on changes in the size structure of the population over time. The difficulty arises because these objectives can seldom be met simultaneously due to the parameterisation of MFCL, which is detailed in McKechnie (2014) and again raised in Section 5. This can be summarised briefly as follows: there appears to be spatial variation in fish sizes within model regions; length samples are collected unevenly in space and time such that the samples must be reweighted spatially using either catch (to be representative of the size of fish being removed from the population) or CPUE (to be representative of the size of fish in the population).

The external review of the 2011 bigeye tuna stock assessment recommended prioritising removing fish at the correct size and so recent assessments of albacore, bigeye and yellowfin tuna have adopted this approach (Harley et al., 2015; McKechnie et al., 2017a; Tremblay-Boyer et al., 2017a).

The 2015 assessment of south Pacific albacore adopted the spatial re-weighting of length frequency data that was used for bigeye and yellowfin tuna in the WCPO (McKechnie et al., 2017b), with Scott and McKechnie (2015) outlining the methods used to collate these inputs. An identical approach was utilised for the update of the 2018 assessment with three extra years of data.

Catch and length frequency data for albacore were extracted from SPC data holdings and separated into assessment regions (Figures 1 and 2, depending on the step). There are known to be unexplained shifts in the length frequency distributions for data collected at Pago Pago prior to 1972. These data were removed prior to conducting the analysis.

The length frequency data are collected and made available at a range of different spatial scales. For albacore the majority of the length frequency data are collected at either the scale of $5°\times5°$ cells or $10°\times20°$ cells, a remaining small portion are at either $5°\times10°$ or at approximate $10°\times20°$ cells.

Some of the $10°\times20°$ cells overlap two or more region boundaries. Data for cells that overlapped two region boundaries were divided by 2 and data for cells that overlapped four region boundaries were divided by 4. The size data could then be included in each region without giving greater weight to the overlapping cells.

The motivation for adding Index fisheries to the assessment was to better model the size composition of the stock, rather than the size composition being removed from it. To achieve this, it is therefore desirable for the length frequency samples to be weighted such that the samples from subregions of high abundance are given higher weighting than those from subregions with few fish. The process follows directely from Section 5.1 except that instead of catch, the $W_{i,t}$ now denote the CPUE for cell i, and year-quarter t.

5.1 Re-weighting by Spatial Cell

The calculation of spatial cell weightings and subsequent re-weighting of size composition data followed the approach described in McKechnie (2014). The procedure is described briefly below.

1. Cell weightings were calculated as per Equation 1 by summing the catch ($C_{i,t}$) of albacore (in numbers) for each 10°×20° cell (i=1:M) in a given year-quarter (t=1:T) and dividing it by the total catch for all cells in that year-quarter within an assessment region. The same approach was used when reweighting by CPUE, except that in this case the base unit of the cell-weight was the nominal CPUE in individuals per hook.

2. An additional temporal smooth was applied to the weighting function such that the weighting ($W_{i,t}$) was calculated over a moving time window spanning t-k year-quarters to t+k year-quarters. This was imposed in an attempt to prevent the assessment model from interpreting short term variation in lengths compositions (that may have resulted from variation in distribution of the catch) as variation in the size composition of the population.

$$W_{i,t} = \frac{\sum_{t-k}^{t+k} C_{i,t}}{\sum_{i=1}^{M} \sum_{t-k}^{t+k} C_{i,t}} \tag{1}$$

3. Re-weighted length frequencies were calculated by multiplying the numbers in length bin l (1cm bin width), in each cell-year-quarter by the ratio of the cell weighting to the total number of fish sampled in that cell-year-quarter (Equations 2,3).

4. An upper limit of 1,000 fish was set for a year-quarter to prevent very high sample sizes in some year-quarters from being overly influential. This was achieved by multiplying the area weighting by 1000 and then rescaling the ratios of cell weighting to total sample number to a maximum value of 1.0 within a region and year-quarter for those instances where ratios greater than 1.0 were achieved.

$$p_{i,t} = \min \left[\frac{W_{i,t}.1000}{\sum_l n_{i,t,l}}, 1.0 \right] \tag{2}$$

$$N_{t,l} = \sum_i n_{i,t,l}.p_{i,t} \tag{3}$$

In addition to the procedure described above a limit on the lowest allowable weighting was imposed to reduce the influence of size data from cells with very little catch. For example if a limit of 0.3 were set, only those cells with normalised weightings of 0.3 or greater would be included in the re-weighting procedure. The choice of limit was somewhat arbitrary and based on the trade-off between preventing the loss of data from too many year-quarters (in the event of a high limit) and conversely, preventing undesirable temporal variation in the size compositions when two or more cells with very different size compositions contribute to the region-scale size distribution.

Note that this reweighting process was undertaken separately for each fleet within a region as defined in Table 1. For example, if there was a PICT, DWFN and AUNZ fishery in a region then the aggregation of catch and sizes in each subregion was undertaken only on data for that individual fishery. The resulting time series of length frequencies were applied to the appropriate fishery in the model, with the Index fisheries instead receiving length frequencies reweighted using CPUE rather than catch, which will be outlined in the following section.

The impact of the re-weighting procedure are summarized in Figures 19 to 28. Figures 19 to 21 show how individual samples of length in a given-quarter are modified, but also highlight that the overall temporal trend stays mostly similar except for instances where the temporal coverage of

the length-frequency data is small. Figures 22 to 24 compare the overall size distribution before and after reweighting. Figures 25 to 27 show the quantile distribution of the length samples by fleet category under before and after reweighting. Finally Figure 28 shows the impact of catch *vs.* CPUE weights for the all-fleet aggregated length samples which are used for the Index fisheries.

6 CPUE

Two separate methods were used to produce standardised CPUE indices as inputs to the 2018 stock assessment models. 'Traditional' methods that were identical to those adopted in 2015 were used to produce indices in the early stages of the model development progression, as well as a development of that approach within the CPUE model grid axis. The methods for these are presented in Section 6.3. The second approach was based on geostatistical (spatio-temporal) models that were previously used for bigeye and yellowfin tuna (McKechnie et al., 2017b; Tremblay-Boyer et al., 2017b). The resulting indices were used later in the progression for the stock assessment model, as well as the second level of the CPUE axis. The methods used to estimate these indices are outlined in Section 6.4.1. CPUE were used as a grid axis because this is the first year where we use geostats CPUE methods beyond exploratory analyses, and also because the modelling approach makes different implicit assumptions about stock mixing throughout the range since 'traditional' cpue analyses are applied on a by-region basis (i.e. standardized indices in one region are independent from those of another region), whereby the geotatistical model is applied to the entire assessment region such that by definition, a stock-wide index is fitted to which regions interactions are added (i.e. high stock mixing).

6.1 Pacific-wide longline operational dataset

We use the term operational-level longline data to refer to records of fishing activity at the scale of the longline set. The operational-level longline data set used in this analysis was first described in McKechnie et al. (2015) and has been updated annually since. It contains individual records of fishing activity, whereby on a given day and time a longline was set by a vessel in a particular location at the one degree resolution, and the resulting numbers of fish caught of the four species most likely to be targeted in longline operations: albacore tuna (ALB), bigeye tuna (BET), yellowfin tuna (YFT), and swordfish (SWO). The longline set itself is characterized by a total number of hooks set and by the number of hooks that are set between each intermediate float deployed along the length of the line (hooks-between-floats or HBF). These intermediate floats are used to maintain the fishing gear at a particular depth in the water column. The data set, covering the breadth of the Pacific Ocean from around 45° N to 40° S and over sixty years of fishing (1952-2016) and both domestic and distant-water fishing fleets, comprised more than 10.5 million sets, and when subsetted for the assessment region (Figure 2), 3.4 million sets. Table 2 summarizes the variables present in the dataset.

Maps of average effort and CPUE by one-degree cell by decade are shown in Figures 45 and 46.

6.2 Identification of targeting clusters

The first step of the analysis was to assign individual longline sets to a species targeting group, so that non-albacore targeting sets could be accounted for within the standardization. We used

clustering to classify sets, assuming that in general, sets that target a specific species should have a higher proportion of that species in their catch. Cluster analysis has been recommended by several authors to formally generate this set-wise classification (He et al., 1997; Hoyle et al., 2014), and has been used in previous standardization of bigeye and albacore CPUE in the WCPO (Tremblay-Boyer et al., 2015b; Bigelow and Hoyle, 2012). The clustering analysis relied on the aggregated proportion of the four available species (ALB, BET, SWO and YFT) caught by the same flag in a one degree cell in a given month-year in a region (considered to be a 'trip' for the purpose of clustering). Unlike previous CPUE standardization analyses for South Pacific albacore (Bigelow and Hoyle, 2012), we had access to swordfish catches and chose to include these as there appeared to be significant swordfish targeting events within the dataset. Set composition was aggregated at the trip-level (i.e. flag-cell $\times$ month) to remove between set variations in catch, and better reflect the overall targeting strategy. We used this level of aggregation instead of actual fishing trips as some these can extend over multiple months and target switching may occur over this period (see McKechnie et al., 2015 for more details on the choice of 'trip' for clustering).

Based on previous analyses of these data by McKechnie et al. (2015), we used k-means clustering with the Hartigan-Wong algorithm from the R package stats (R Core Team, 2018), which is efficient when working with large datasets. We note, however, that numerous alternative clustering algorithms can be used in such analyses, and changes in the assignment of sets to targeting clusters under different algorithms cannot be ruled out. The k-means algorithm assigns observations to a pre-set, user-defined number of groups ('clusters') to maximize the Euclidean distance between the group means ('cluster centre', the mean proportion of each species in the catch of sets belonging to the cluster). For each region we ran the algorithm with 2, 3 and 4 clusters since we only had 4 variables (one for each target species). We found that, in general, the three-cluster configuration did the best at capturing key spatial and temporal patterns in species targeting. For 2018 regional structure regions 2 and 5 though, the three-cluster configuration added little extra value so we retained the two-cluster classification of these regions. Each set was thus assigned a targeting cluster ID, and this variable was used as a covariate in the CPUE standardisations described below.

Clustering results are shown here from Figures 29 to 44 at various levels of aggregation, and comparing the two- to four-cluster levels of classification. Figures 29 to 33 show the average composition of targeting cluster under each of the configuration, highlight that most sets in tropical regions (1 and 4) are assigned to bigeye and yellowfin targeting, whereas sets in more temperate regions (2, 3 and 5) mostly target albacore, with a small proportion in regions 3 and 5 having distinct swordfish targeting. Figures 34 to 38 show the temporal trends in the proportion of trips in cluster over time, including the increase in albacore targeting in tropical regions (1 and 4) in recent decades. Figures 39 to 44 show the spatial distribution of clusters over time under the three-cluster configuration.

6.3 'Traditional' CPUE analysis methods

We are deriving CPUE for albacore tuna, but having access to the catches of all four species yields insight into the vessel's targeting behaviour, and can inform the exclusion, or retention, of the set in the standardization analyses. In 2015 the targeting cluster (see above) was used as a filter to only retain sets targeting albacore, which was problematic in tropical regions where for some year-quarters there were very few sets remaining after this filter was applied. Here, we instead use cluster as a covariate in the standardization, except for sets in swordfish clusters (occurring in 2018 regions 3 and 5 only), which show clear targeting patterns and were removed.

The other key difference from the standardization performed for the 2015 assessment is that Japanese longline operational data were included in the current analysis, which allowed the historical coverage of the dataset to be extended significantly.

These two key differences in the standardization were explicitly mapped as steps in the stepwise analysis for the assessment, see Figures 14 and 15 in Tremblay-Boyer et al. (2018).

Data grooming was applied prior to the standardization to remove sets outside of the temporal or spatial span of the analysis, or with improbable records. Sets with missing logdates, sets with 0 hooks, sets where there were more albacore caught than the number of hooks, sets from years before 1960, sets outside of the regions defined for the stock assessment, and sets where there was not at least one individual of ALB, BET, YFT or SWO caught, were all excluded from the dataset. Additional filters were applied to ensure that covariates used in the analysis reached a minimum sample size over the time-span of the assessment. More specifically, year-quarter effects were only estimated for year-quarters with at least 50 sets, 5×5 cells were only retained if they had at least 20 sets, and, when vessel effects were estimated, only sets belonging to a 'core fleet' were kept within which vessels were active for at least 4 quarters with 20 sets or more.

Once these filters were applied we proceeded with the standardization by first fitting GLM models to each region independently with negative binomial or delta-lognormal errors as described below, then extracted the year-quarter effects from these GLMs to use as standardized indices for a given region. All GLM models were fitted in R *via* the TMB package (Kristensen et al., 2016).

6.3.1 Negative binomial error distribution

Albacore catch-in-numbers, n_i, is modelled with effort as an explanatory variable with

$$n_i \sim \text{NegBin}(\mu_i, \theta) \tag{4}$$

$$\log(\mu_i) = \beta_0 + \beta_{yq[i]} + ... + \beta_H \log h_i \tag{5}$$

where μ_i is the mean albacore catch in count for set i, θ is the size parameter, $\beta_y q$ are the year-quarter coefficients, and '...' are coefficients for the levels of additional factor variables included for set i. Parameter β_H is the coefficient for a linear relationship between log mean count and $\log h_i$, the logarithm of hooks (divided by 100) fished for set i. We parameterize the negative binomial likelihood based on the mean μ_i and variance $V_i = \mu_i + \mu_i^2/\theta$.

6.3.2 Delta-lognormal/hurdle models

The probability of having a set with non-zero catch and the CPUE of the catch when positive are modelled separately using a binomial and a log-normal distribution, respectively. The binomial GLM uses a binary response variable (y_i ; 1 =≥ 1 fish caught, or a 0 = zero fish caught in set i)

$$y_i \sim \text{Bernoulli}(p_i) \tag{6}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{yq[i]} + ... + \beta_H h_i \tag{7}$$

where p_i is the probability of at least one albacore individual being caught in set i, the logit link function is used to express this probability in terms of the linear predictor and the model coefficients

have the same interpretation as in Section 6.3.1. β_H is the coefficient for the continuous variable of hooks fished (divided by 100) for set i on the natural rather than log-scale used above for the Negative binomial.

The positive component of the delta log-normal model was them modelled using a Gaussian GLM given by:

$$\log c_i \sim \text{Normal}(\log \mu_i, \sigma^2) \tag{8}$$
$$\log \mu_i = \beta_0 + \beta_{yq[i]} + ... \tag{9}$$

where $\log c_i$ is the log-CPUE of ALB (number caught divided by hundred hooks) and the parameters in the linear predictor are interpreted as above.

Model fit was assessed by observing the distribution of residuals against model assumptions, and ensuring there were no persistent trends against the covariates used in the model. Models with negative binomial error distributions had overall better diagnostics which is also what was noted in 2015. Therefore we proceeded with these models to use as final indices for each region, under the full covariate configuration (cluster + vessel) for consistency with the 2015 approach. Standard errors for each year-quarter were estimated using the canonical approach as described in McKechnie et al. (2015).

Stepwise models showing the impact of the addition of successive covariates for the 2018 regional structure are shown in Figure 47. Diagnostics for the final models are shown in Figures 48 to 52.

6.4 Geostatistical CPUE standardisation

6.4.1 Summary of geostatistical model

Geostatistical approaches to CPUE standardization have become more prevalent in recent years, due in part to improvements in algorithms which have made them more computationally efficient, and increased exposure in the fisheries literature (Thorson et al., 2015; Shelton et al., 2014, see also Petitgas, 2001 for an earlier discussion). Geostatistics explicitly models the spatial relationship between the response variable, that is, the fact that observations that occur closer in space are more likely to be similar. This allows the spatial autocorrelation to be removed, which increases the precision in estimates and makes it easier to identify a relationship between response and candidate explanatory variables.

The approach we use here is adapted from the R package SpatialDeltaGLMM developed by Thorson et al. (2015). We give a brief description of how it is applied to the operational longline dataset below, and refer the reader to the original reference for more technical details (see also Thorson et al., 2016). The key difference with the traditional model is that the 5×5 cell effect is modelled as a 'geostatistical surface' instead of each cell effect being estimated independently of one another. Also, sets from all regions are included in the same model, instead of running individual GLM models for each region.

The geostatistical surface ω_s was fitted assuming a Matérn covariance matrix which is used to model the spatial autocorrelation, i.e. how the correlation between observations changes as distance between them increase, with anisotropy (which means the relationship does not have to change in

the same rate in all directions) and with a temporal interaction fitted as a random effect, $\epsilon_{s,t}$. The surface also requires the definition of knots s which are points where the effects are estimated (shape of the correlation surface between knots assumed to be piecewise linear; see Figure 53 for the configuration used in the current analysis). Each observation in the dataset then gets assigned to the knot which is the closest to them. The ensemble of these knots are referred to as a mesh and can be configured to have different features, notably in how densely knots are distributed in edge areas. As for normal CPUE standardizations, covariates can be added and we used here the targeting cluster (obtained from the analysis described in Section 6.2) and the vessel (but noting this last covariate was not retained in the final analyses, see Section 6.4.4).

A delta-lognormal GLM was then fitted to the occurrence and catch rates observed for each set. We explored alternative error distributions, notably the negative binomial for consistency with the other indices developed here, but the delta-lognormal error is the one that converged the most reliably and was therefore retained for final indices.

The prediction for the response variable in their respective link space (binomial: logistic; positive: log), corresponding to observed set i, is:

$$P_i = \beta_t + \omega_s + \epsilon_{s,t} + \zeta_c + \nu_v$$

whereby, for each observation i, β is the year-quarter coefficient at t, ω is the coefficient estimated by the geostatistical surface for knot s (i.e. the knot to which observation i is the closest), ϵ is the time-interation effect for the geostatistical surface at time t, ζ_c is the targeting cluster assigned to the set, and ν_v is the vessel ID. Predictions of standardized abundance for i then excludes the value for the covariates linked to catchability, here the cell, the targeting cluster and, when applicable, the vessel, but otherwise retains the other predictors of density in space and time:

$$P_{k,t} = \beta_t + \epsilon_{s,t}$$

Density at knot k is then the product of back-transformed (R) binomial and positive effects:
$D_{k,t} = R_{bino|k,t} \times R_{pos|k,t}$

All analyses were performed in R (R Core Team, 2018) and GLMs were run with TMB (Kristensen et al., 2016) using algorithms developed in the libraries SpatialDeltaGLMM (Thorson et al., 2015) and INLA (Rue et al., 2017).

6.4.2 Subsampling

The large size of the longline datataset made it impractical to use it in its entirety for the algorithm to run in a reasonable time. The density of observations is much higher towards the tropics and declines towards higher latitudes, and towards the east. The eastern component is especially under-represented. Since geostatistical approaches make the assumption that the value of the response variable is independent from the intensity of sampling, we elected to distribute evenly-spaced N_K knots in the dataset using a k-means algorithm on the extent of the surface covered by observations (see Figure 53). An alternative would have been to distribute knots so that they represent effort

densities (i.e. knots around the equator would have been closer together than in temperate areas) (see discussion in Pilling and Brouwer, 2017).

For a given model run we specified a number of sets N to be retained and these were distributed evenly across the N_k knots. This effectively resulted in a lower sampling rate for knots in highly fished areas, but allowed more observations to be retained at the edges of the distribution. It was also a way to reduce the violation of the model assumption that sampling rates are independent from the response variable. All results presented here used the knot-subsampling approach.

For exploration purposes and to allow the model to run in under 1 hour, we first used a very low subsampling rate where only 10,000 records were retained, but, unless otherwise mentioned, all models presented here had at least one million records in the dataset (i.e. $\sim$ 33% sampling) with 200 knots under mesh configuration shown in Figure 53.

Figure 54 presents the proportion of effort remaining by knot under the sub-sampling scheme used to produce the final indices.

6.4.3 Standardized indices of abundance

Given the spatial effect is estimated at the level of the knot and there are many knots in a region, the lognormal and binomial components by knot for the region first have to be combined before the final indices for each region are built using the usual delta-lognormal approach. This is done by weighting the year-interaction effect of each knot in the region as a function of the knot-area *within* the region (since knots near region boundaries can have cells in more than one region). The year-quarter lognormal or binomial effect for a region thus becomes the stock-wide year-quarter index β_t added to an average year-quarter interaction for the region, obtained from a knot-area weighted mean year-interaction $\epsilon_[s,t]$ for the knots present in the regions.

The area of the knot in a region is the sum of the areas of the region's one-degree-cells assigned to the given knot. The area for each 1 degree cell was obtained *via* the PBSmapping package (Schnute et al., 2017) which calculates area in km^2 by first converting coordinates into northings-eastings.

Final standardized indices of abundance for the albacore assessment were obtained by extracting all fitted effects for an observation except those that impacted catchability (i.e. the cluster and, when applicable, the vessel effect), building year-quarter-region mean effects as described above, back-transforming them and using their product following a standard delta-lognormal standardization approach. The indices were built directly within the TMB algorithm which allowed to estimate standard errors by year-quarter-region directly as part of the modelling procedure (see Thorson et al., 2016 for more details).

The overall year-quarter-region effect for the lognormal or binomial component of the model is the sum of all densities at t weighted by the knot area within each region:

$$I_{t,R} = \beta_t + \sum_{k=1}^{N_R} \frac{A_k \times \epsilon_{k,t}}{A_R}$$

where N_R is the total number of knots in region R and A_R is the region area.

Stepwise models showing the impact of the addition of successive covariates are shown in Figure 55. QQ-plots for the final models are shown in Figure 59, and a comparison of the predicted catch rates

per decade *vs.* those observed is shown in Figures 60 and 61.

6.4.4 Inclusion of vessel ID

One variable that has consistently been found to be influential in standardizations is the vessel identifier, that is, a variable that identifies which fishing sets are performed by a given vessel. This variable represents an array of operational features for which the fine-scale data are frequently not available, such as vessel features, gear type/configuration used, the experience of the skipper, the fishing strategy used by the crew, etc.

The extended time series of operational data used here brought new challenges in terms of including a vessel effect. In many regions, the bulk of the effort from 1950 to the 1980s does not have unique vessel identifier information. This problem also occurs in other t-RFMOs (Hoyle et al., 2016) and has proved an ongoing issue to the calculation of both point-estimates and measures of variation for standardized indices since this level introduces an implicit temporal bias. Although attempts were made by the nations concerned to investigate these information gaps (outlined in OFP (2015) and further discussed in Pilling and Brouwer (2017), further information could not be obtained, although some additional information on fleet ID (e.g. coastal, offshore, distant water) is available.

The temporal bias introduced when including the missing vessel ID becomes especially prevalent when the standardization is done at the scale of the assessment region, as done here under the geostatistical approach. When traditional CPUEs are fitted by a region, the missing vessel IDs essentially get a different level for each region, which at least means that there is some regional differentiation in how this level is estimated, but here the missing vessel has a single stock-wide effect that has a significant impact on standardized indices (Figure 55, rightmost column). Essentially, the early drop in CPUE seen in most regions becomes sharply pronounced, with indices sometimes tripling in value for a given year-quarter. This can be linked directly to the missing (usually Japanese) IDs: in short, the missing Japanese vessel level is assigned a very low effect compared to other vessels as it 'operates' throughout the South Pacific for 3 decades, presumably targeting a range of species that are not albacore. In contrast, starting in the 1980s many more vessel IDs enter the fleet (Figure 58), such that the estimated vessel effects are much more precise and representative of a single set performed by that ID, since the effort represents a single vessel instead of an entire fleet. Once vessel IDs appear, the average vessel effect estimated thus starts rising for both the binomial and lognormal components (Figures 56 and 57). Since those vessels effects are then standardized out, it results in a sharper drop in the CPUE.

In the immediate future, these issues could be improved on by fitting missing vessel ID $\times$ region factors instead of a single vessel ID. We also explored using flags as factor but had issues with the hessian matrix being unstable. Both of these options should be explored in future analyses. For now we elected to exclude vessel as a covariate from the final standardized models. Of note also, including vessel ID in the geostatistical analysis also brought about additional problems in terms of the subsampling, as it was challenging to come up with a scheme where vessel units were preserved (i.e. we want to retain the time series of individual vessels when subsampling) and also respect the knot allocation assumptions. As computer power improves it might not be necessary to subsample the dataset anymore (especially for the South Pacific), but there are some advantages in subsampling in that it helps distribute effort more evenly across space. If subsampling is performed in future analyses care should be taken when considering the optimal way to account for vessel in the subsampling scheme.

6.5 Comparison between traditional and geostatistical approaches

The South Pacific albacore assessment was the first WCPO tuna assessment utilising geostatistical indices of abundance within an assessment used to provide management advice (i.e. not just as a sensitivity analysis). As a result, a specific CPUE-related axis was included within the model grid. The two methods (traditional and geostatistical) resulted in standardized indices whose trends were sufficiently different (see Figure 62) that they had a slight impact on reference points (e.g. $SB_{recent}/SB_{F=0}$ and $SB_{latest}/SB_{F=0}$). For instance, geostatistical indices resulted in a median $SB_{recent}/SB_{F=0}$ about 5 points higher over structural uncertainty grid runs (see Section 7.5.2 in Tremblay-Boyer et al., 2018).

The preferred CPUE standardisation approach is that which results in standardized indices that most closely capture trends in true abundance, a metric we cannot have access to, by definition. The alternative strategy would be to build a simulation model that closely captures the features present in the catch-and-effort dynamics of the Pacific and test both CPUE methods to identify the one that performs best in the objective of representing 'true' abundance, as done by Campbell et al. (2017). We note this type of work can be quite resource intensive, and methods need to be designed carefully in order to enable a valid comparison between CPUE standardization approaches.

From a practical standpoint the geostatistical approach has many desirable features in addition to the technical fact that it improves estimates for spatial covariates. It can be applied at the scale of the assessment region, and the year-quarter effect is an interaction of the spatial surface, which means that there can be a year-quarter index for each region covering the time-span of the assessment. Because it is a single model, it is easier to change between region structures if that is a feature of the assessment that has to be investigated. The standard errors for the year-quarter indices per region can be estimated directly as part of the modelling procedure, so no additional step is required to produce canonical CVs. Finally the regional weights are also estimated as part of the modelling procedure, which means that no external analysis is required to produce them. On the other hand the models can be technically challenging to run and to diagnose, but continuous investments in the improvement of methods from year to year should smooth these initial issues.

It is unclear whether the difference in indices seen between the results of the traditional and geostatistical approaches are due to the estimation of the 5° cell effect, or the fact that the dataset was subsetted by region for the traditional approach which meant the estimated regional indices were allowed to vary completely independently from one another. It would not be possible to replicate the traditional approach at the South Pacific scale as a whole because of computational issues. However, to compare the two approaches we ran simple traditional models using the same knots as those used for the geostatistical approach and a cluster effect, with a delta-lognormal error structure. Comparing the values of the cell effects estimated with these two approaches (Figures 63 and 64) shows that the effects are somewhat similar, except for the positive component of the geostatistical approach which estimates slightly lower knot effects in region 3. This indicates that a major driver of the differences between the standardized indices under the two approach might be the subsetting of the data by region. A future avenue for investigation could be to increase the flexibility of the year-quarter-knot interactions for the geostatistical model to see if we can increase the regional differences in the standardized indices estimated by region. On the other hand, a well-mixed stock would be expected to have minimal differences in abundance trends between regions, and current knowledge about stock structure could also inform whether it is appropriate, or not, to allow indices to vary between regions without constraints.

7 Regional weights

The 'regional weights' (McKechnie et al., 2014) are empirical estimates of the proportion of the population in each region of the assessment area vulnerable to a specified fishing gear (typically longline). Without regional weights, models often have too much flexibility in where they assign recruits and move individuals among regions, leading to unrealistic exchange rates and population sizes among regions. Regional weights are therefore used to adjust the standardised mean-centered catch-per-unit-effort (CPUE; numbers of fish per hundred hooks) indices used in the assessment, which in conjunction with the assumption of shared catchability among these fisheries, and the effort deviation penalties, provide constraints on the relative abundance among regions over the time-period that the regional weights were calculated.

Regional weights have been utilised in all recent WCPO stock assessments for tuna, including for south Pacific albacore, and yellowfin and bigeye tuna (Harley et al., 2015; Tremblay-Boyer et al., 2017a; McKechnie et al., 2017a). The use of geostatistical models to develop standardized indices of abundance allows the estimation of regional weights directly as part of the modelling procedure, but for consistency with previous assessments we also used 'traditional' regional weights as developed for the 2015 assessment by (Tremblay-Boyer et al., 2015a). The traditional approach can be briefly summarised as: aggregate-level relative abundance (CPUE) data are collated over a period of time where spatial differences in catchability (due to targeting etc.) are thought to be minimised (here we use 1975-2016) and spatial coverage of data is adequate; spatial surfaces of relative abundance are estimated for the entire assessment region in each year-quarter over a defined period using generalised additive models (GAMs), and may include catchability covariates; the predictions of this spatial surface are used to sum abundance over all cells in a region (including imputed values for cells with missing data) and the region-specific values are scaled to sum to 1; and finally, the regional weights used in the stock assessment are calculated as the region-specific means over the time period 1975-2016, chosen to exclude the 1960s-early 1970s where effort was sparse across most of the assessment region.

For the geostatistical approach, as mentioned above, the same model was used as that to standardize the CPUE, except that the cell-effect was not removed when predicting time-series of biomass by knot by region. Otherwise, the same steps as for the 'traditional' approach were performed, that is, the regional weight for each region is the area-weighted sum of all the predicted cell abundances, scaled across assessment regions to sum to 1 across the period 1975-2016, during which the data had better coverage. Only cells with a SST of 16° C or more were included as part of the abundance predictions for the southernmost regions, as the southmost latitudes had very little effort (especially in region 5) yet were predicted to have very high abundance based on the high observed CPUE right up ∼42° S in region 5 (see Figure 46). There were no observations to confirm whether those predicted abundances were indeed realistic, but it was assumed that the lack of fishing effort in that part of the assessment region throughout the time-span of the assessment likely indicated that local catch rates were in fact low.

The geostatistical approach for the regional weights has the advantage that it does not require the additional step of standardizing the CPUEs, mean-centering them, and then rescaling them according to regional weights from a separate analysis. For the current assessment we ended up using the two-step approach but future iterations should use the scaling directly produced via the geostatiscal approach.

Three configurations of CPUE indices and regional weights were utilised in the 2018 stock assess-

ment of south Pacific albacore, with regional weights compared in Figure 65. The early steps in the progression from the 2015 assessment to the 2018 diagnostic case model (Tremblay-Boyer et al., 2018) aimed to replicate the methods used in 2015 and so CPUE analyses were calculated using the 'traditional' method outlined in Section 6. Consequently, the 2015 regional weights were used for those early steps, and the first step with the 2018 regional structure, which also had traditional CPUEs, used the 2015 regional weights but combined to account for the merged region structure (i.e. 2018 region 1 = 2015 region 1 + region 4, etc.)(see Figures 14 and 15 in Tremblay-Boyer et al., 2018). The next step switched to geostatistical CPUEs, which were scaled with the geostatistical regional weights (though we also explored intermediate steps combining geostatistical CPUEs with 'traditional' regional weights and found the influence of this new approach was minimal). All subsequent model runs (including one-off sensitivities) with geostatistical CPUE used the geostatistical regional weights. Finally, all runs within the structural uncertainty grid, including those using the traditional CPUE approach, used geostatistical regional weights to ensure that any differences in outputs could be attributed to temporal patterns in CPUE.

8 Tagging data inputs

Tagging data for the 2018 stock assessment consisted of tag releases and returns from a South Pacific Albacore Research Group tagging programme in the mid-1980s and SPC albacore tagging programmes conducted during the austral summers of 1990–1992 and 2009–2010 (Figure 12; from Hoyle et al., 2012). Albacore were captured primarily by trolling and tagged using standard tuna tagging equipment and techniques by trained scientists and scientific observers. During the 1980s and 1990s, the majority of tag releases were made by scientific observers on-board New Zealand and US troll vessels fishing in New Zealand waters and in the central South Pacific subtropical convergence zone region. The more recent tagging was conducted in New Zealand waters.

The construction of the tagging files (hereafter .tag files) for use in MFCL for the 2018 assessment follows very closely the recent methodology used for skipjack (McKechnie et al., 2016), and bigeye and yellowfin tuna (McKechnie et al., 2017b). The raw .tag data files constructed from SPC databases underestimate recapture rates of tagged fish due to: not all recaptured tags being usable in MFCL (e.g. tags without recapture dates, location or vessel identification often cannot be attributed to a model fishery and are therefore excluded from the stock assessment); some fish shed tags immediately, or some time after tagging; and some tagging-induced mortality occurs such that tagged fish assumed to be present in the population will have already died. This results in fishing mortality being underestimated by this data component, which will potentially lead to underestimates of biomass.

The methodology of correcting releases and filtering of the data for input into MFCL is covered in detail by McKechnie et al. (2017b), and so we provide only the following brief overview of the process and refer readers to that study for relevant details, especially the formulae used to correct release numbers which are presented in Section 3.4 of that report. The concise details of the methodology used to construct the .tag file are:

- The "raw" .tag file for MFCL is produced using software written in FoxPro (Long, 1994) and only includes usable tags (those that can be assigned to a fishery and time period).
- A parallel programme written in the R statistical language (R Core Team, 2018) uses identical SQL queries with the exception that it extracts all recaptures (usable and unusable) and

produces a "full" .tag file.

- The usability ratio is calculated as the ratio of usable to total recaptures at the scale of the length bin within a tagging release event.
- The number of releases is then scaled down by these release-group- and length bin-specific correction ratios.

We also further adjusted the tag release numbers downwards by a factor of 0.5 to account for likely mortality of albacore due to the stress of capture, handling and tagging. Initial tagging mortality operates in a similar fashion to non-reporting and was therefore dealt with by adjusting the releases. The choice of 0.5 as the adjustment factor was somewhat arbitrary, but was influenced by recent work on tagger effects on tagged fish survival and associated correction factors in what are thought to be more robust tropical tunas (Berger et al., 2014). In that study it was found that the median correction factors for tagger effects were 0.68–0.76 for the tropical tunas. Given that albacore are believed to be more sensitive to capture and handling, it was felt that a stronger correction was appropriate in this case. The 2015 stock assessment also tested a factor of 0.7 and found that this resulted in relatively minor impacts on model results.

In contrast to the tropical tuna stocks in the WCPO, no information on tag reporting rates in the form of tag seeding trials were available for south Pacific albacore. This prevented the construction of informative tag reporting rate priors, and so uninformative, diffuse priors with a mean of 0.1 were used for all reporting rate groupings.

Note that there were no additional tag releases subsequent to the 2015 assessment although an additional **7** recaptures were included in the 2018 data. These relate to fish tagged around New Zealand in 2009 that have been recaptured in longline fisheries after the terminal year of the 2015 assessment (which was 2013).

9 Maturity-at-length

The maturity function is routinely input to stock assessment models as a fixed vector of age-specific values as there is very little information to inform the model of its shape. A new development in MULTIFAN-CL this year allows to input length-based spawning potential, which is then converted internally within the model to an age-based vector dependent upon the growth.

We specified spawning potential-at-length as the product of sex ratio and maturity-at-length, rescaled to 1. Sex ratio at length was obtained by fitting a spline to observed sex-ratio from SPC-held longline observers' data, stratified by flag and 10° cell and weighted by longline catch to account for uneven observer coverage amongst fleets in space (Figure 66). Longline observers' data only covered lengths from 70 to 110 cm so we extrapolated to cover the stock assessment range by (1) setting sex ratio for lengths < 70cm at the 70cm value and (2) assuming sex ratio for females declined linearly from the value observed at 110cm to 0 at 130cm (Figure 67, blue line in left panel). The maturity proportion at length was obtained from the weighted maturity ogives presented in (Farley et al., 2014) (see Figure 10 therein), smoothed *via* a logistic curve (Figure 67, red line in left panel). The final input to MULTIFAN-CL was a product of those two ogives, rescaled so that the maximum was 1 (Figure 67, right panel).

References

Berger, A. M., McKechnie, S., Abascal, F., Kumasi, B., Usu, T., and Nicol, S. J. (2014). Analysis of tagging data for the 2014 tropical tuna assessments: data quality rules, tagger effects, and reporting rates. WCPFC-SC10-2014/SA-IP-06, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Bigelow, K. A. and Hoyle, S. D. (2012). Standardized CPUE for South Pacific albacore. WCPFC-SC8-2012/SA-IP-14, Busan, Republic of Korea, 7–15 August 2012.

Campbell, R., Zhou, S., Hoyle, S., Hillary, R., Haddon, M., and Auld, S. (2017). Developing innovative approaches to improve CPUE standardisation for Australia's multispecies pelagic longline fisheries.

Farley, J. H., Hoyle, S. D., Eveson, J. P., Williams, A. J., Davies, C. R., and Nicol, S. J. (2014). Maturity ogives for south pacific albacore tuna (*Thunnus alalunga*) that account for spatial and seasonal variation in the distributions of mature and immature fish. *PLoS ONE*, 9(1).

Harley, S. J., Davies, N., Tremblay-Boyer, L., Hampton, J., and McKechnie, S. (2015). Stock assessment of south Pacific albacore tuna. WCPFC-SC11-2015/SA-WP-06, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

He, X., Bigelow, K. A., and Boggs, C. H. (1997). Cluster analysis of longline sets and fishing strategies within the hawaii-based fishery. *Fisheries Research*, 31:144–158.

Hoyle, S., Hampton, J., and Davies, N. (2012). Stock assessment of albacore tuna in the South Pacific Ocean. WCPFC-SC8-2012/SA-WP-04, Busan, Republic of Korea, 7–15 August 2012.

Hoyle, S. D., D. Kim, D., Lee, S., Matsumoto, T., Satoh, K., , and Yeh, Y.-M. (2016). Collaborative study of tropical tuna cpue from multiple Indian ocean longline fleets in 2016. Technical Report IOTC–2016–WPM07–11, 7th Working Party on Methods, Victoria, Seychelles.

Hoyle, S. D., Langley, A. D., and Campbell, R. A. (2014). Recommended approaches for standardizing CPUE data from pelagic fisheries. WCPFC-SC10-2014/SA-IP-10, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5):1–21.

Long, J. (1994). *FoxPro 2.6 for Windows: developer's guide (2nd ed)*. Sams Pub, Carmel, Ind.

McKechnie, S. (2014). Analysis of longline size frequency data for bigeye and yellowfin tunas in the WCPO. WCPFC-SC10-2014/SA-IP-04, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Harley, S. J., Davies, N., Rice, J., Hampton, J., and Berger, A. (2014). Basis for regional structures used in the 2014 tropical tuna assessments, including regional weights. WCPFC-SC10-2014/SA-IP-02, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

McKechnie, S., Ochi, D., Kiyofuji, H., Peatman, T., and Caillot, S. (2016). Construction of tagging data input files for the 2016 skipjack tuna stock assessment in the western and central Pacific Ocean. WCPFC-SC12-2016/SA-IP-05, Bali, Indonesia, 3–11 August 2016.

McKechnie, S., Pilling, G., and Hampton, J. (2017a). Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-WP-05, Rarotonga, Cook Islands, 9–17 August 2017.

McKechnie, S., Tremblay-Boyer, L., and Harley, S. J. (2015). Analysis of Pacific-wide operational longline CPUE data for bigeye tuna. WCPFC-SC11-2015/SA-WP-03, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

McKechnie, S., Tremblay-Boyer, L., and Pilling, G. (2017b). Background analyses for the 2017 stock assessments of bigeye and yellowfin tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-IP-06, Rarotonga, Cook Islands, 9–17 August 2017.

OFP (2015). Report of the workshop on operational longline data. Technical Report WCPFC-SC11-2015/SA-IP-01, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Petitgas, P. (2001). Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, 2(3):231–249.

Pilling, G. and Brouwer, S. (2017). Report from the SPC pre-assessment workshop, Noumea, April 2017. Technical Report WCPFC-SC13-2017/SA-IP-02, Rarotonga, Cook Islands, 9–17 August 2017.

Pilling, G. and Brouwer, S. (2018). Report from the SPC pre-assessment workshop, Noumea, April 2018. Technical Report WCPFC-SC14-2018/SA-IP-01, Busan, Korea, 8–16 August 2018.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, J., Harley, S., Davies, N., and Hampton, J. (2014). Stock assessment of skipjack tuna in the Western and Central Pacific Ocean. WCPFC-SC10-2014/SA-WP-05, Majuro, Republic of the Marshall Islands, 6–14 August 2014.

Rue, H., Riebler, A. I., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Reviews of Statistics and Its Applications*, 4(March):395–421.

Schnute, J. T., Boers, N., and Haigh, R. (2017). *PBSmapping: Mapping Fisheries Data and Spatial Analysis Tools*. R package version 2.70.4.

Scott, R. D. and McKechnie, S. (2015). Analysis of longline length frequency compositions for South Pacific albacore tuna. WCPFC-SC11-2015/SA-IP-07, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(11):1655–1666.

Thorson, J., Pinksy, M., and Ward, E. (2016). Model-based inference for estimating shifts in species distribution, area occupied, and center of gravity. *Methods Ecol. Evol.*, 7(8):990–1008.

Thorson, J., Shelton, A., Ward, E., and Skaug, H. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for west coast groundfishes. *ICES J. Mar. Sci. J. Cons*, 72(5):1297–1310.

Tremblay-Boyer, L., Hampton, J., McKechnie, S., and Pilling, G. (2018). Stock assessment of South Pacific albacore tuna in the WCPO. Technical Report WCPFC-SC14-2018/SA-WP-05.

Tremblay-Boyer, L., McKechnie, S., and Harley, S. J. (2015a). Spatial and fisheries structure and regional weights for the 2015 south Pacific albacore tuna (*Thunnus alalunga*) assessment. WCPFC-SC11-2015/SA-IP-07, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Tremblay-Boyer, L., McKechnie, S., and Harley, S. J. (2015b). Standardized CPUE for south Pacific albacore tuna (*Thunnus alalunga*) from operational longline data. WCPFC-SC11-2015/SA-IP-03, Pohnpei, Federated States of Micronesia, 5–13 August 2015.

Tremblay-Boyer, L., McKechnie, S., Pilling, G., and Hampton, J. (2017a). Stock assessment of yellowfin tuna in the Western and Central Pacific Ocean. WCPFC-SC13-2017/SA-WP-06, Rarotonga, Cook Islands, 9–17 August 2017.

Tremblay-Boyer, L., McKechnie, S., Pilling, G. M., and Hampton, J. (2017b). Geostatistical analyses of operational longline CPUE data. WCPFC-SC13-2017/SA-WP-03, Rarotonga, Cook Islands, 9–17 August 2017.

10 Tables

Table 1: Definition of fisheries for the MULTIFAN-CL South Pacific albacore tuna stock assessment.

Fishery	Nationality	Gear	2018 regions	2015 fleet label
1. DWFN LL 1	Distant-water Fishing Nations	Longline	1	LL ALL 1 and 4
2. PICT.AZ LL 1	Pacific Island Countries and Territories	Longline	1	LL ALL 1 and 4
3. DWFN LL 2	Distant-water Fishing Nations	Longline	2	LL ALL 2 and 5
4. PICT LL 2	Pacific Island Countries and Territories	Longline	2	LL ALL 2 and 5
5. AZ LL 2	Australia/New Zealand	Longline	2	LL ALL 2 and 5
6. DWFN LL 3	Distant-water Fishing Nations	Longline	3	LL ALL 3 and 6
7. PICT LL 3	Pacific Island Countries and Territories	Longline	3	LL ALL 3 and 6
8. AZ LL 3	Australia/New Zealand	Longline	3	LL ALL 3 and 6
9. DWFN LL 4	Distant-water Fishing Nations	Longline	4	LL ALL 7
10. PICT.AZ LL 4	Pacific Island Countries and Territories	Longline	4	LL ALL 7
11. DWFN LL 5	Distant-water Fishing Nations	Longline	5	LL ALL 8
12. PICT.AZ LL 5	Pacific Island Countries and Territories	Longline	5	LL ALL 8
13. All TR 3	All nationalities	Troll	3	TR ALL 3 and 6
14. All TR 5	All nationalities	Troll	5	TR ALL 8
15. All DN 3	All nationalities	Driftnet	3	DN ALL 3 and 6
16. All DN 5	All nationalities	Driftnet	5	DN ALL 8
17. Index LL 1	Index fishery	Longline	1	–
18. Index LL 2	Index fishery	Longline	2	–
19. Index LL 3	Index fishery	Longline	3	–
20. Index LL 4	Index fishery	Longline	4	–
21. Index LL 5	Index fishery	Longline	5	–

Table 2: Summary and coverage of potential explanatory variables available in the dataset

Variable	% records	Retained in GLM	Note
Date of set	100	No	
Year	100	Yes	Used in clustering
Quarter	100	Yes	
Month	100	No	Used in clustering
Geographic coordinate (1° or 5°)	100	Yes	
Vessel identification	93	Yes	Used in clustering
Hooks-between-floats, HBF	65	No	
Set start time	74	No	
Effort in hooks	100	Yes	

11 Figures

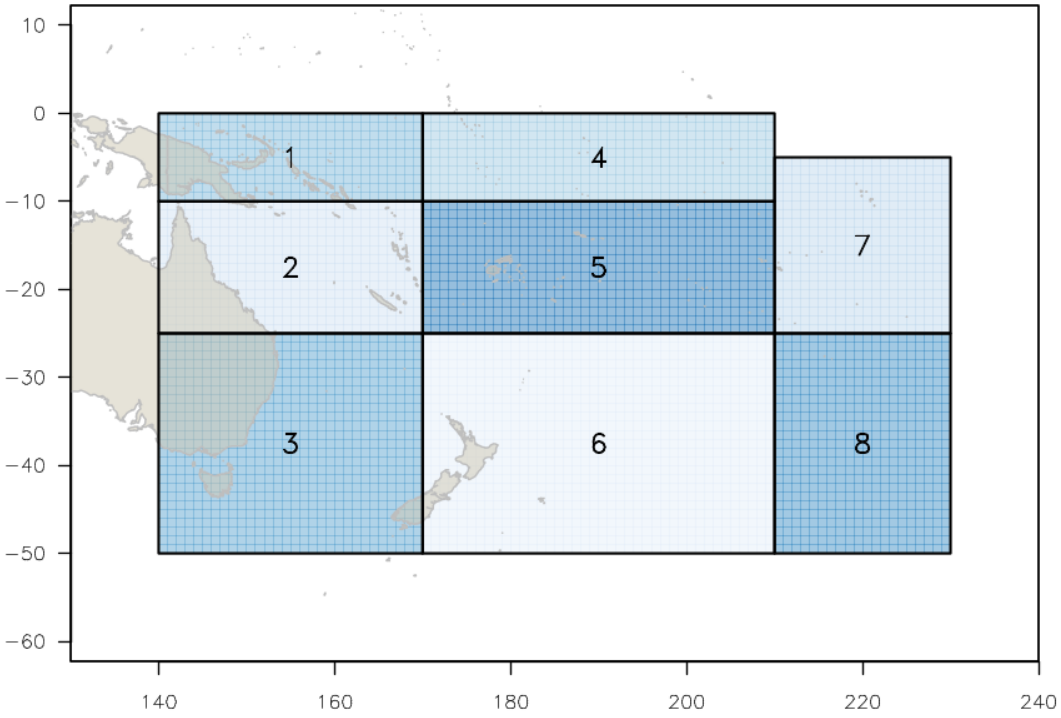

Figure 1: Map of the assessment regional structures used in the 2015 stock assessment ('2015 regional structure').

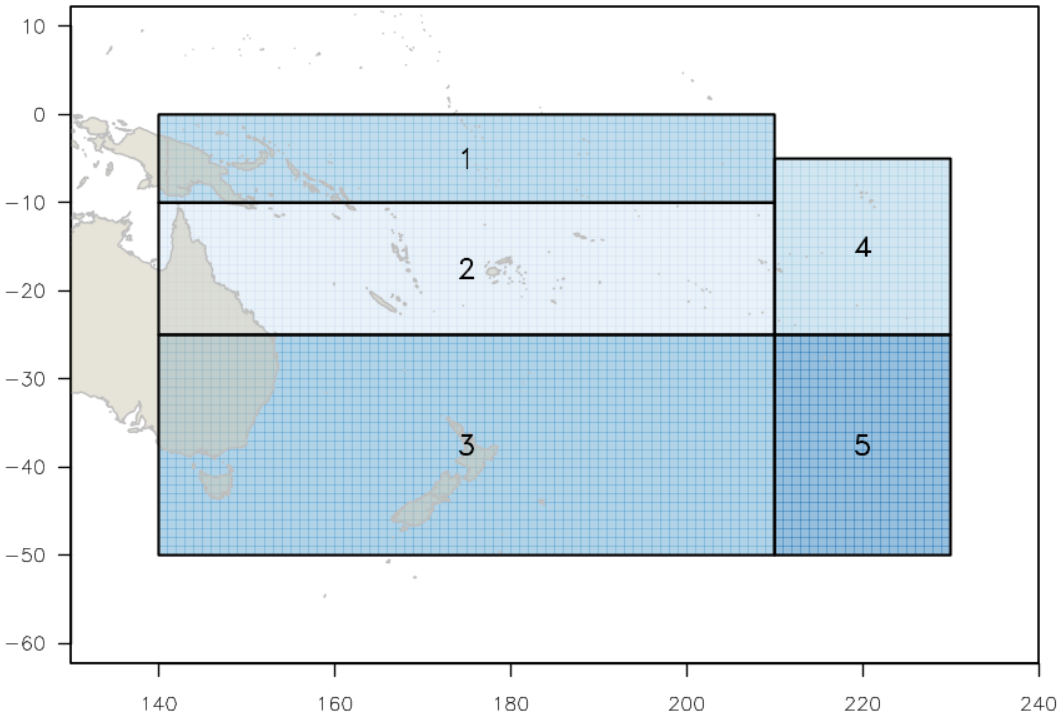

Figure 2: Map of the assessment regional structures used in the 2018 stock assessment ('2018 regional structure').

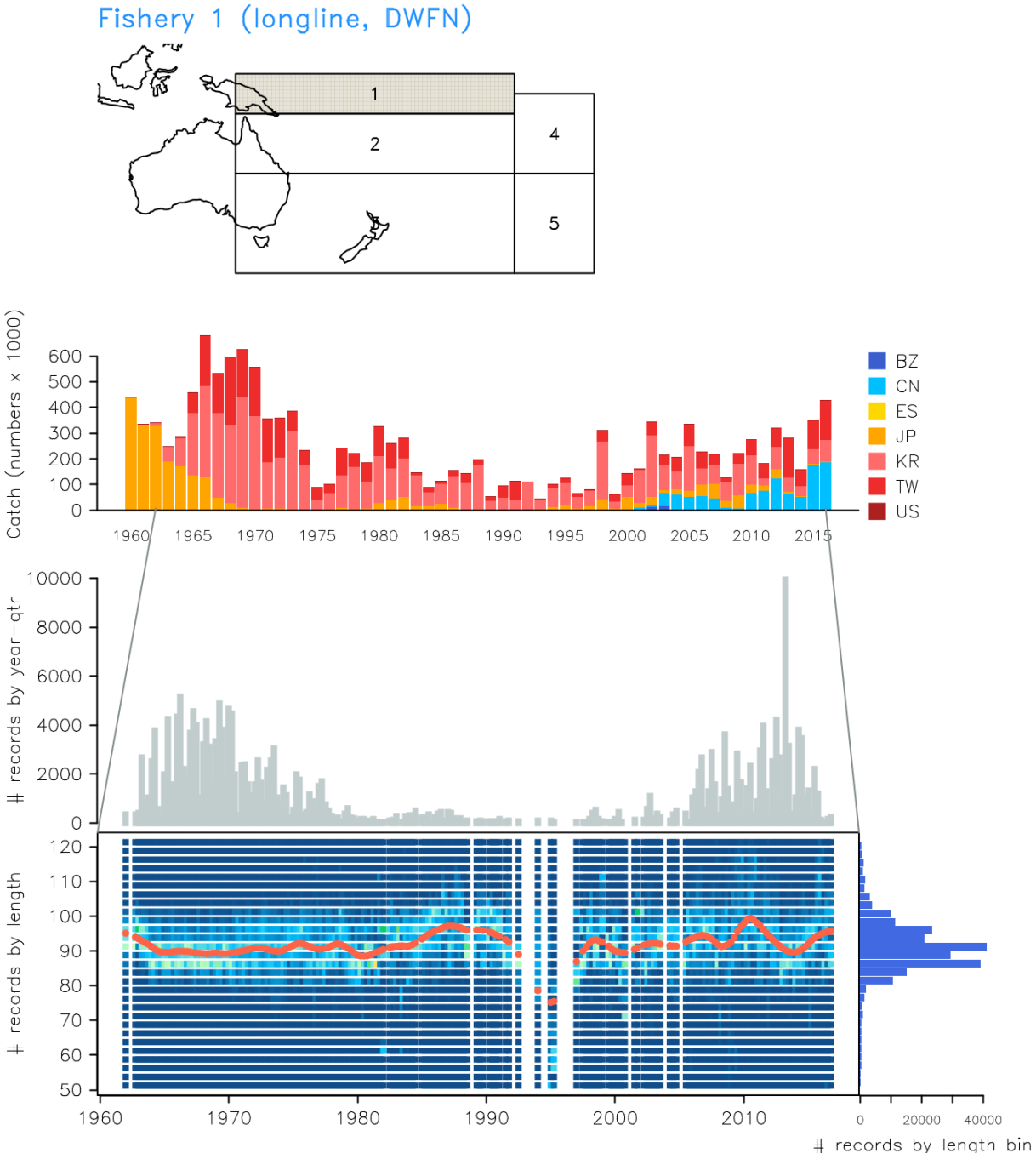

Figure 3: Summary of raw data available for fishery 1 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

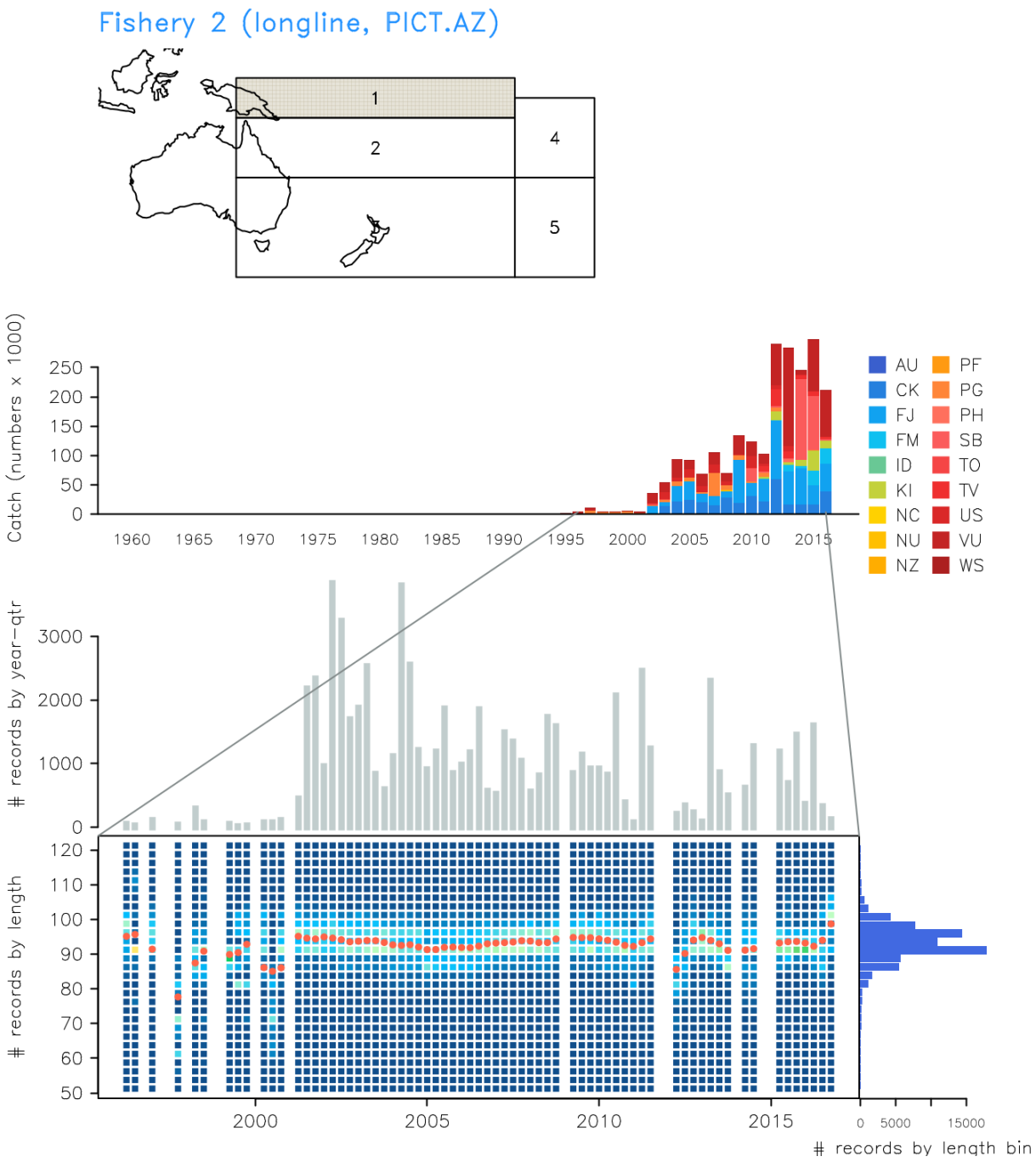

Figure 4: Summary of raw data available for fishery 2 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

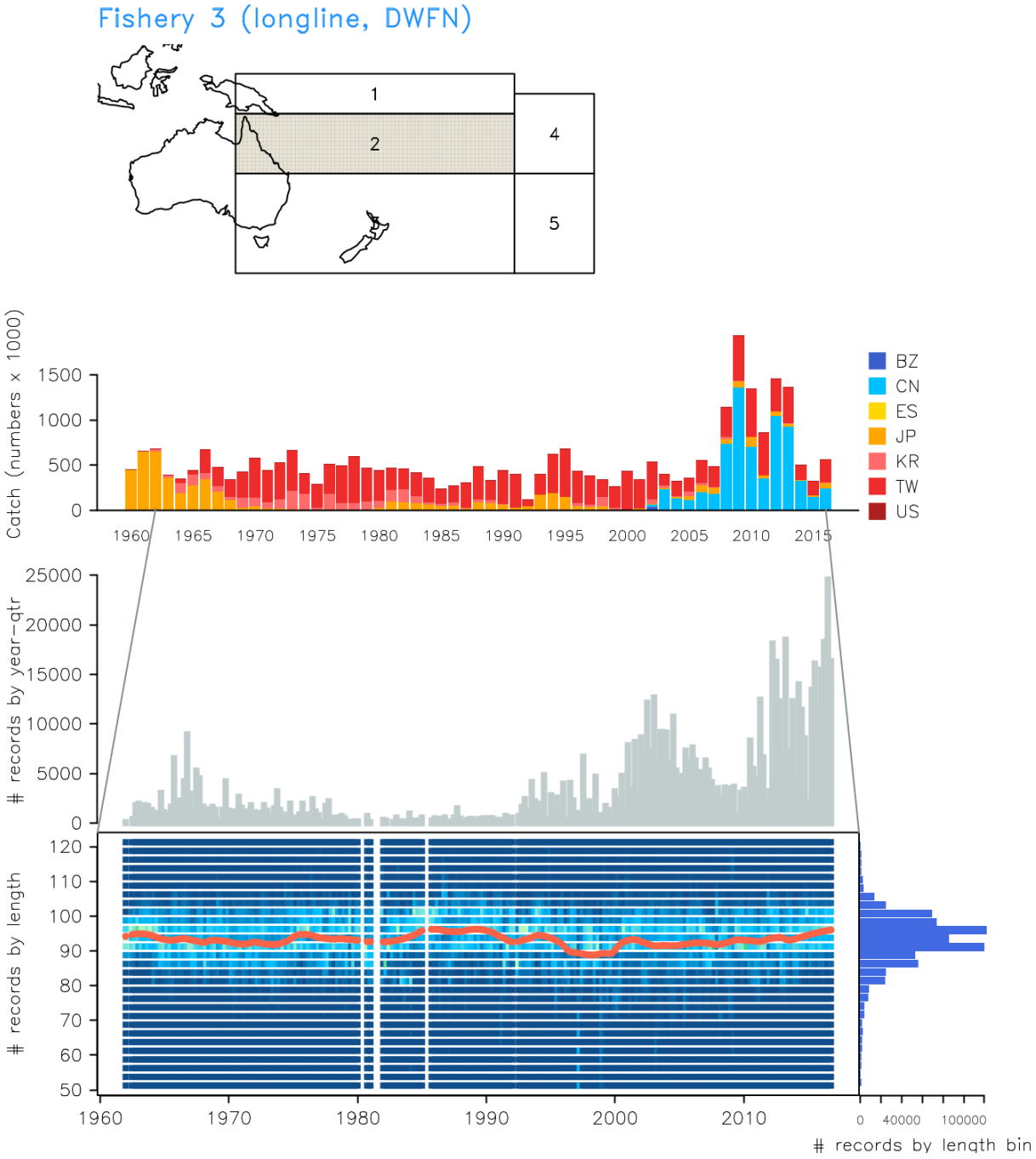

Figure 5: Summary of raw data available for fishery 3 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

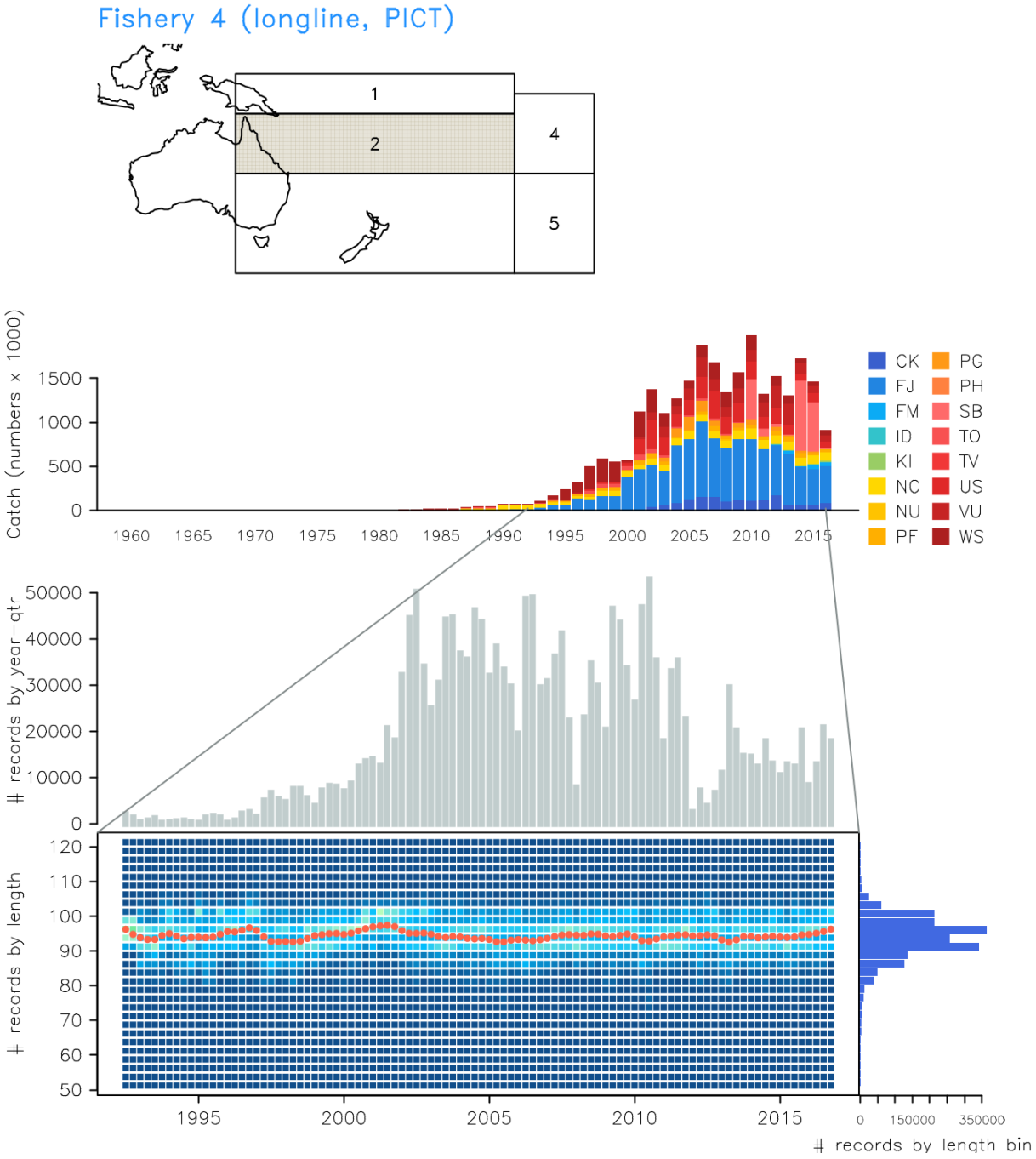

Figure 6: Summary of raw data available for fishery 4 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

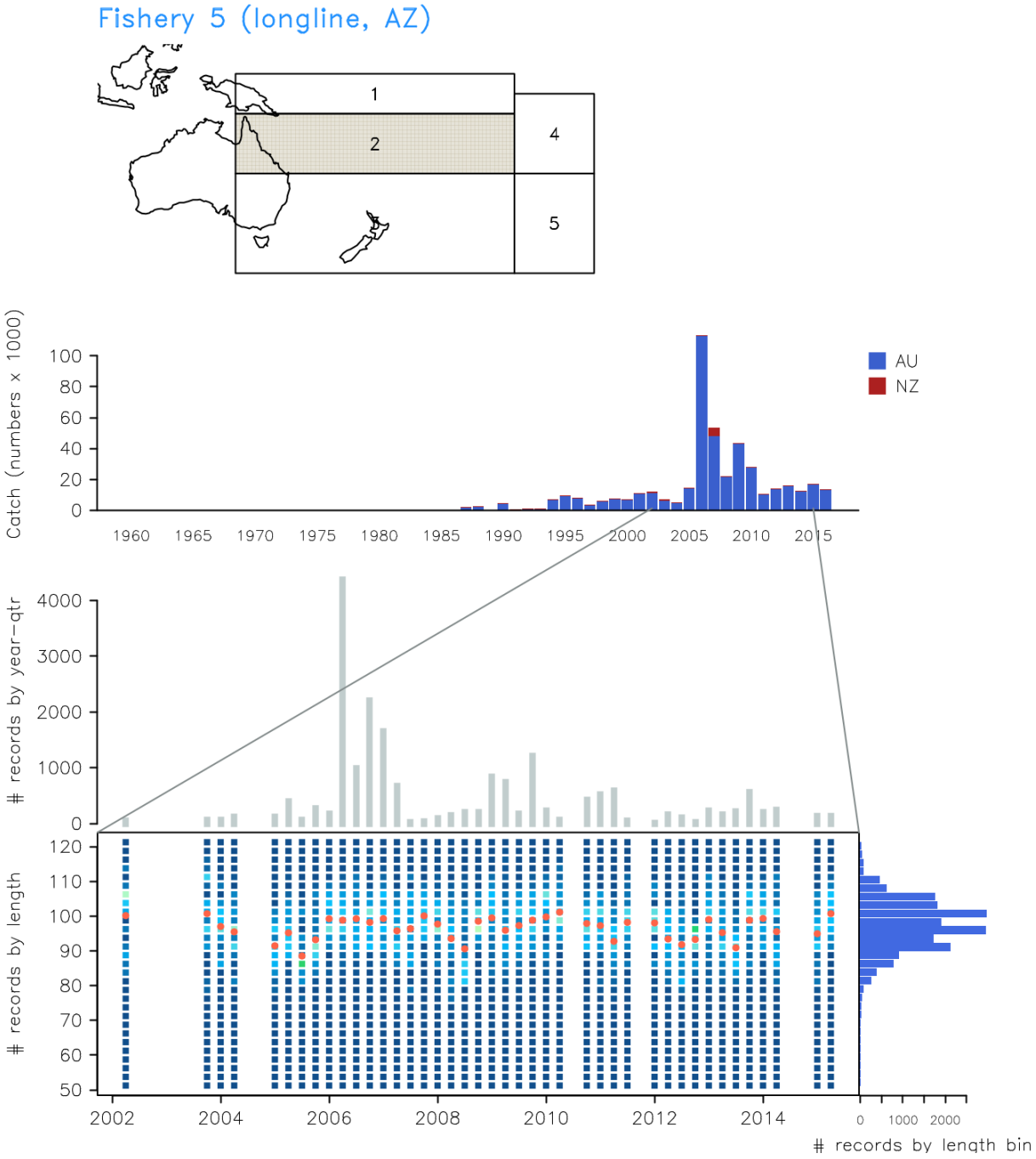

Figure 7: Summary of raw data available for fishery 5 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

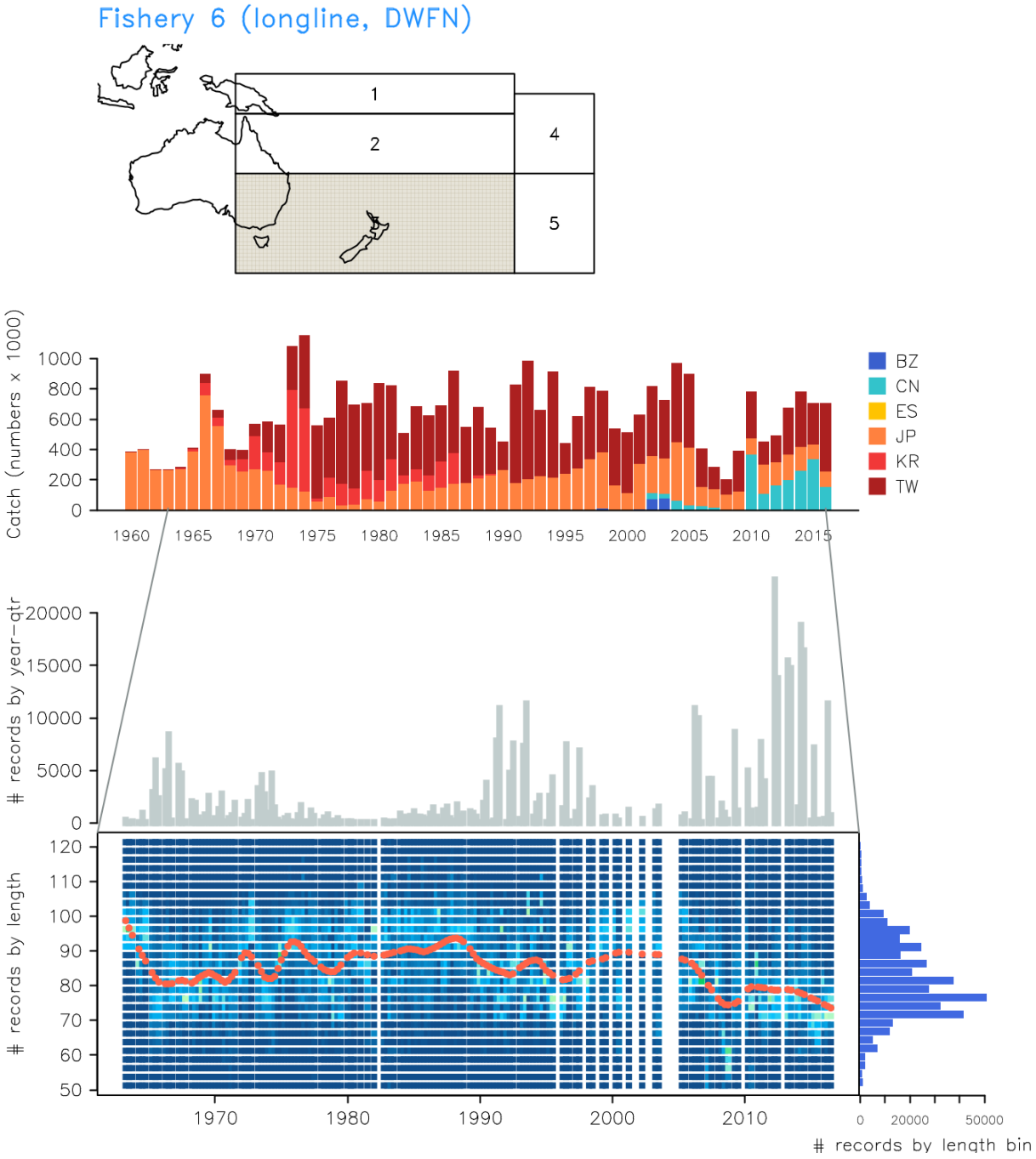

Figure 8: Summary of raw data available for fishery 6 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

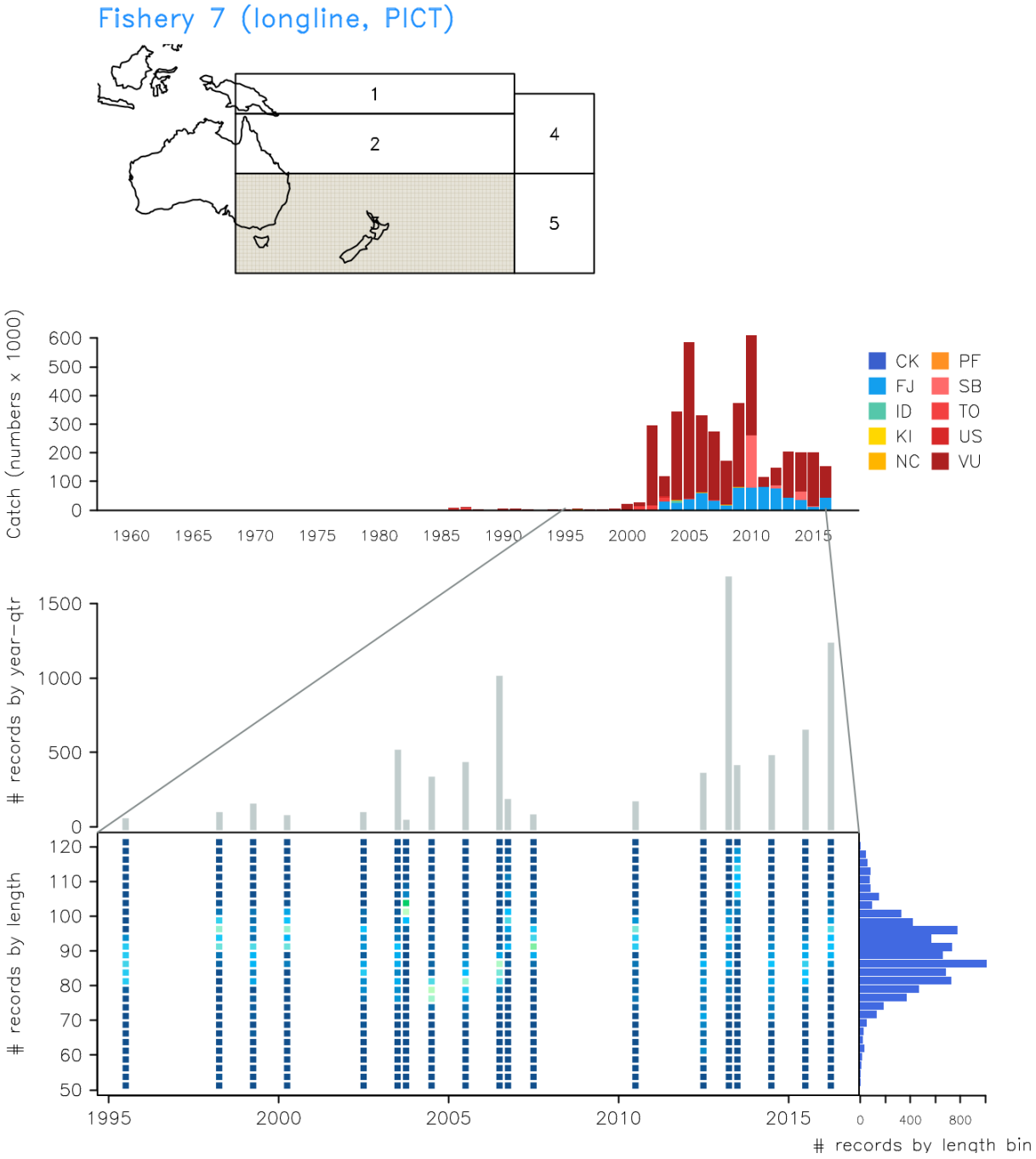

Figure 9: Summary of raw data available for fishery 7 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

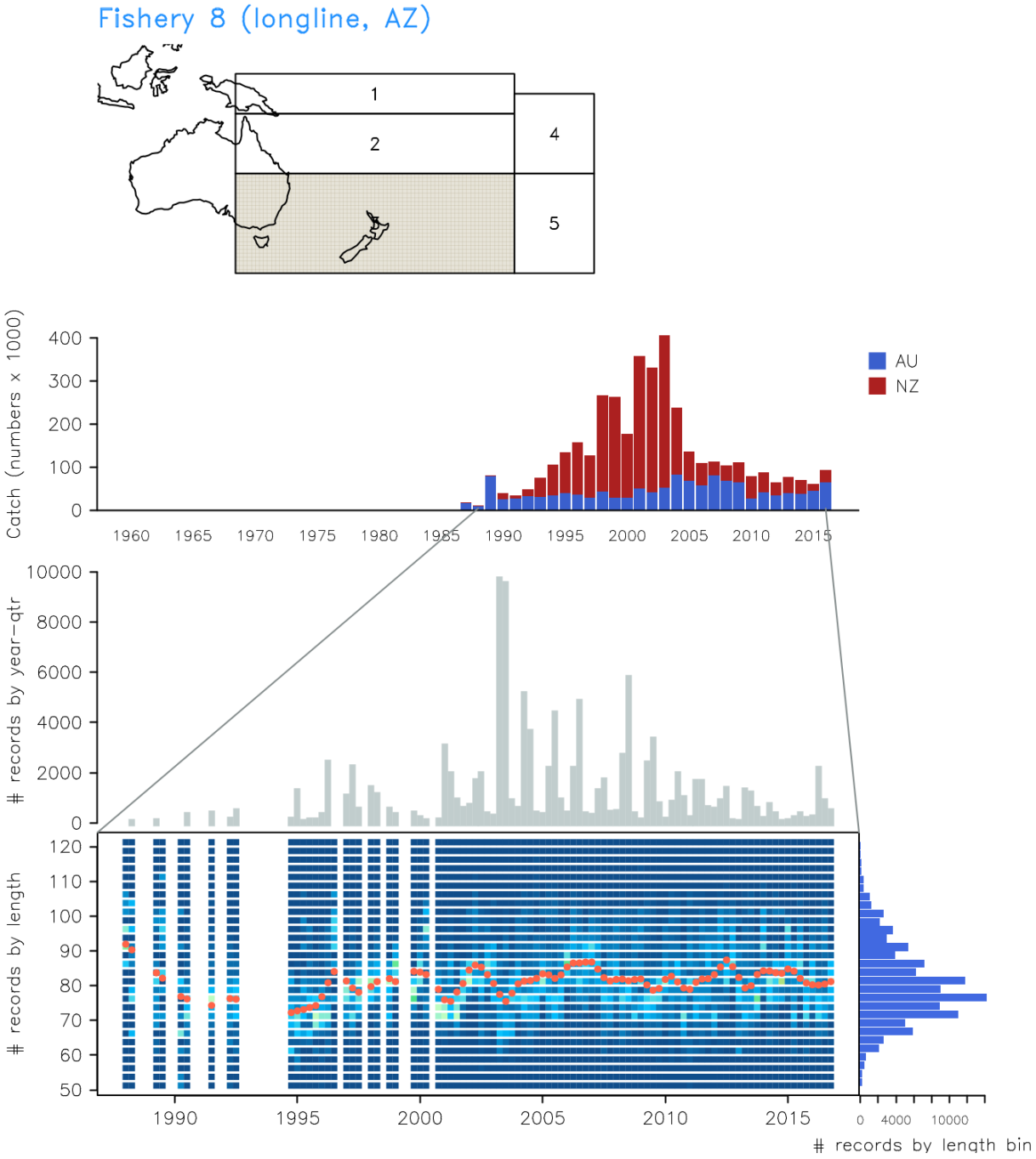

Figure 10: Summary of raw data available for fishery 8 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

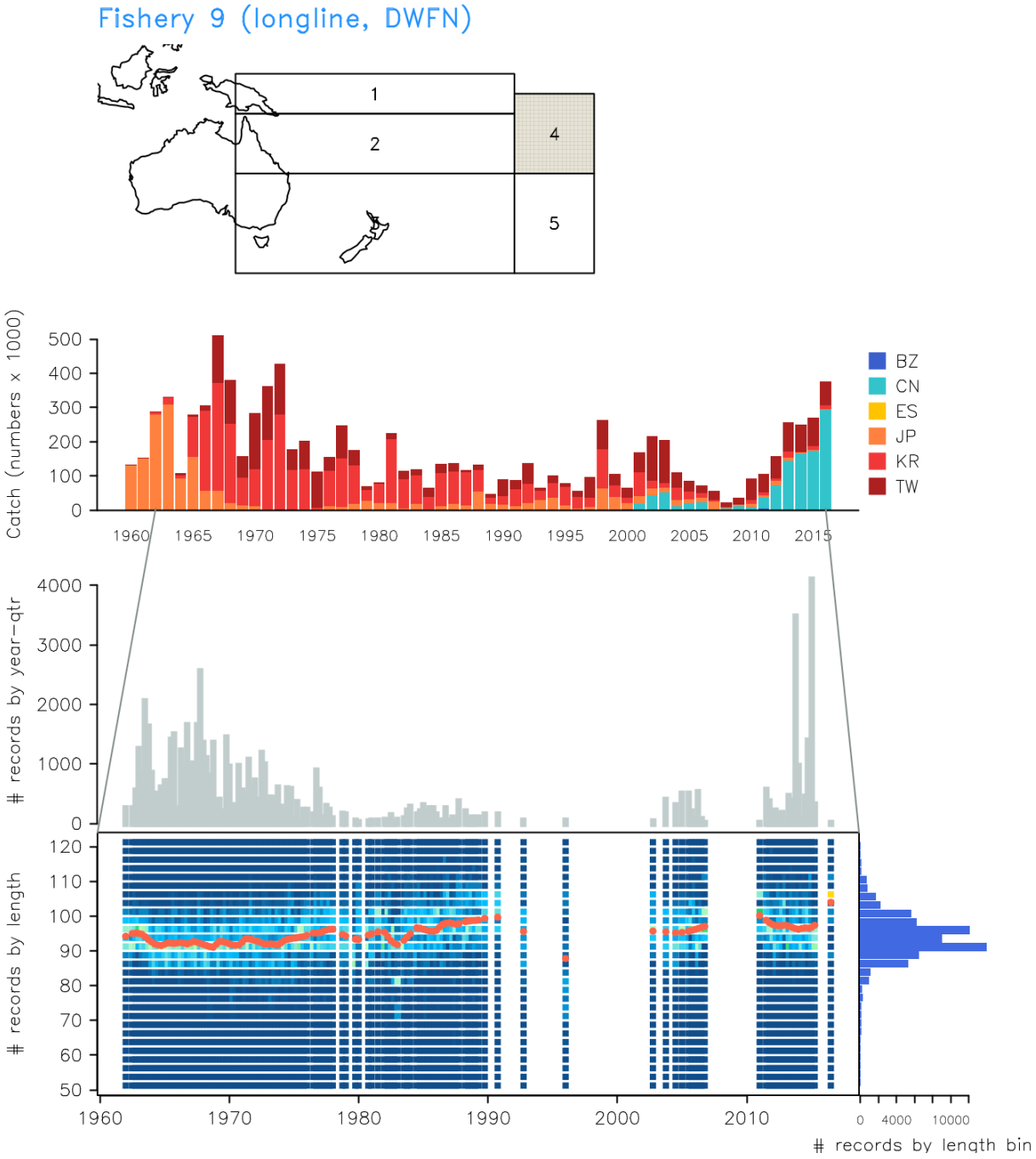

Figure 11: Summary of raw data available for fishery 9 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

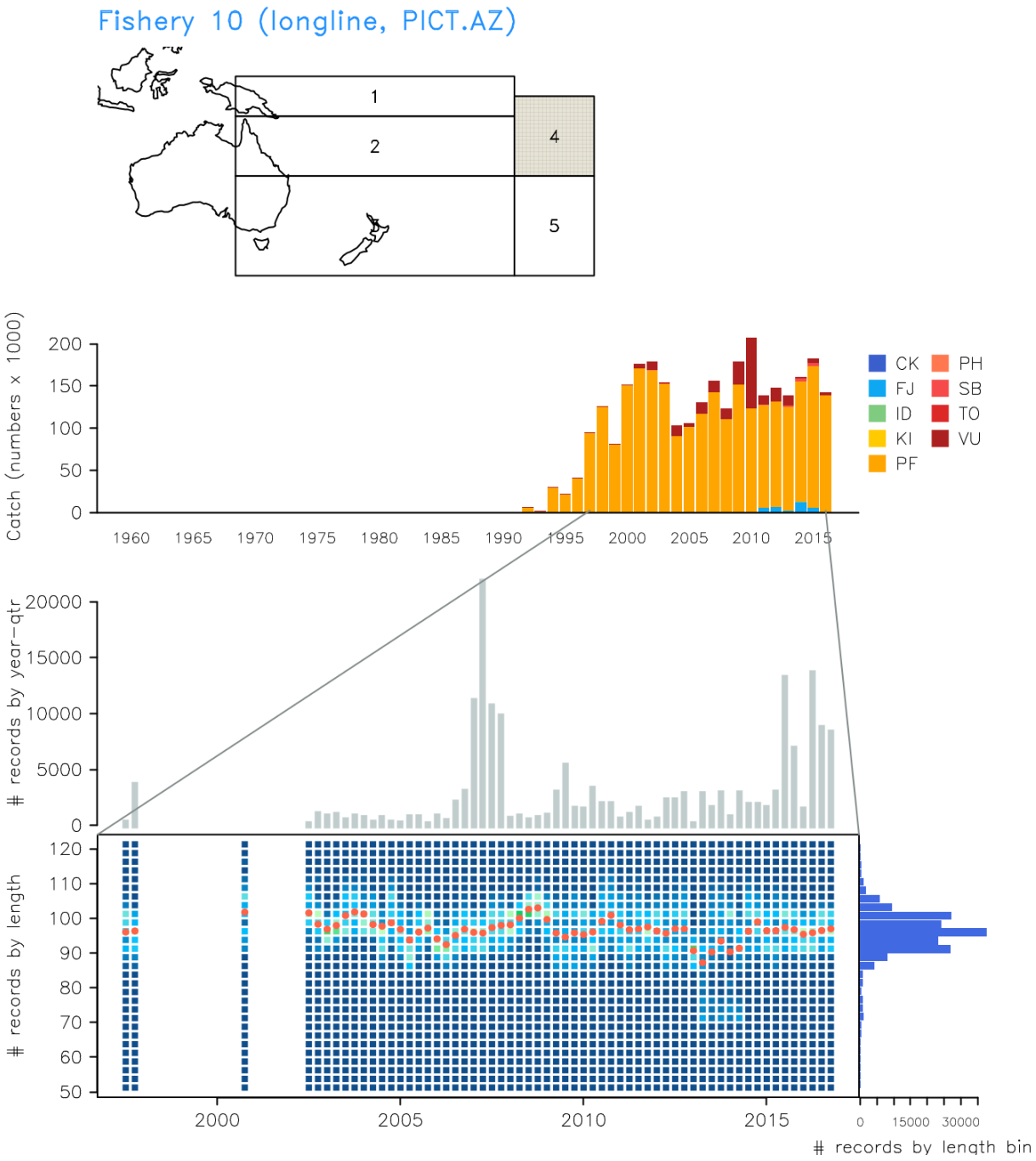

Figure 12: Summary of raw data available for fishery 10 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

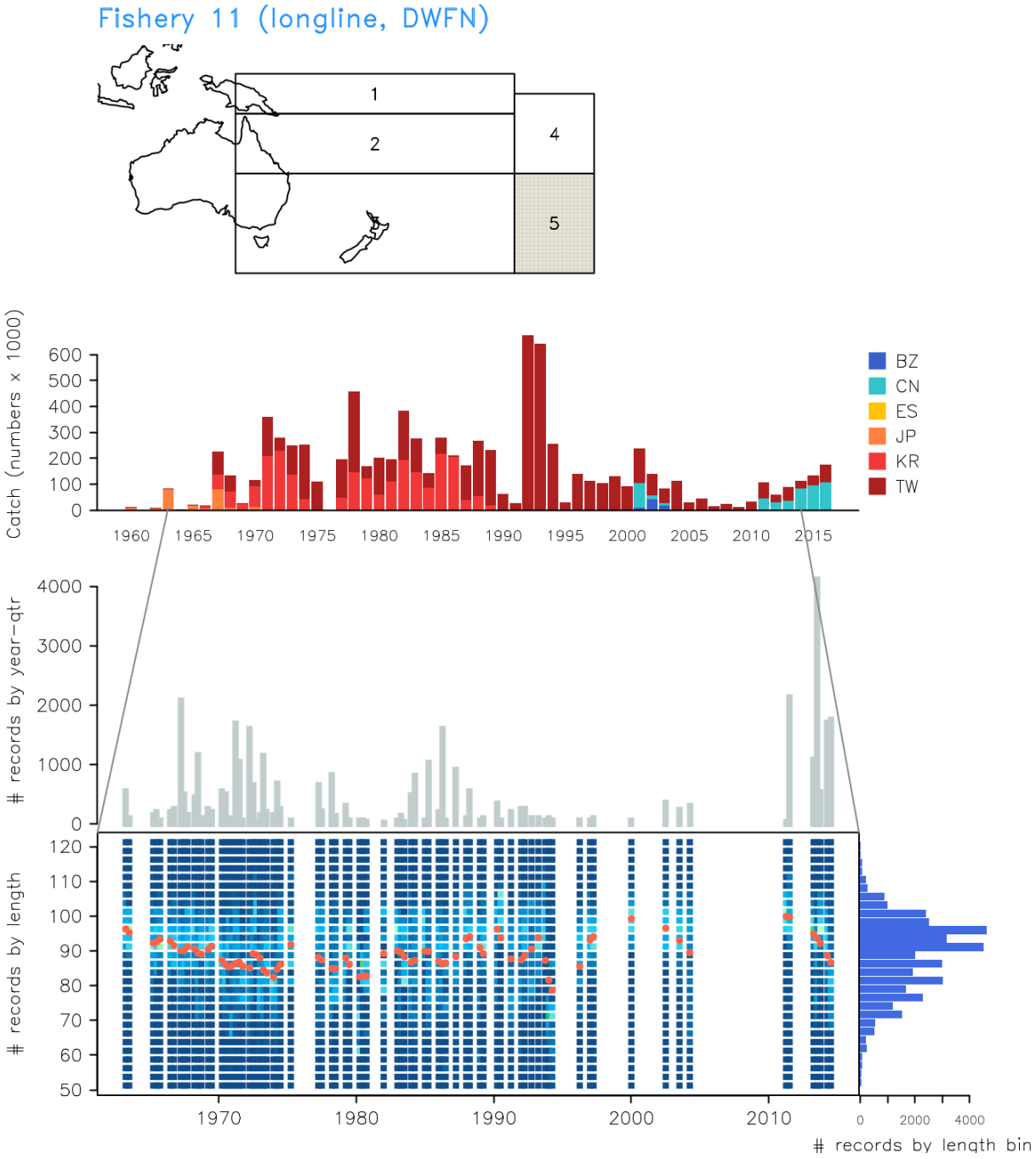

Figure 13: Summary of raw data available for fishery 11 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

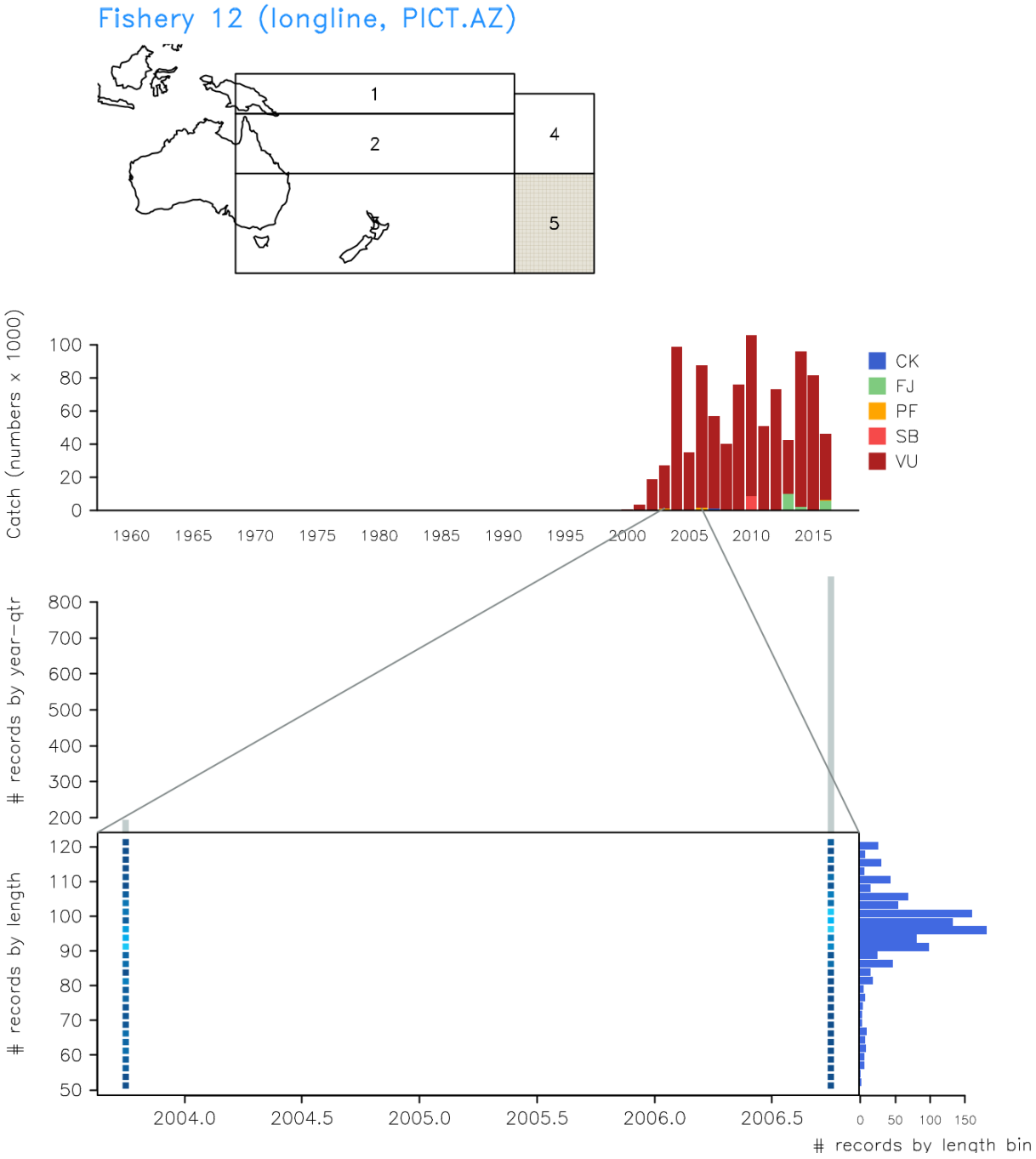

Figure 14: Summary of raw data available for fishery 12 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

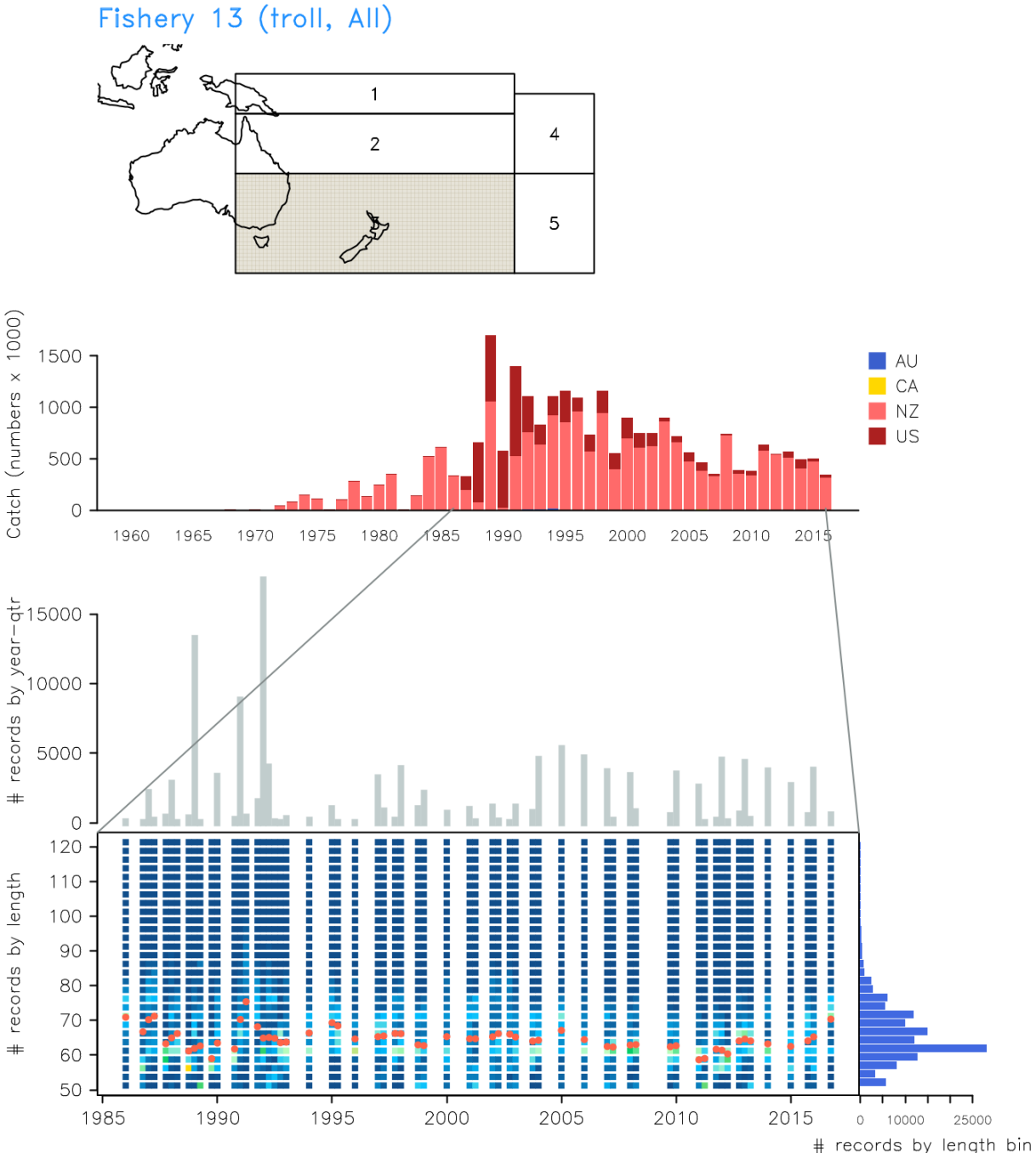

Figure 15: Summary of raw data available for fishery 13 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

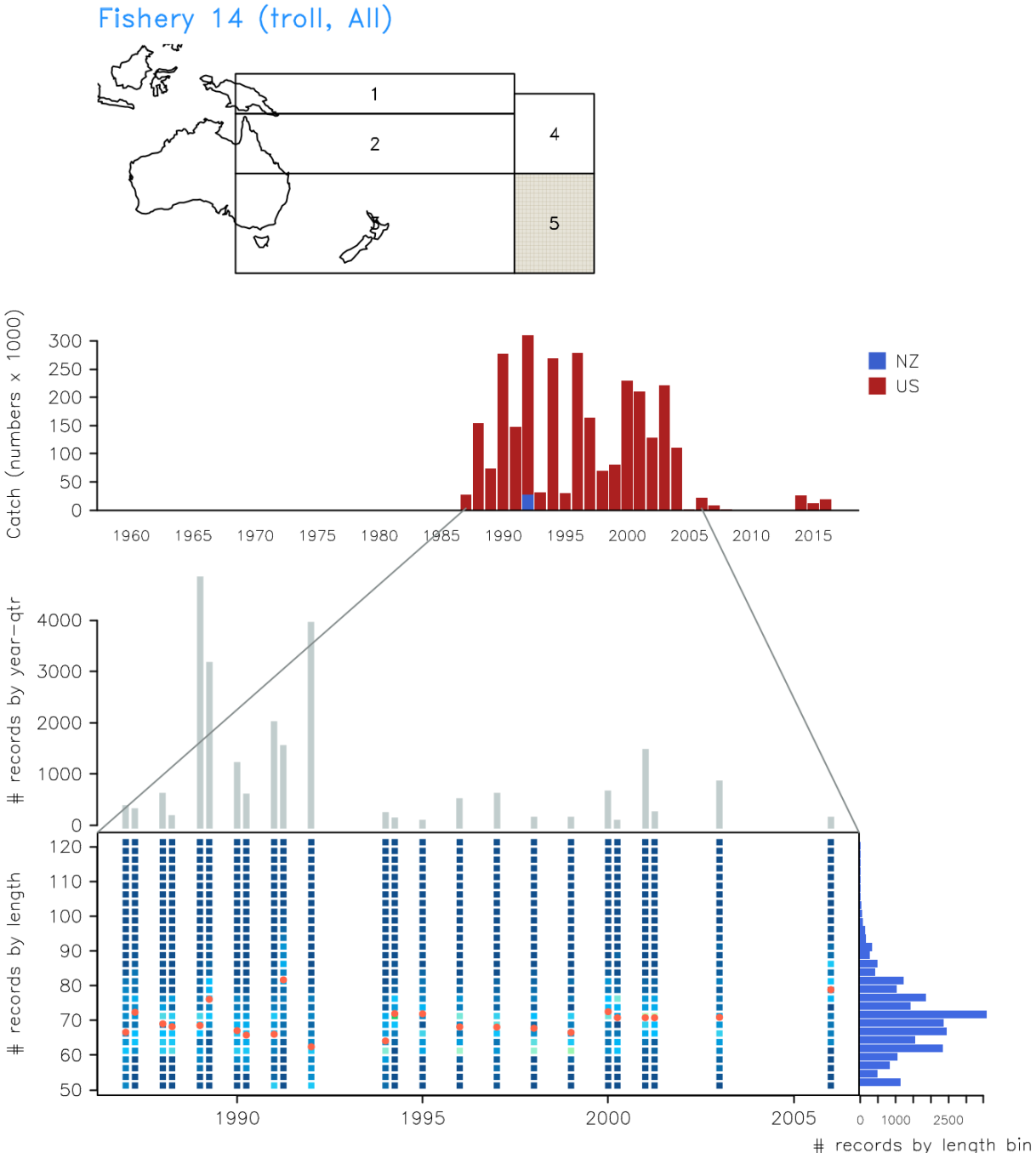

Figure 16: Summary of raw data available for fishery 14 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

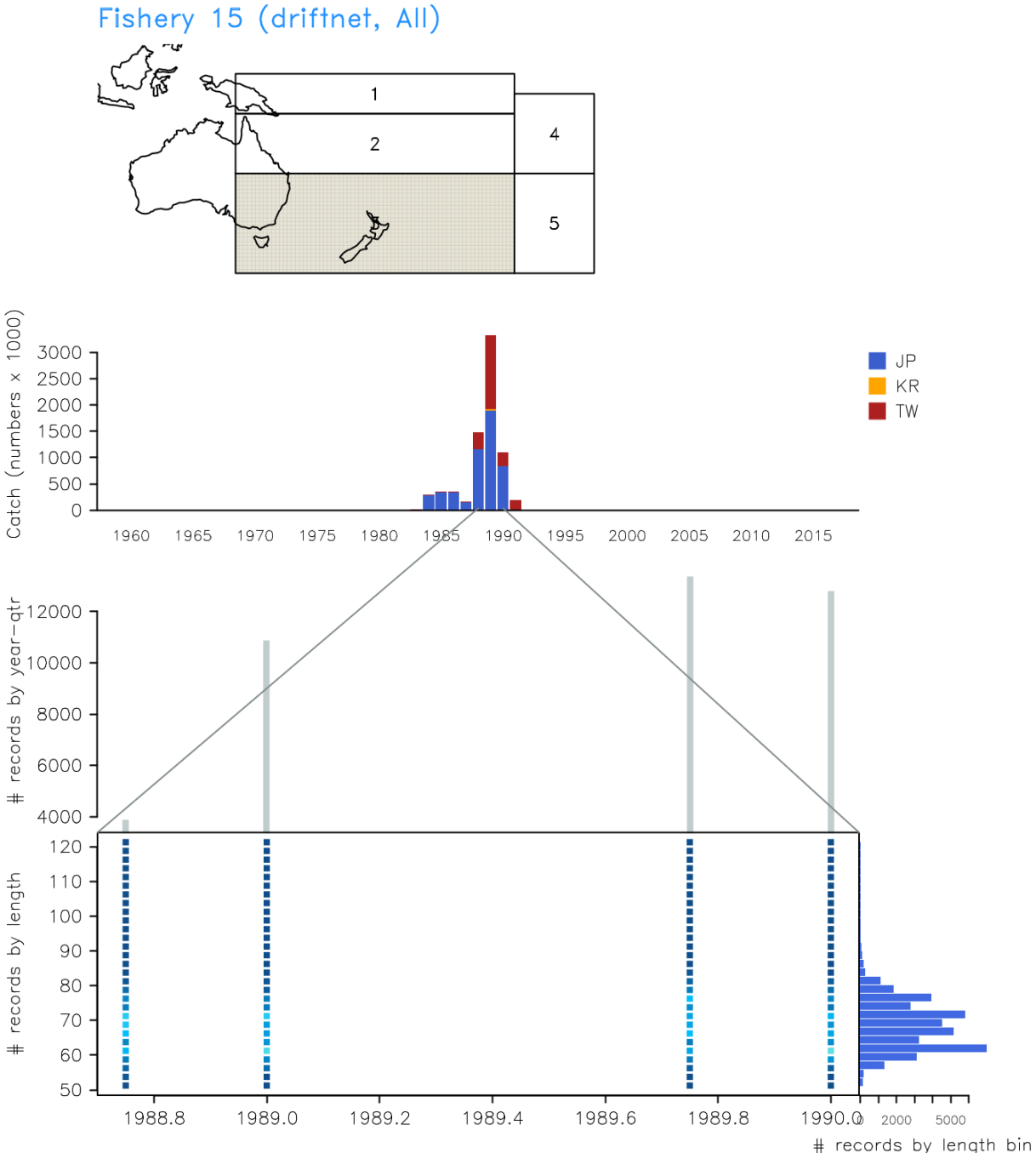

Figure 17: Summary of raw data available for fishery 15 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

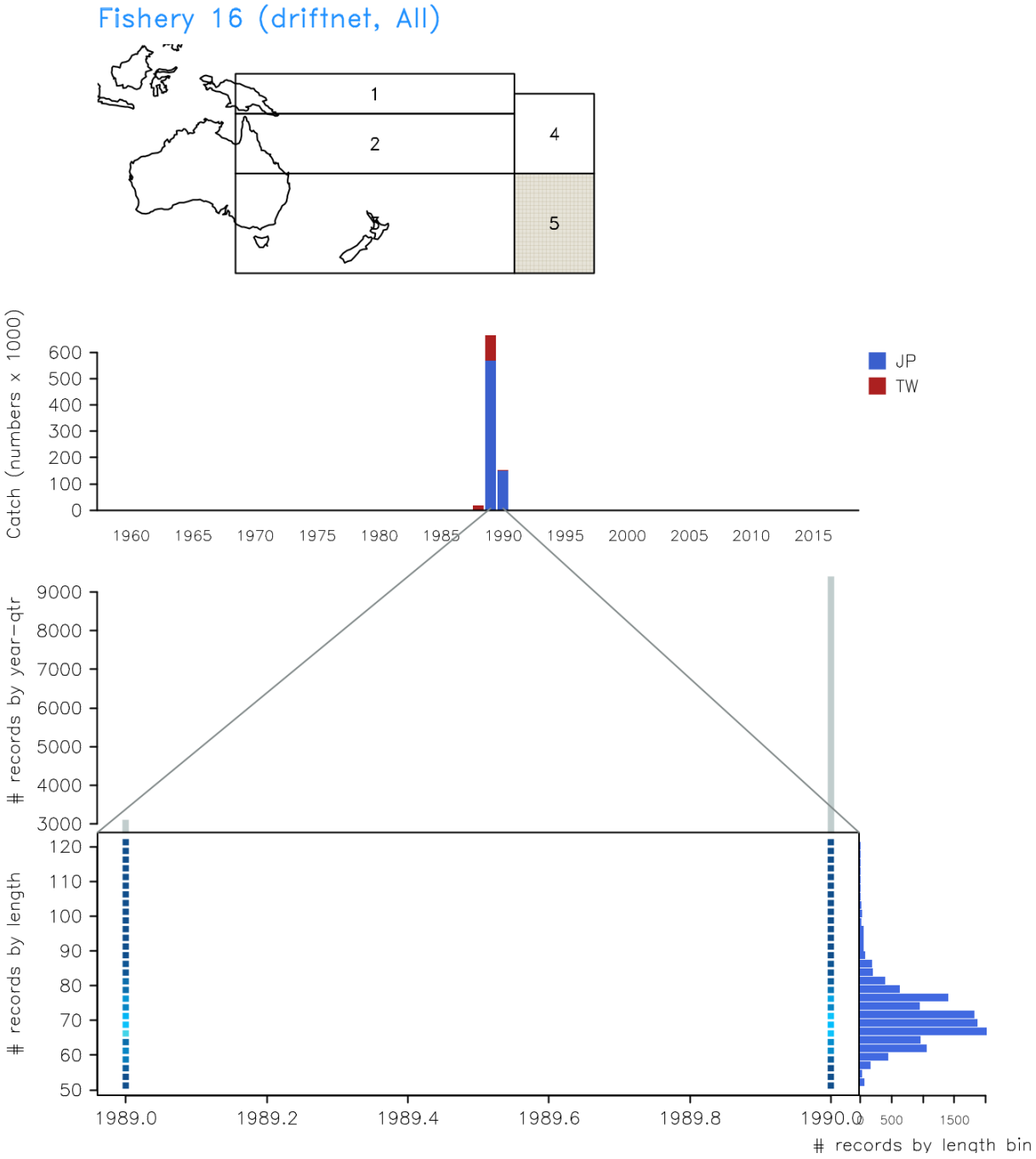

Figure 18: Summary of raw data available for fishery 16 of the 2018 albacore stock assessment. The panels display: the region of occurrence (top left), the annual catch by fleet within the fishery, in individuals (top middle panel), the annual number of fish with measured length (bottom middle panel), trends in length composition data with the median highlighted in red (bottom), and the overall size distribution over the time-span of the fishery (bottom right).

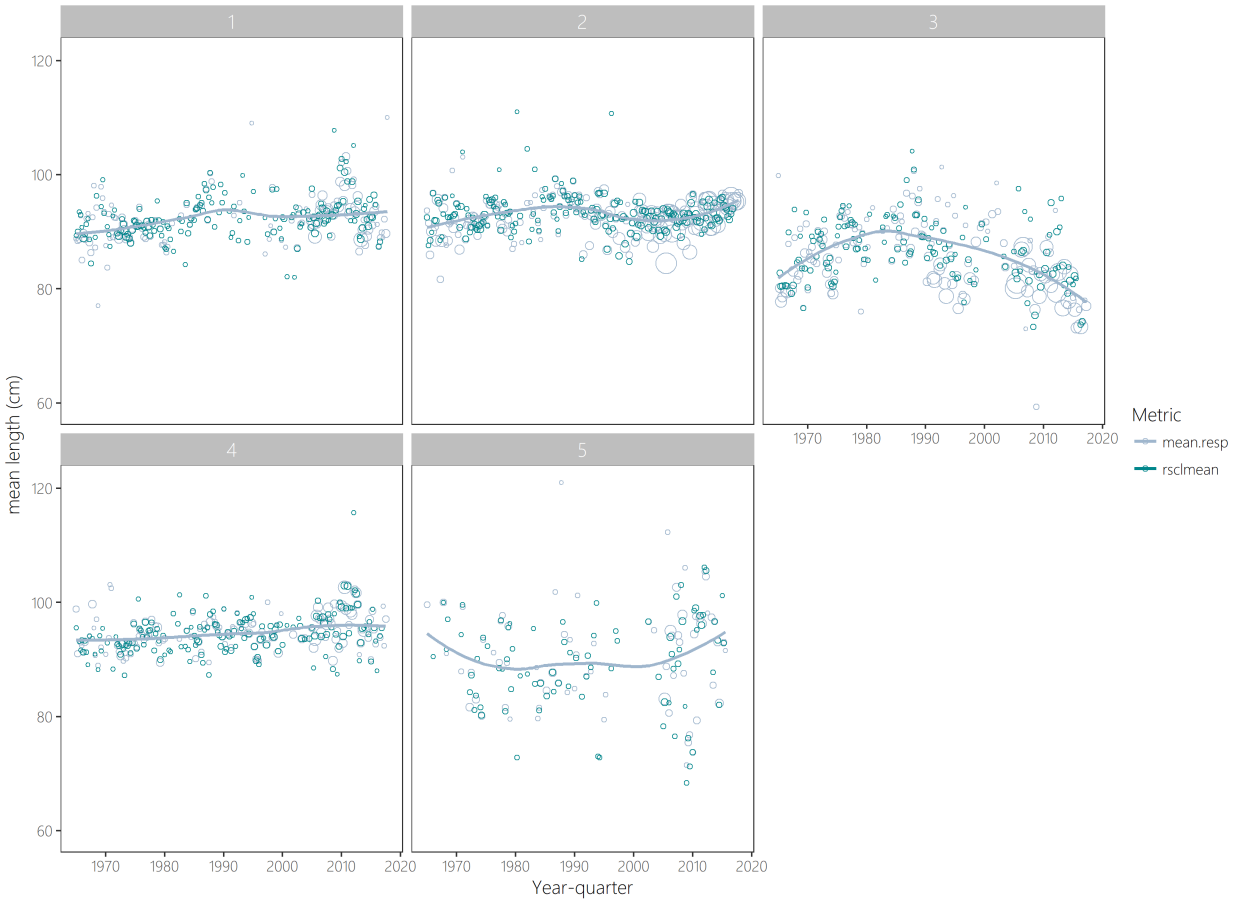

Figure 19: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the distant-water fishing nations: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

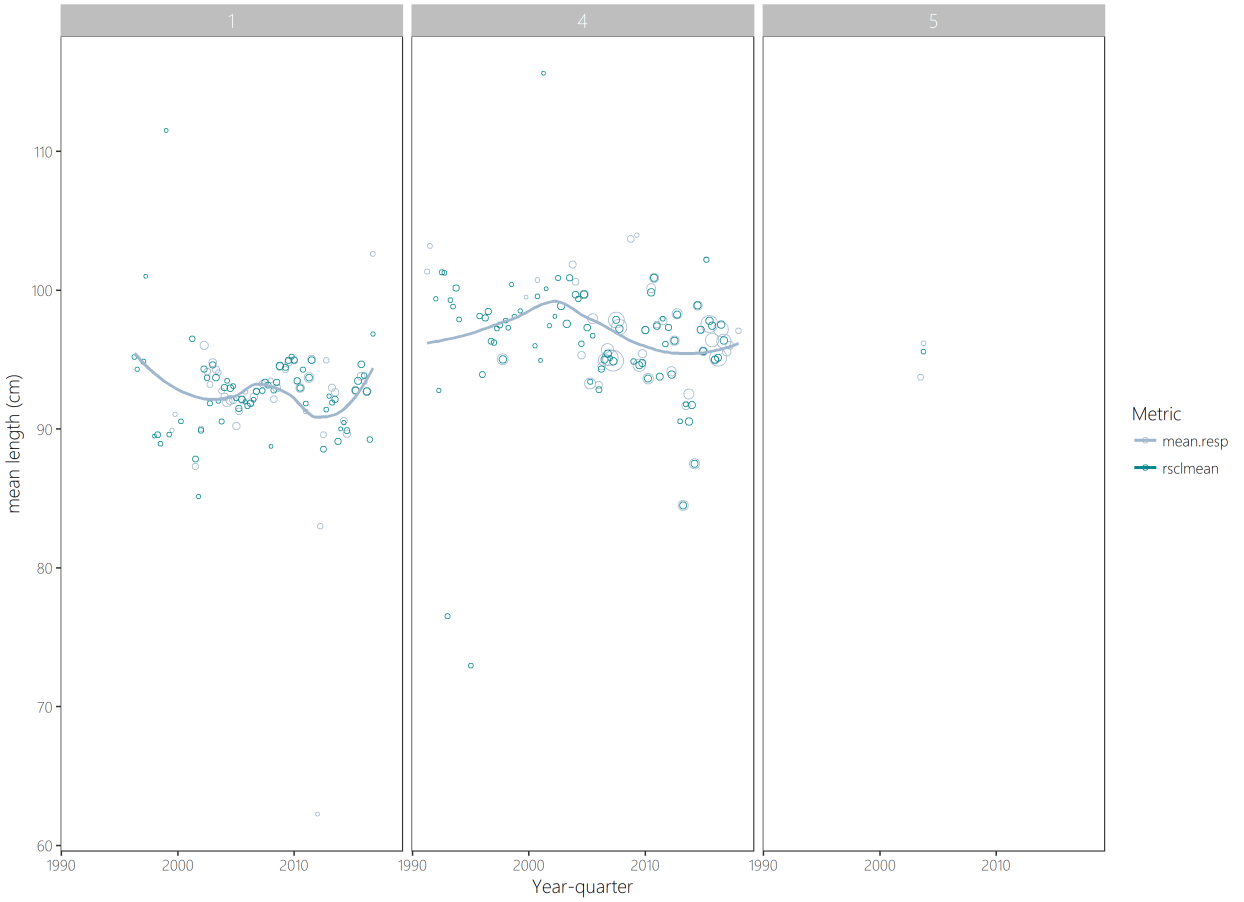

Figure 20: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the Pacific Island Countries and Territories: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

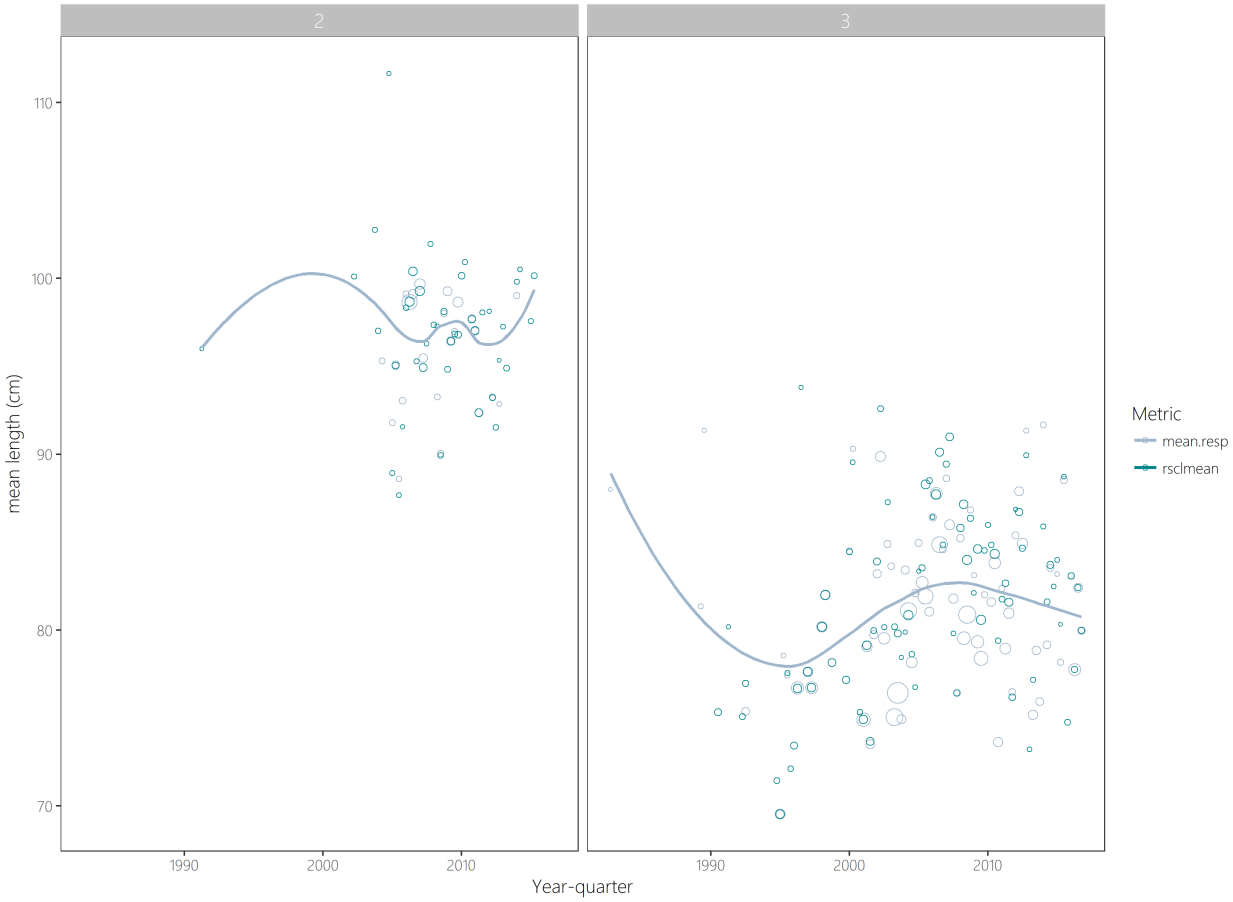

Figure 21: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the Australia and New Zealand fleets: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

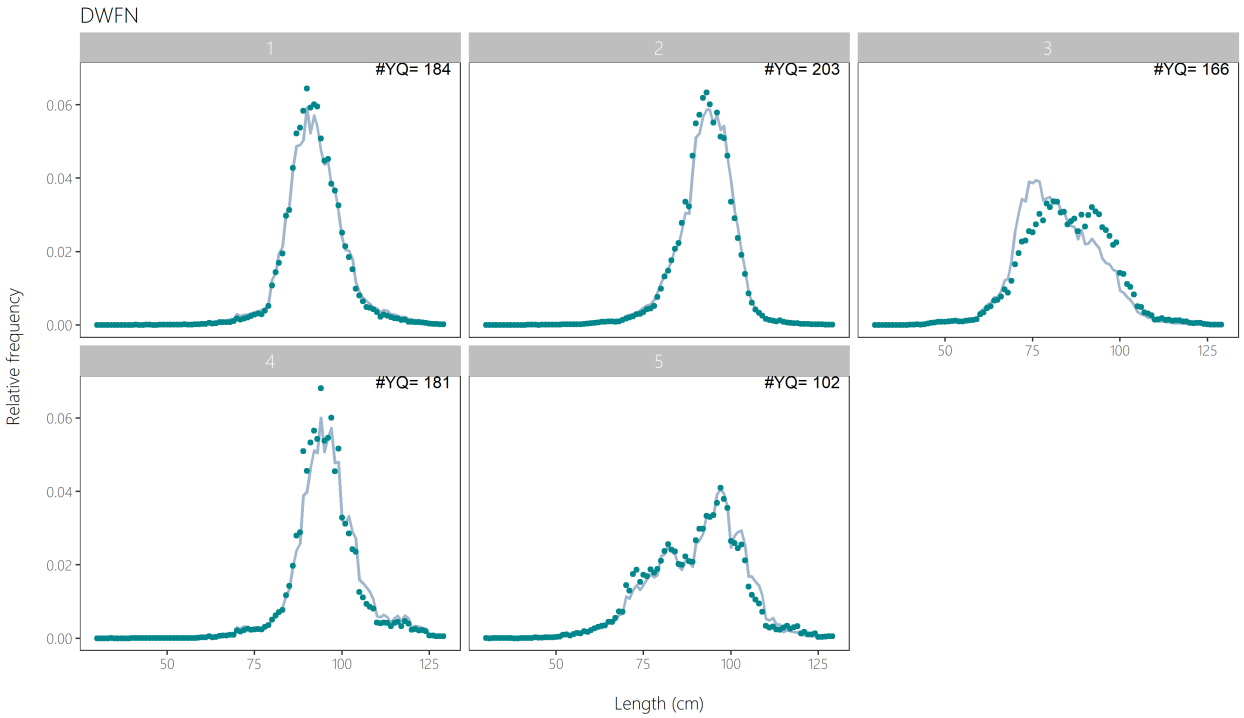

Figure 22: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the distant-water fishing nations: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

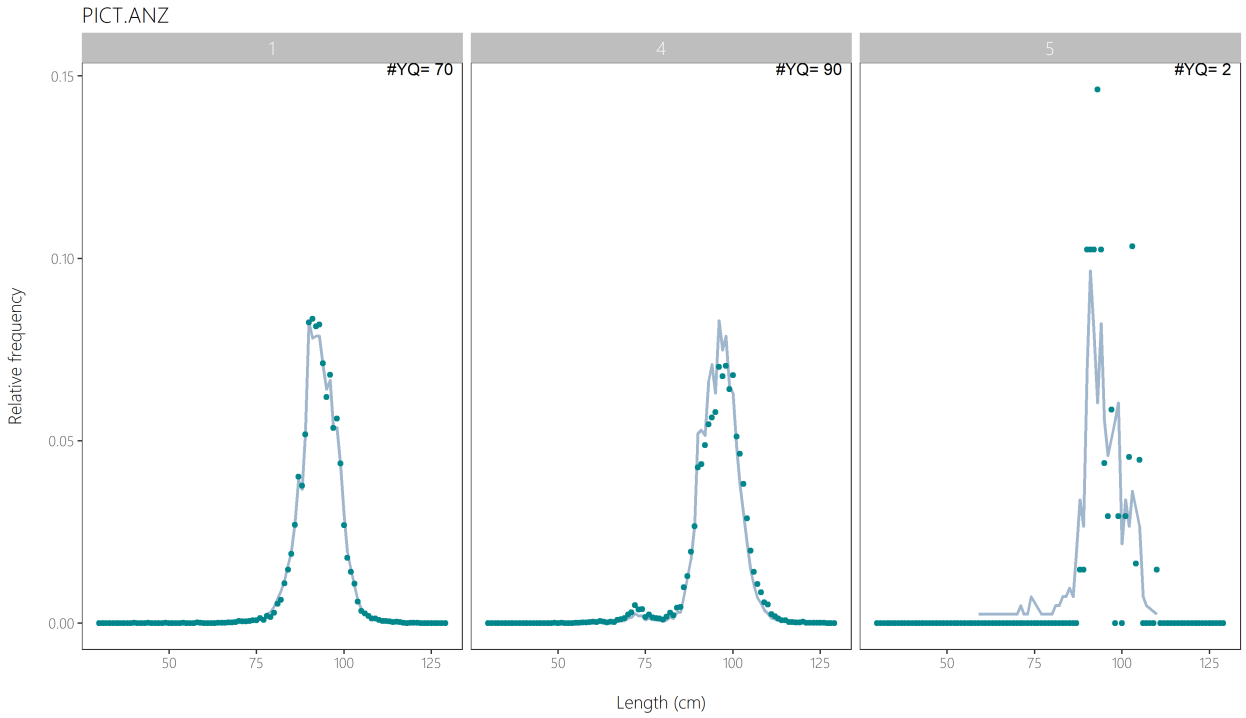

Figure 23: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the Pacific Island Countries and Territories: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

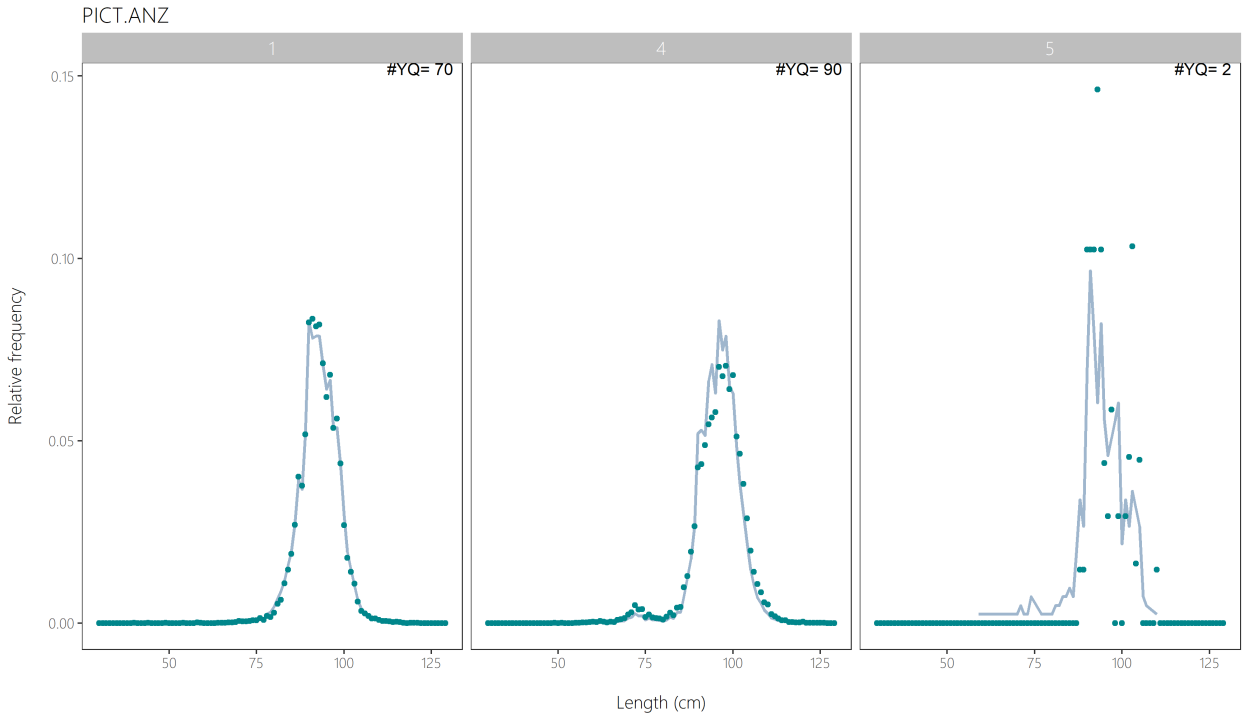

Figure 24: Density plots of raw (light blue) and re-weighted (dark green) albacore length frequency data by region for the Australia and New Zealand fleets: Re-weighted LFs are based on flag weighting for an 11 quarter time period and a 10% threshold for the lowest total weighting.

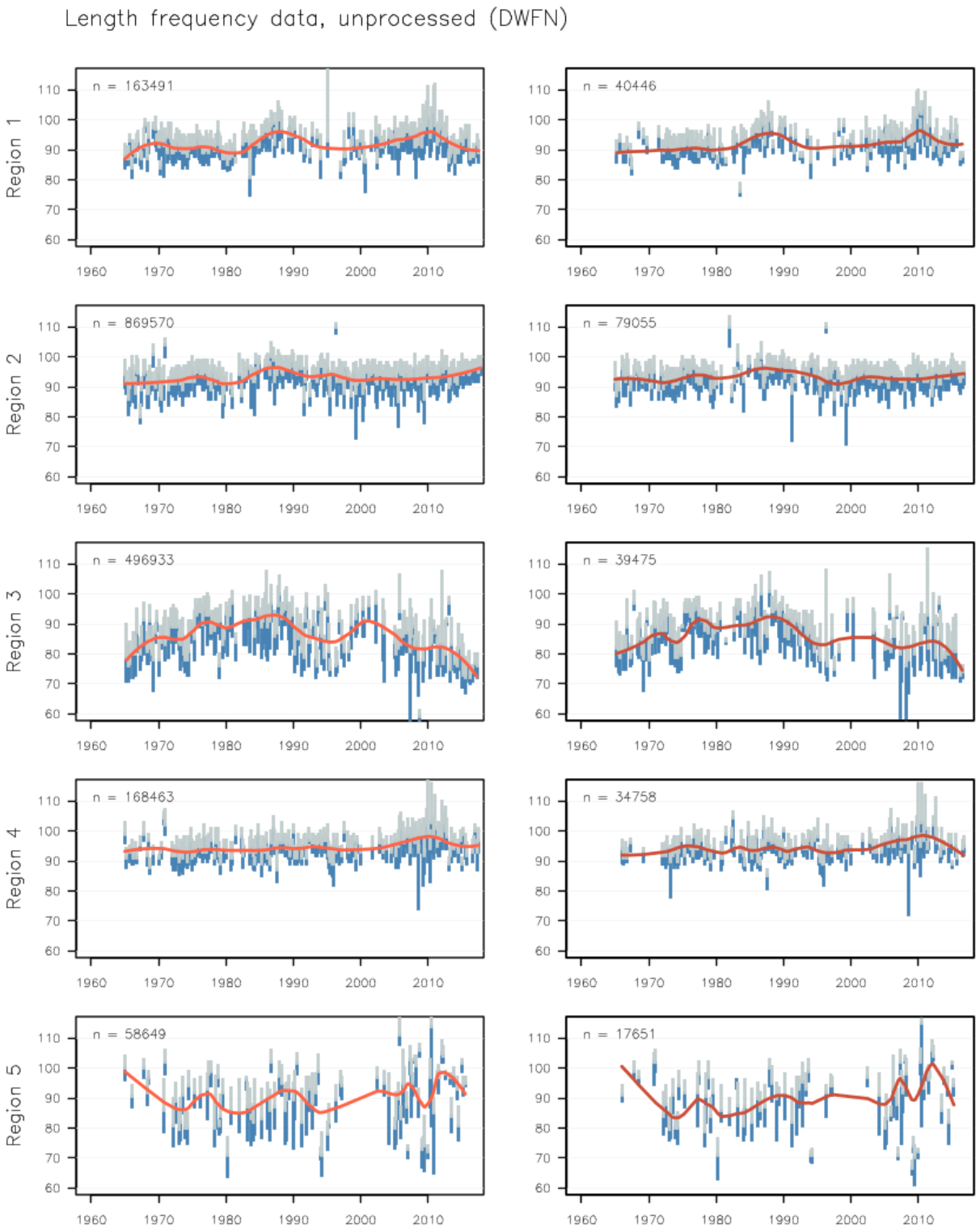

Figure 25: Comparison of 25th and 75th quantile in length distribution (in cm) by region for the distant-water fishing nations before and after reweighting, with a spline fitted to the median added in red.

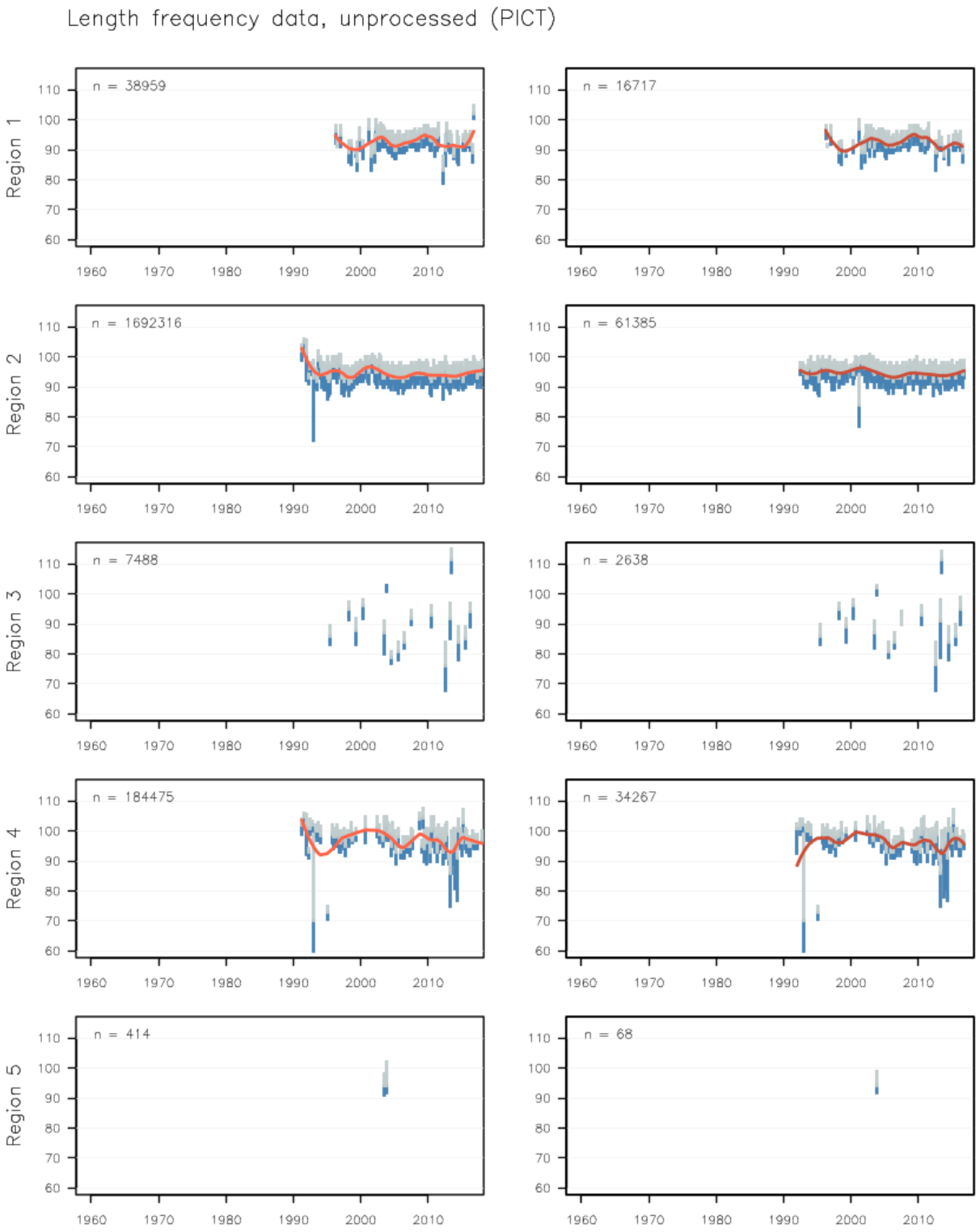

Figure 26: Comparison of 25th and 75th quantile in length distribution (in cm) by region for the Pacific Island Countries and Territories before and after reweighting, with a spline fitted to the median added in red.

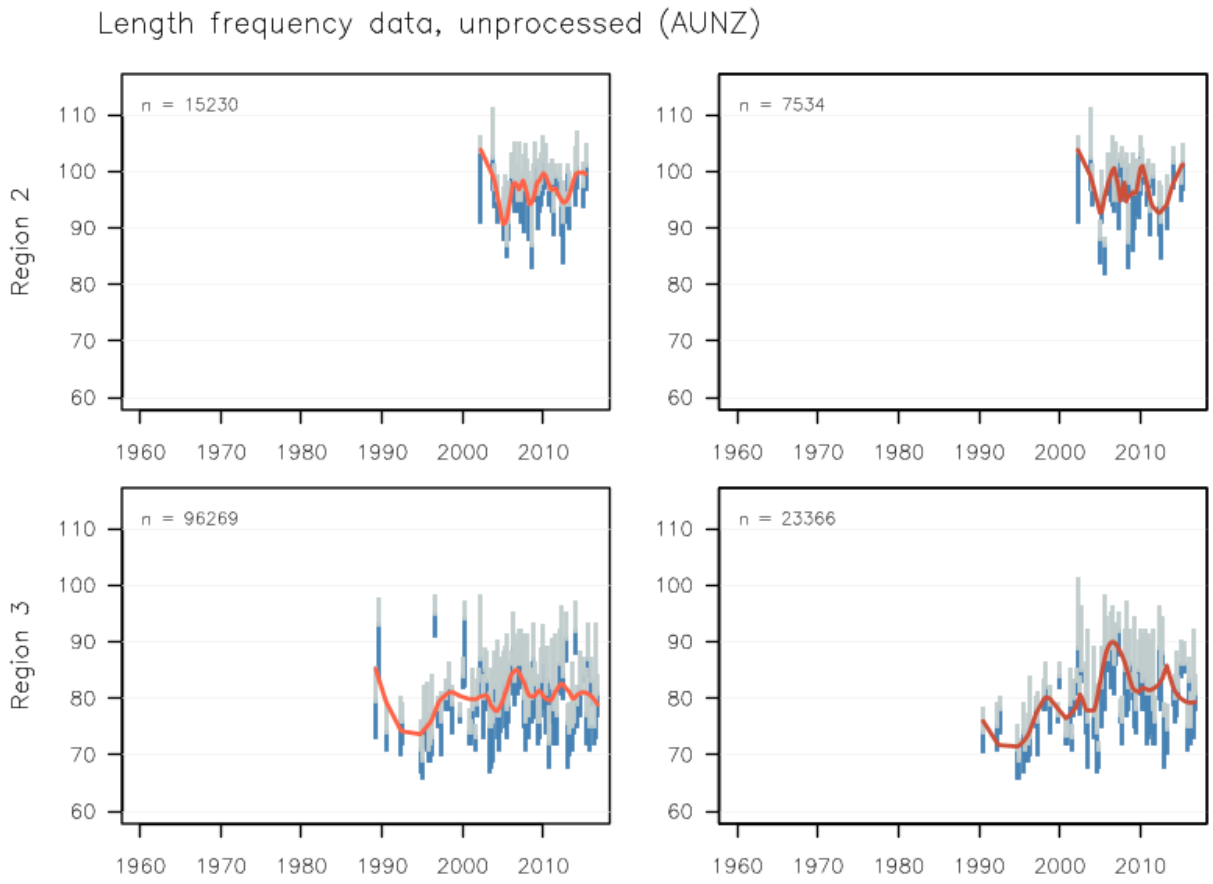

Figure 27: Comparison of 25th and 75th quantile in length distribution (in cm) by region for the Australia and New Zealand fleets before and after reweighting, with a spline fitted to the median added in red.

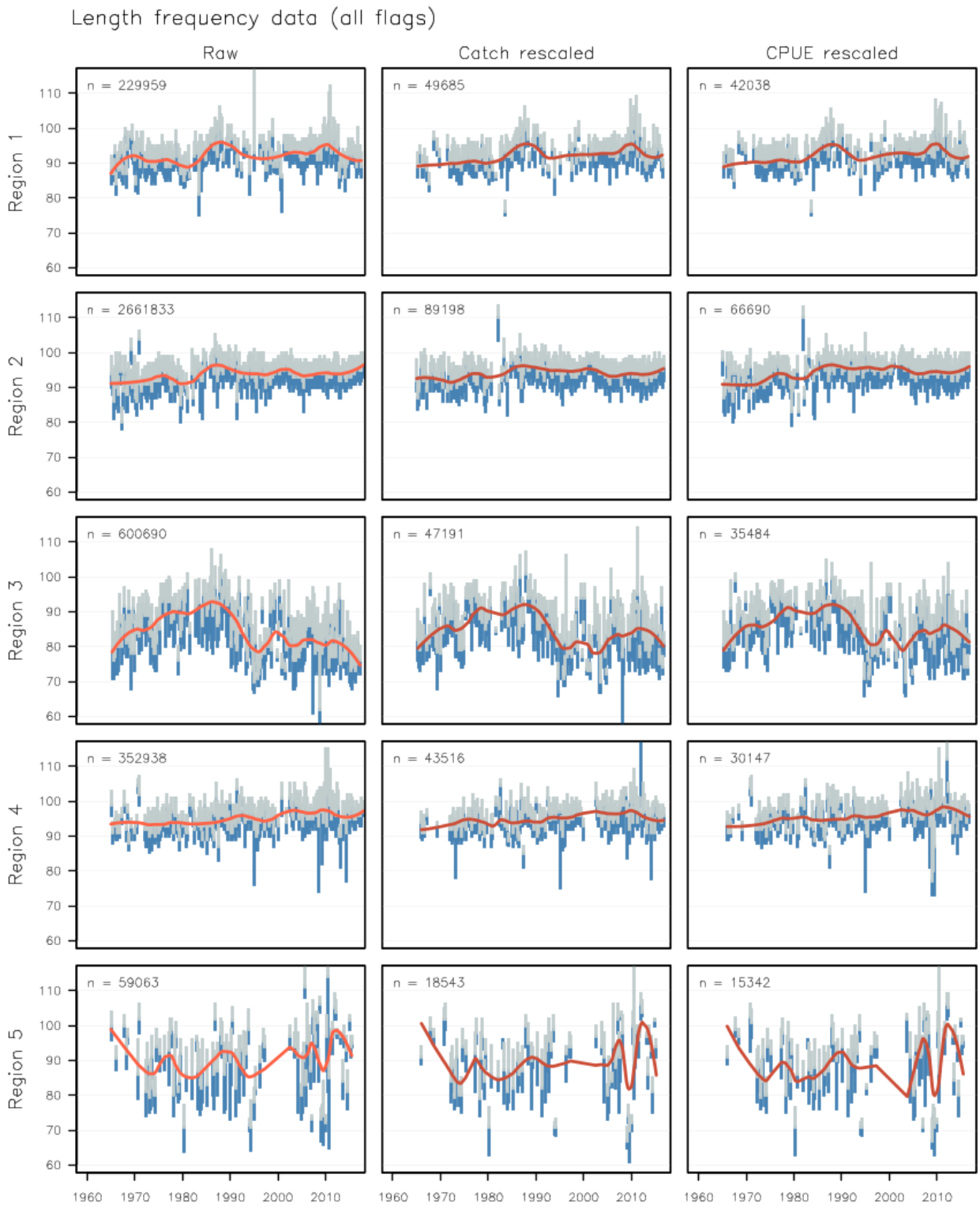

Figure 28: Comparison of 25th and 75th quantile in length distribution (in cm) by region, all-fleet combined, before and after reweighting with catch (center) and CPUE (right) weights, with a spline fitted to the median added in red. The CPUE-reweighted length distribution was used as the size data input for the 'Index' fisheries in the 2018 South Pacific albacore assessment.

Figure 29: Depiction of the clustering results for region 1. The three panels represent clustering undertaken by imposing 2, 3 and 4 clusters, respectively. The different bars within a panel indicate the proportion of the four species caught over all sets included in that cluster, with the name on the x axis indicating the most dominant species. The width of the bars are scaled to represent the number of trips included in that cluster.

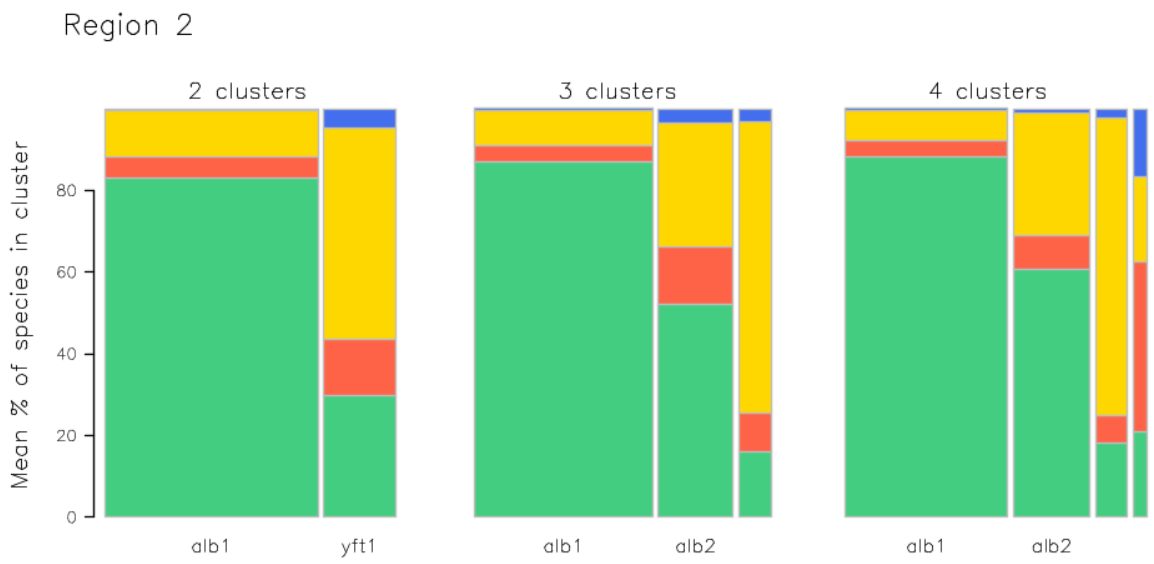

Figure 30: Depiction of the clustering results for region 2. The three panels represent clustering undertaken by imposing 2, 3 and 4 clusters, respectively. The different bars within a panel indicate the proportion of the four species caught over all sets included in that cluster, with the name on the x axis indicating the most dominant species. The width of the bars are scaled to represent the number of trips included in that cluster.

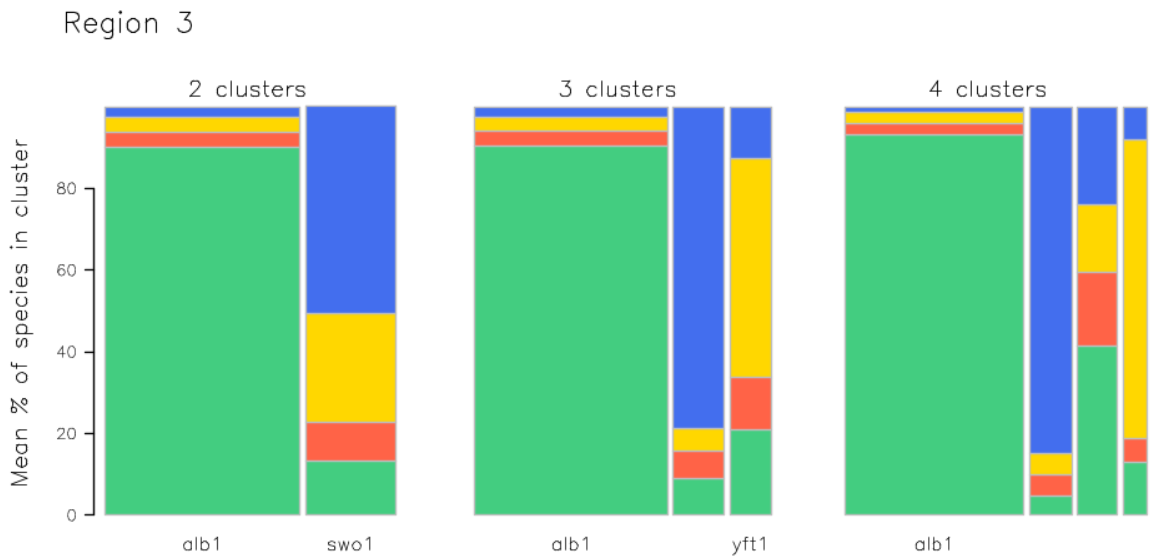

Figure 31: Depiction of the clustering results for region 3. The three panels represent clustering undertaken by imposing 2, 3 and 4 clusters, respectively. The different bars within a panel indicate the proportion of the four species caught over all sets included in that cluster, with the name on the x axis indicating the most dominant species. The width of the bars are scaled to represent the number of trips included in that cluster.

Figure 32: Depiction of the clustering results for region 4. The three panels represent clustering undertaken by imposing 2, 3 and 4 clusters, respectively. The different bars within a panel indicate the proportion of the four species caught over all sets included in that cluster, with the name on the x axis indicating the most dominant species. The width of the bars are scaled to represent the number of trips included in that cluster.

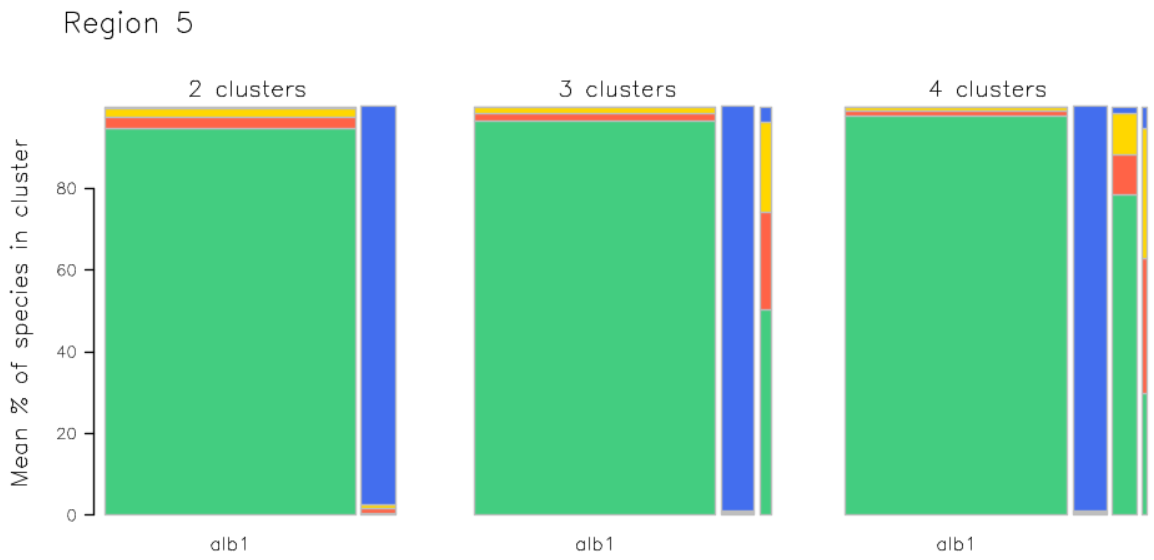

Figure 33: Depiction of the clustering results for region 5. The three panels represent clustering undertaken by imposing 2, 3 and 4 clusters, respectively. The different bars within a panel indicate the proportion of the four species caught over all sets included in that cluster, with the name on the x axis indicating the most dominant species. The width of the bars are scaled to represent the number of trips included in that cluster.

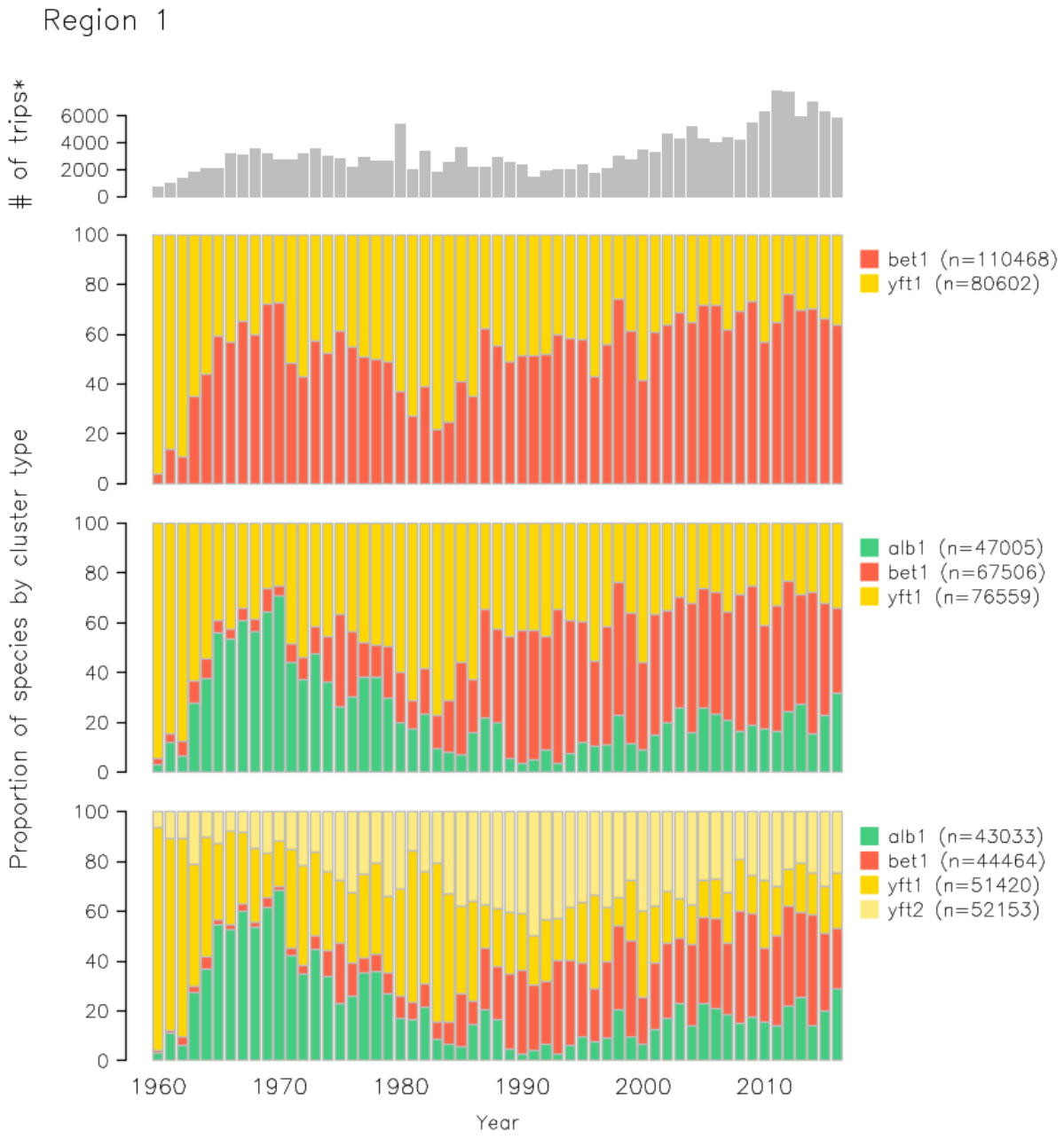

Figure 34: Time-series of cluster membership for the 2, 3 and 4 cluster models in region 1, with the colour matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

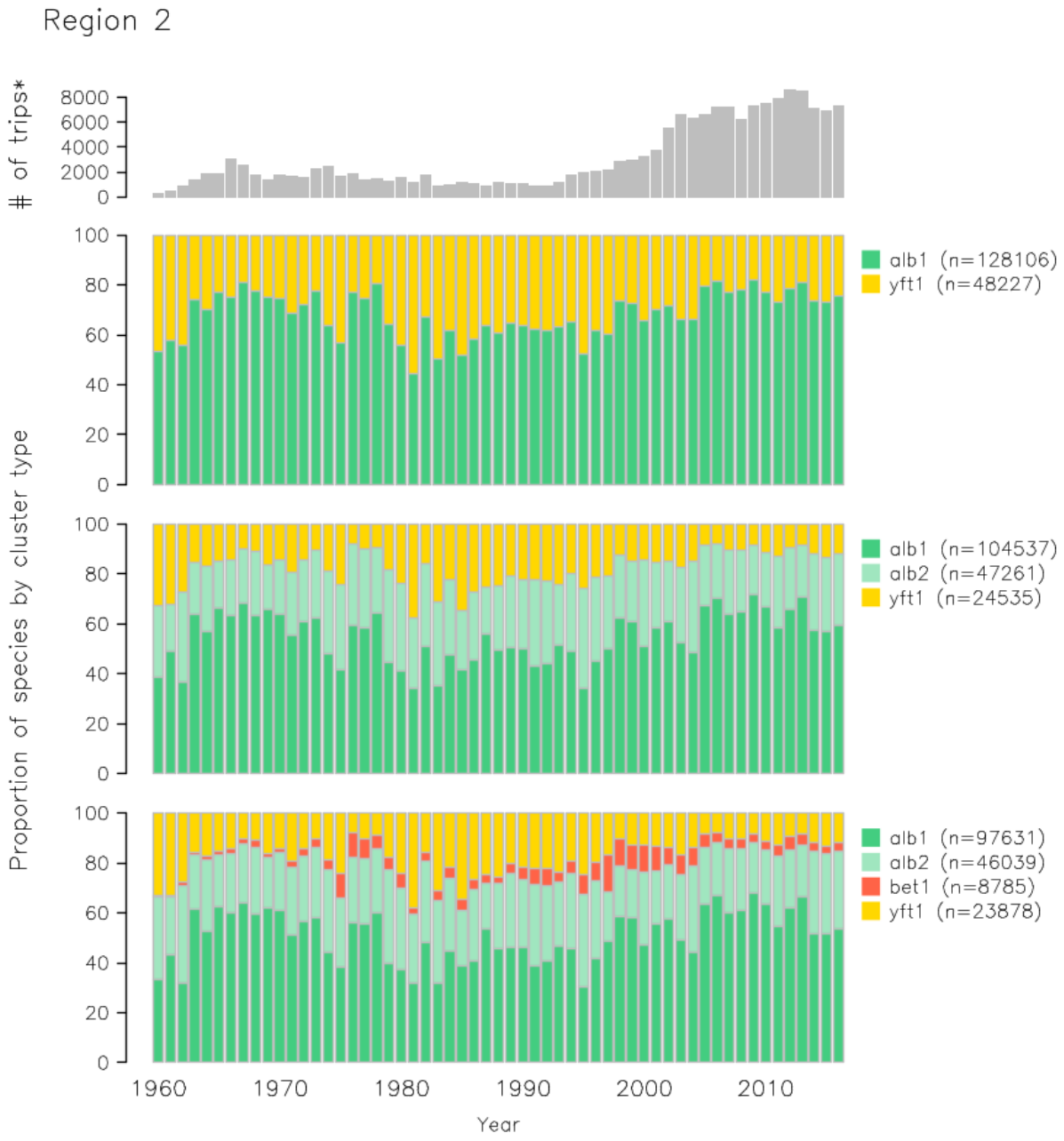

Figure 35: Time-series of cluster membership for the 2, 3 and 4 cluster models in region 2, with the colour matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

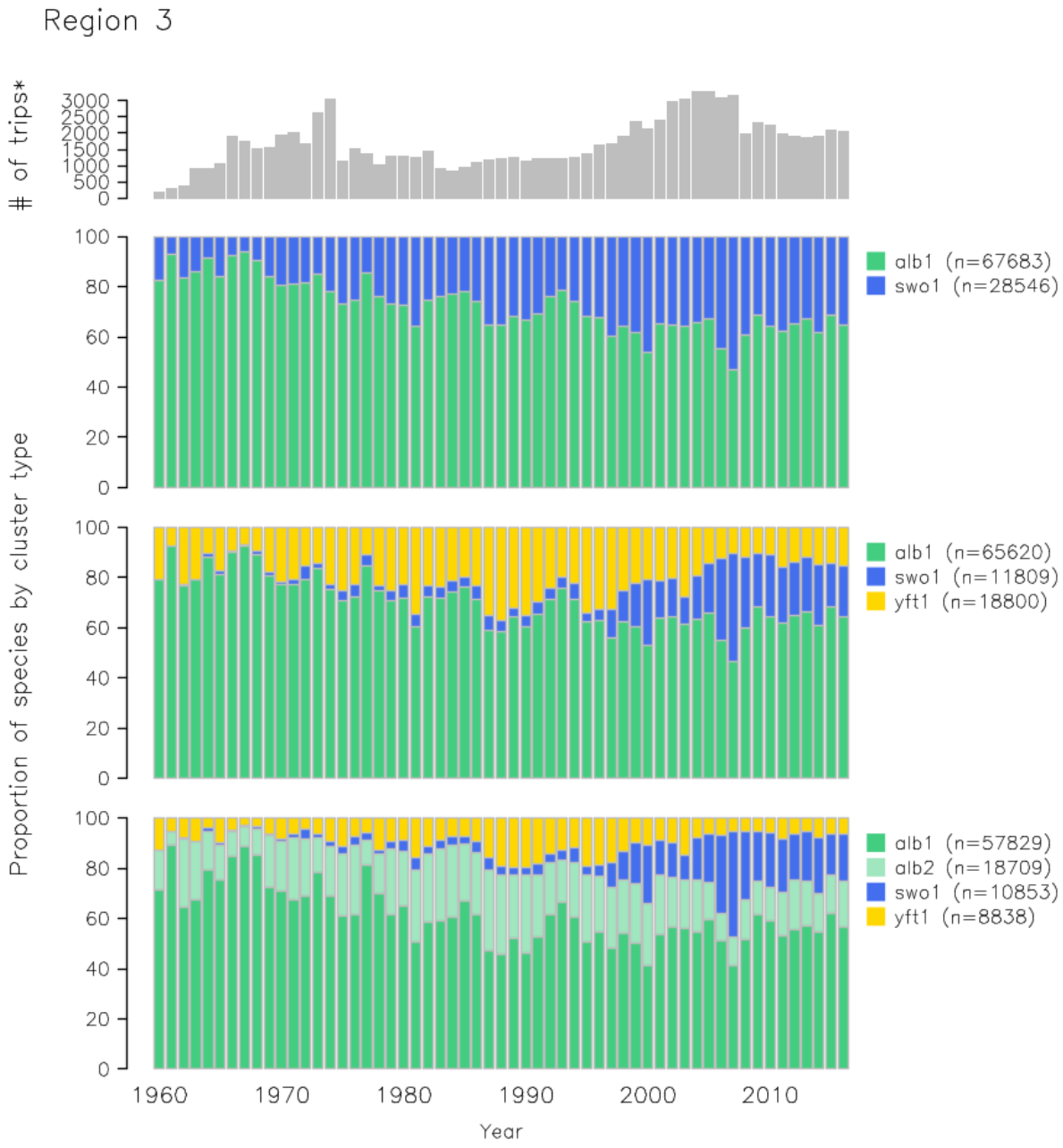

Figure 36: Time-series of cluster membership for the 2, 3 and 4 cluster models in region 3, with the colour matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

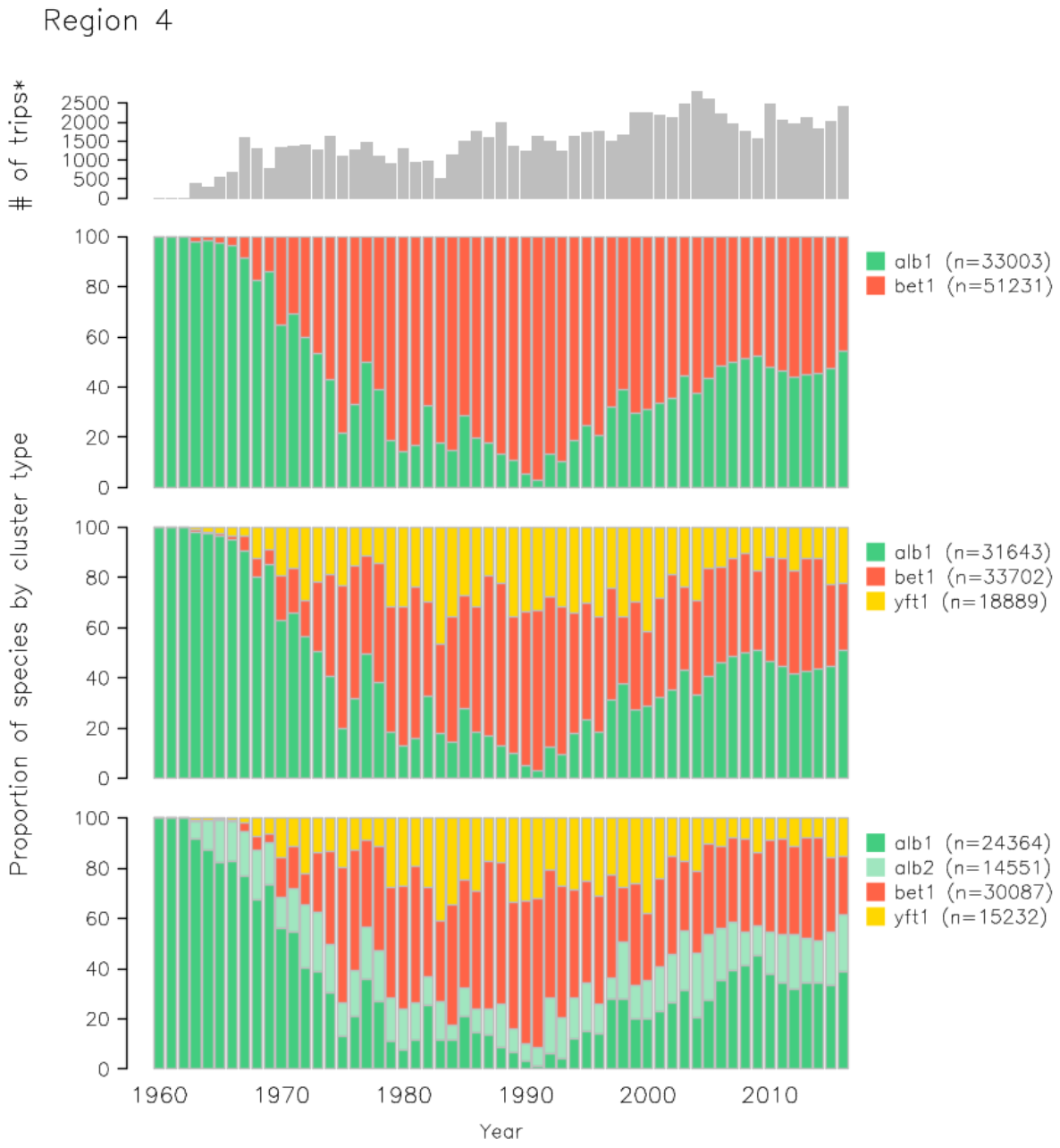

Figure 37: Time-series of cluster membership for the 2, 3 and 4 cluster models in region 4, with the colour matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

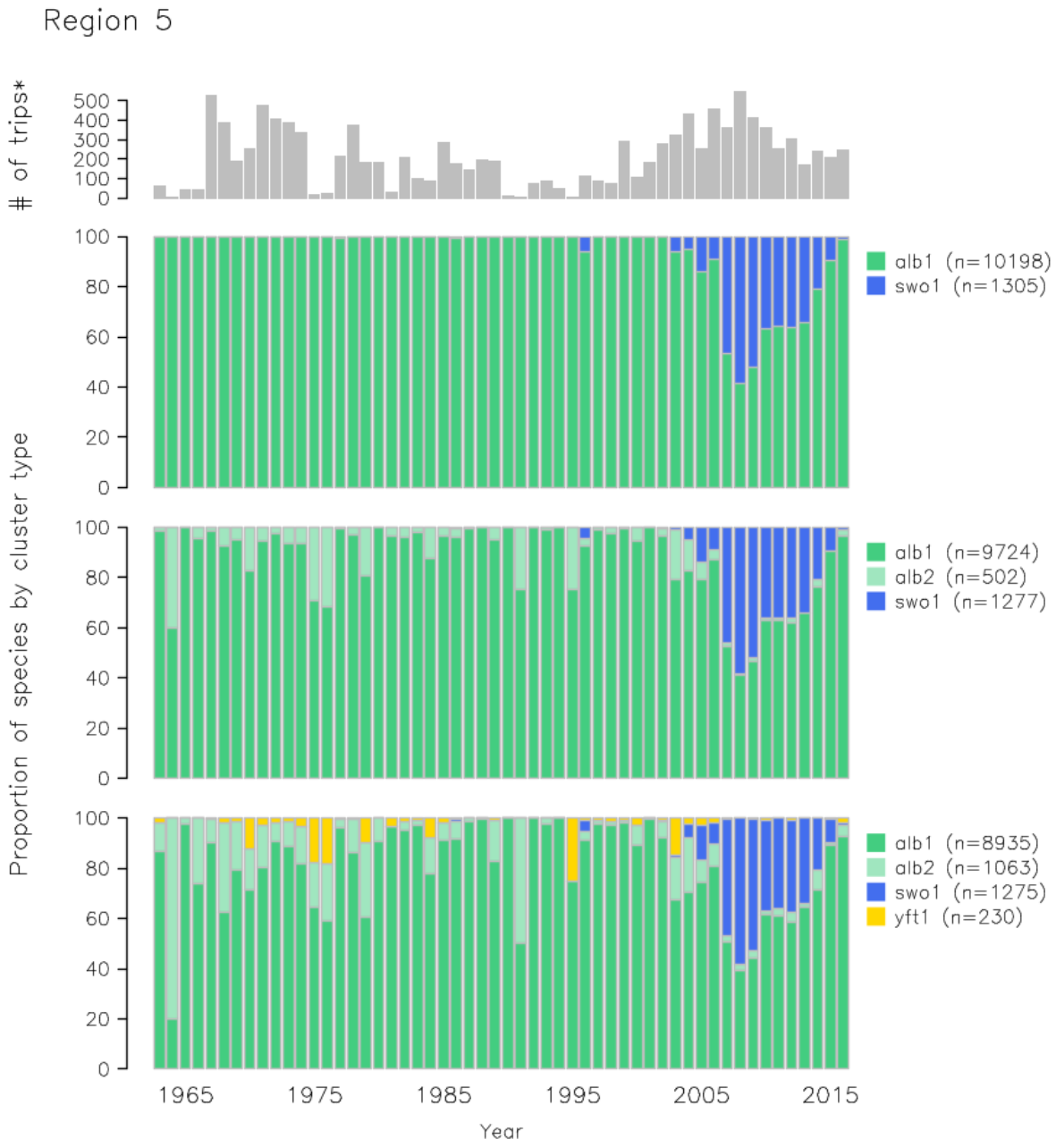

Figure 38: Time-series of cluster membership for the 2, 3 and 4 cluster models in region 5, with the colour matching the most dominant species in the cluster and the top panel indicating the number of trips available in each year.

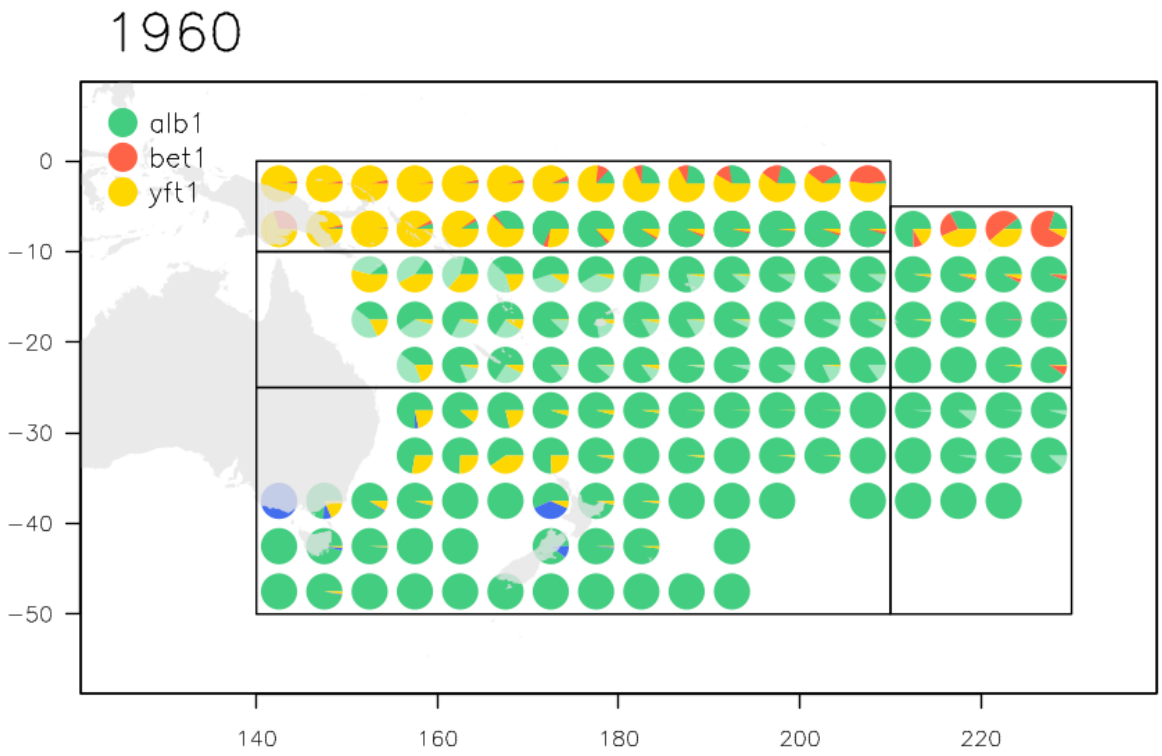

Figure 39: Map of cluster composition at the 5 degree scale, by decade (in this case 1960-69), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to.

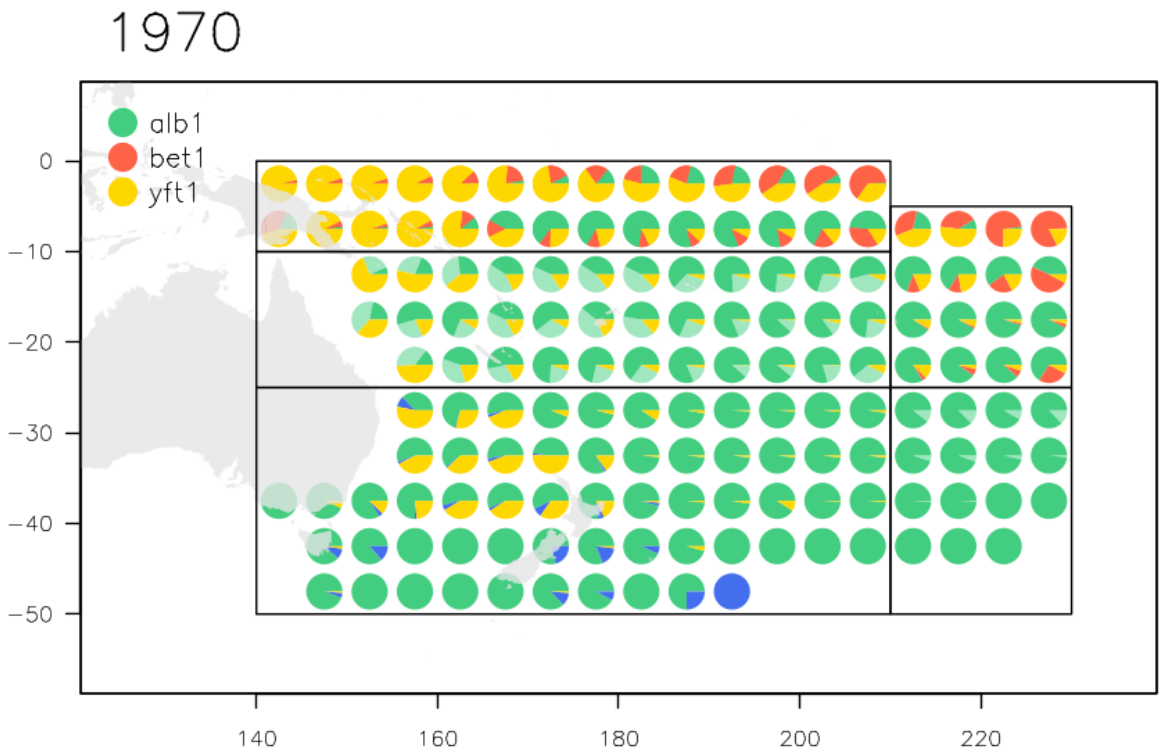

Figure 40: Map of cluster composition at the 5 degree scale, by decade (in this case 1970-79), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to..

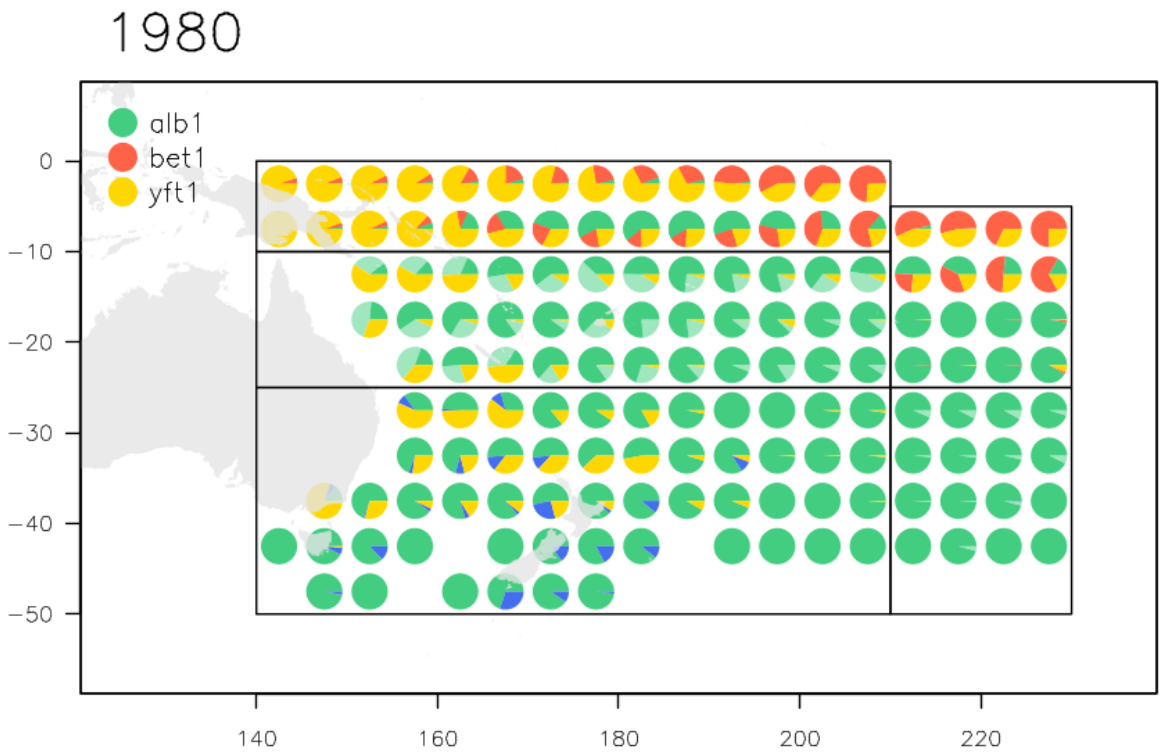

Figure 41: Map of cluster composition at the 5 degree scale, by decade (in this case 1980-89), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to..

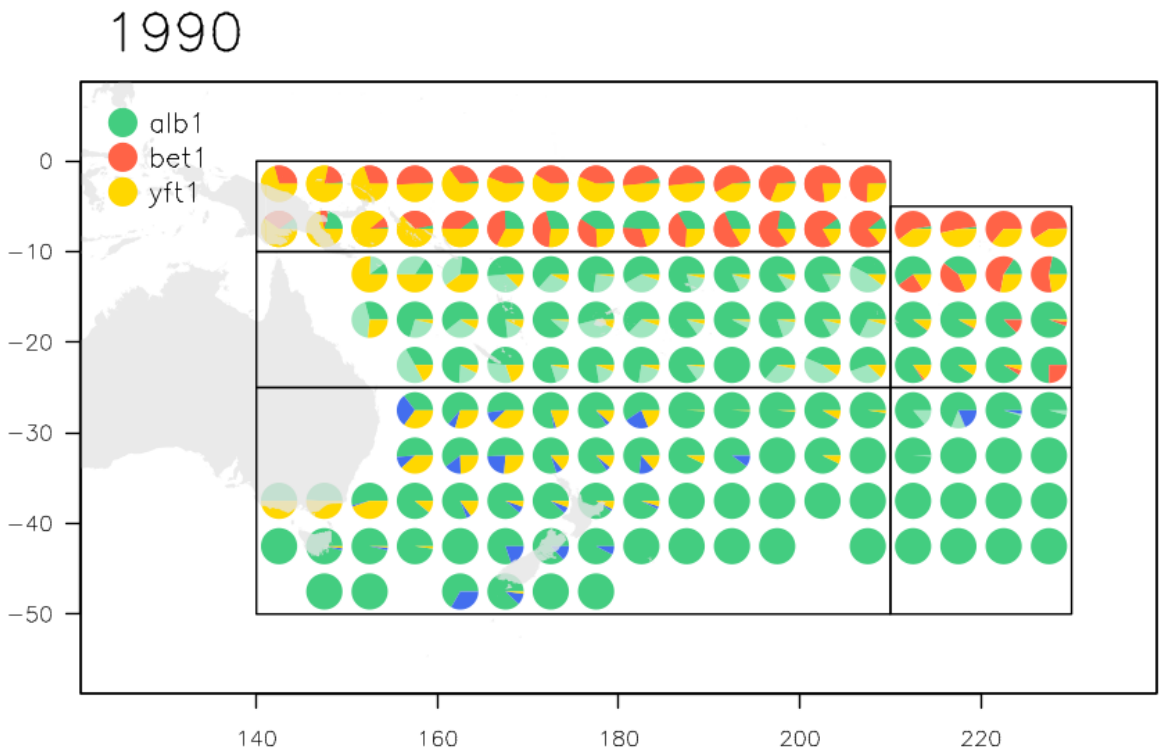

Figure 42: Map of cluster composition at the 5 degree scale, by decade (in this case 1990-99), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to..

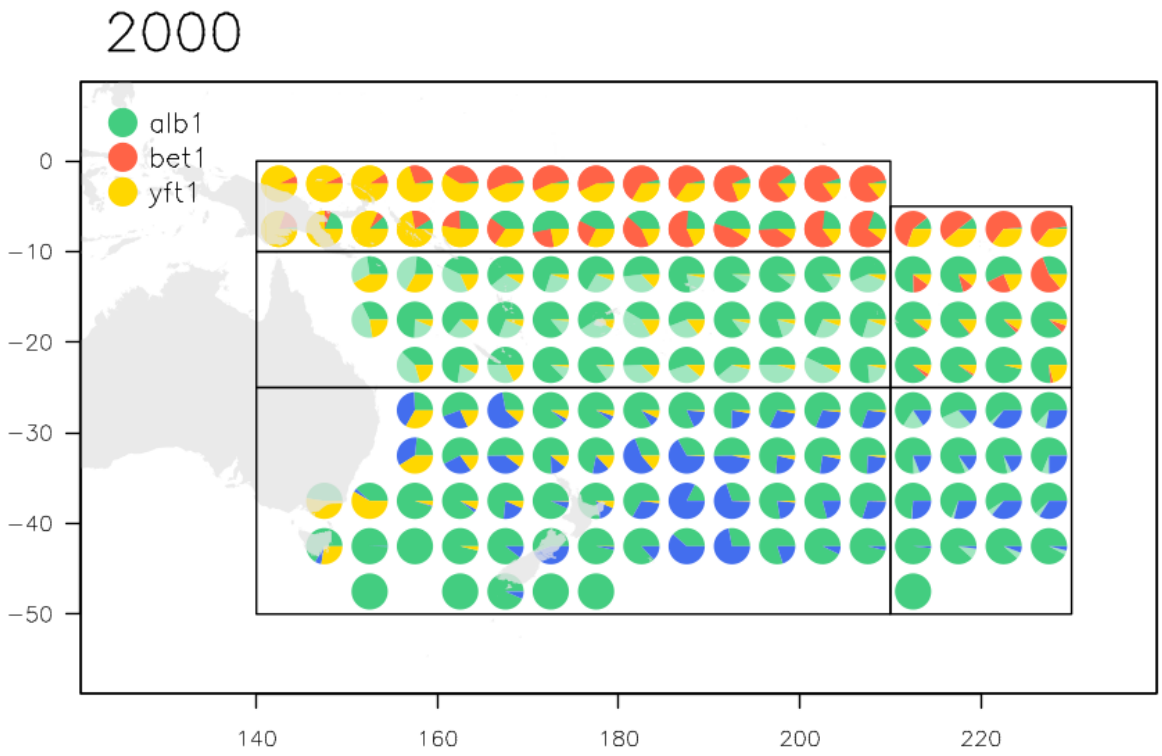

Figure 43: Map of cluster composition at the 5 degree scale, by decade (in this case 2000-09), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to..

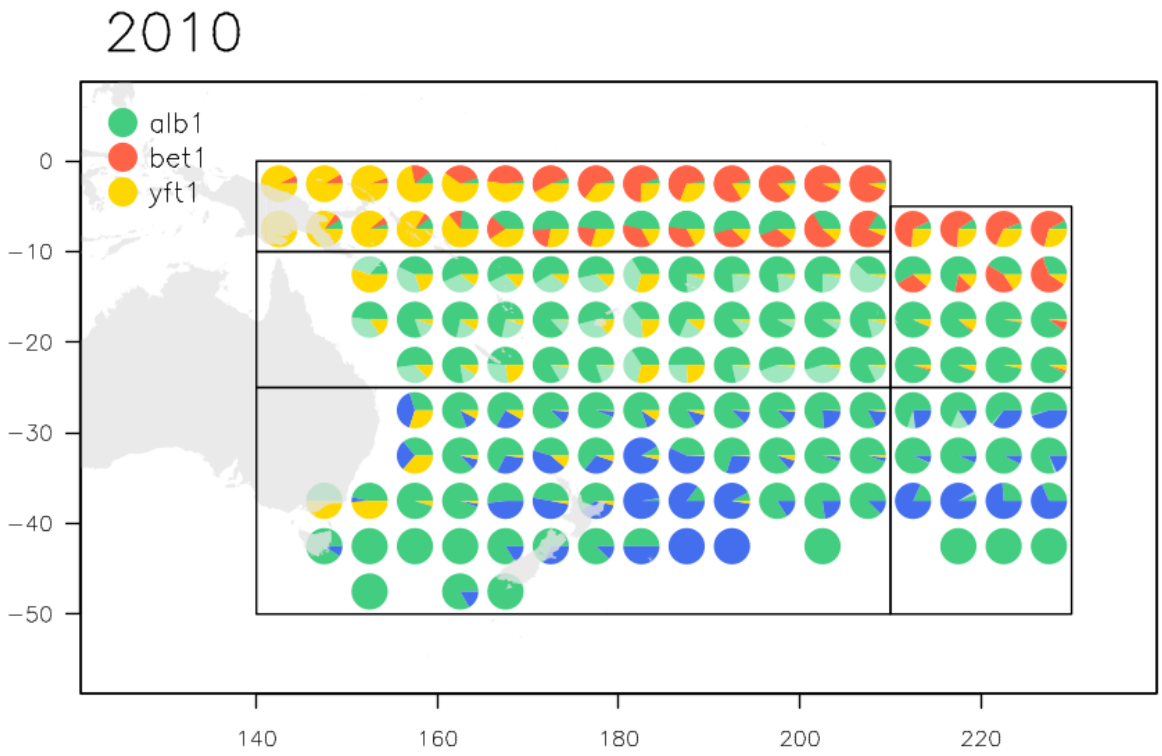

Figure 44: Map of cluster composition at the 5 degree scale, by decade (in this case 2010-16), with region-specific cluster numbers where the colours of the segments indicate the clusters to which the data belongs to..

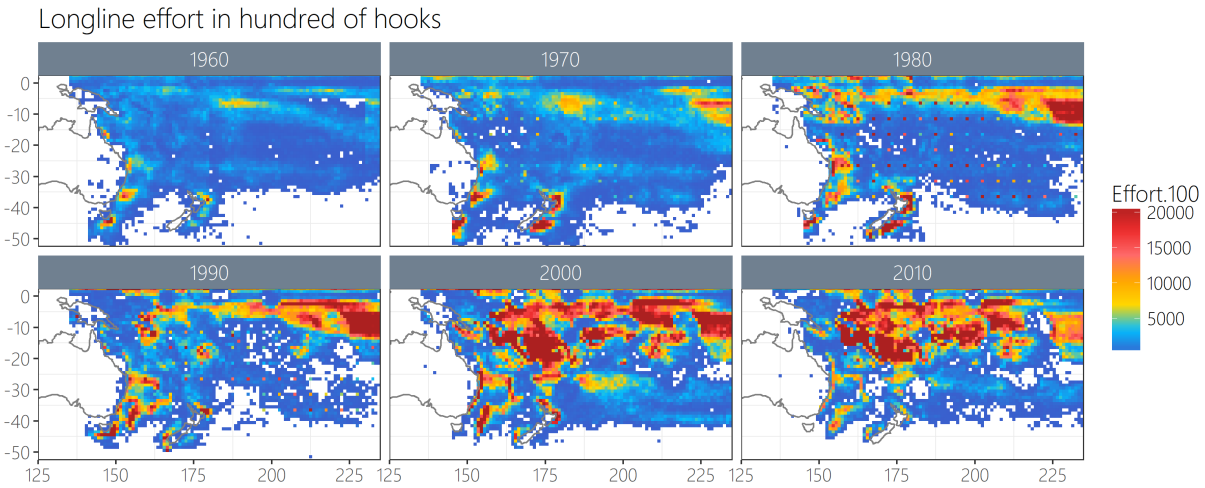

Figure 45: Map of effort in hooks for the South Pacific aggregated across fleets at the one-degree resolution.

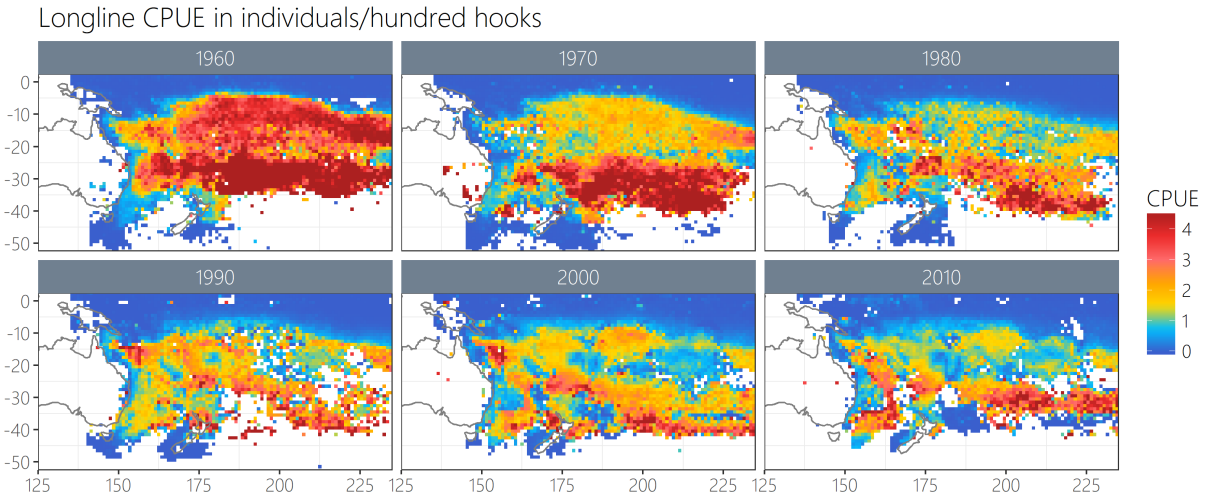

Figure 46: Map of CPUE in individuals per hook for the South Pacific aggregated across fleets at the one-degree resolution.

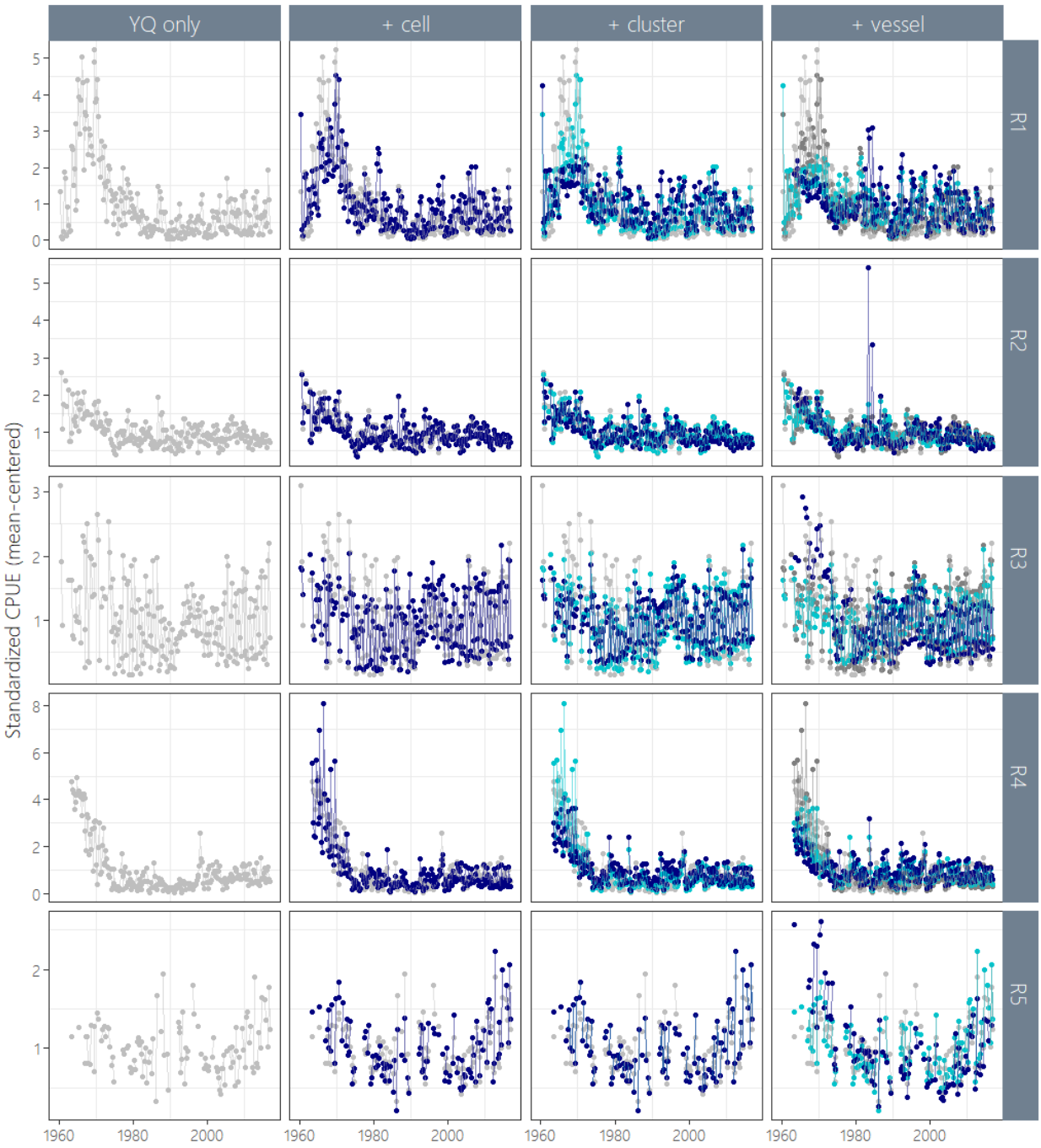

Figure 47: Step plots showing the effects of removing individual factors from the GLM model with respect to the estimated CPUE indices produced using the 'traditional' CPUE approach. The left hand panel shows the model fitted with just a year-quarter factor, the panel second from the left shows the model with the cell factor added (dark blue line; grey line is the year-quarter-only model), the next panel shows the model with the cluster factor added (dark blue line; light blue line is the cell + year-quarter model, etc.) and finally the right panel shows a model with the vessel factor added (dark blue line; the light blue line is the previous model - cluster + cell + year-quarter, etc.)

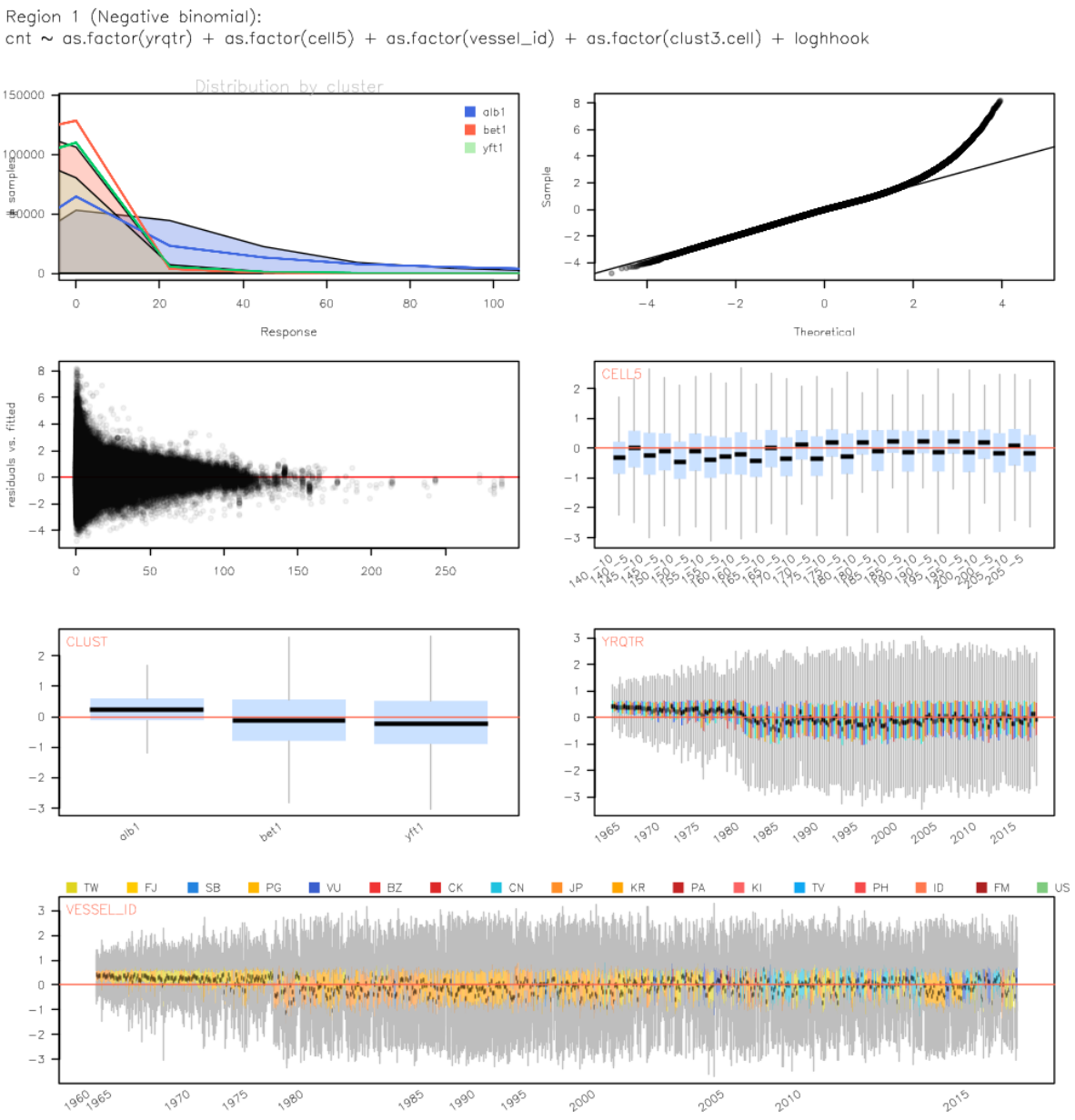

Figure 48: Diagnostic plots of fitted GLM models for region 1 (negative binomial distribution) showing characteristics of the model residuals and comparisons between observed and simulated data.

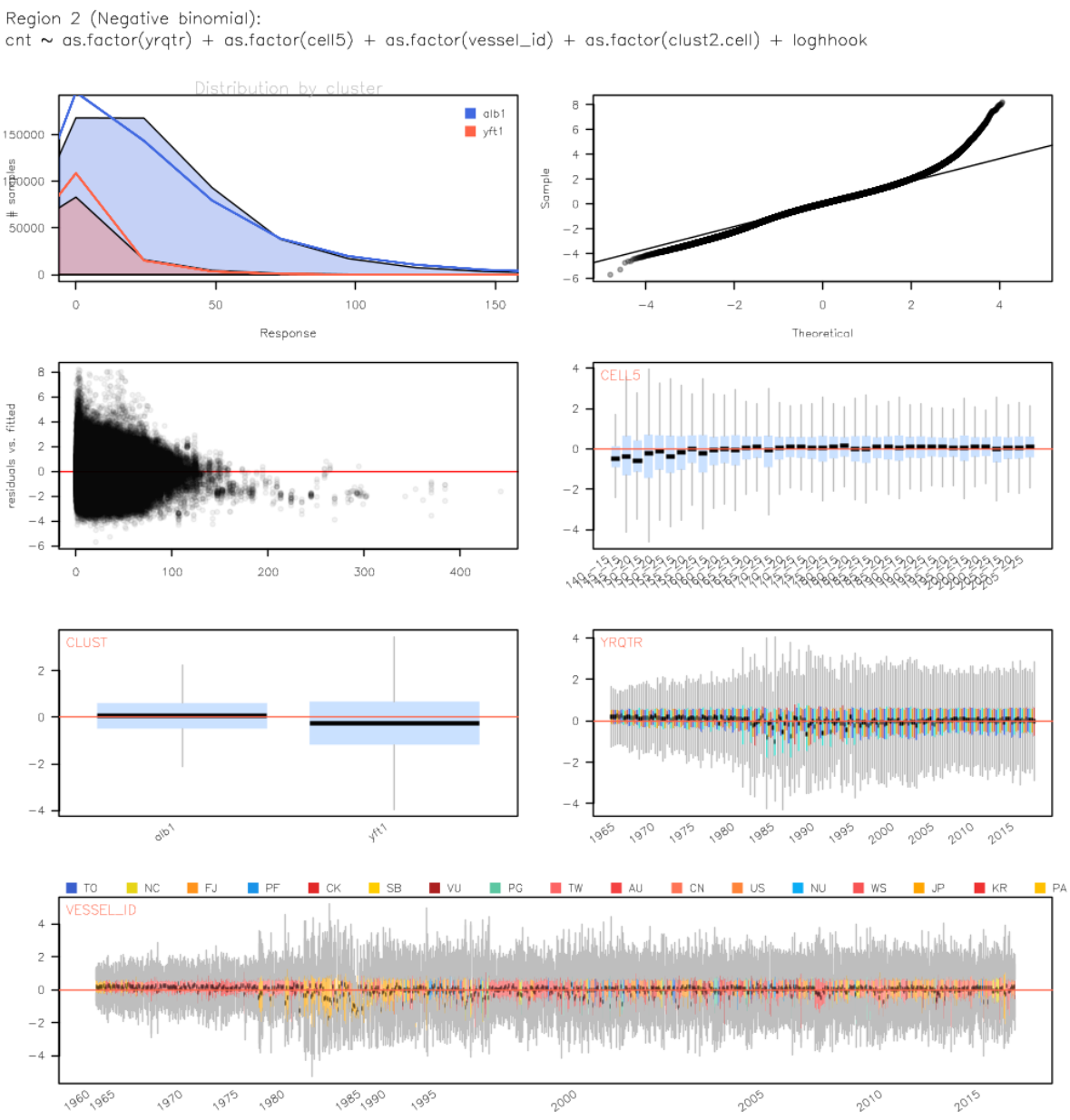

Figure 49: Diagnostic plots of fitted GLM models for region 2 (negative binomial distribution) showing characteristics of the model residuals and comparisons between observed and simulated data.

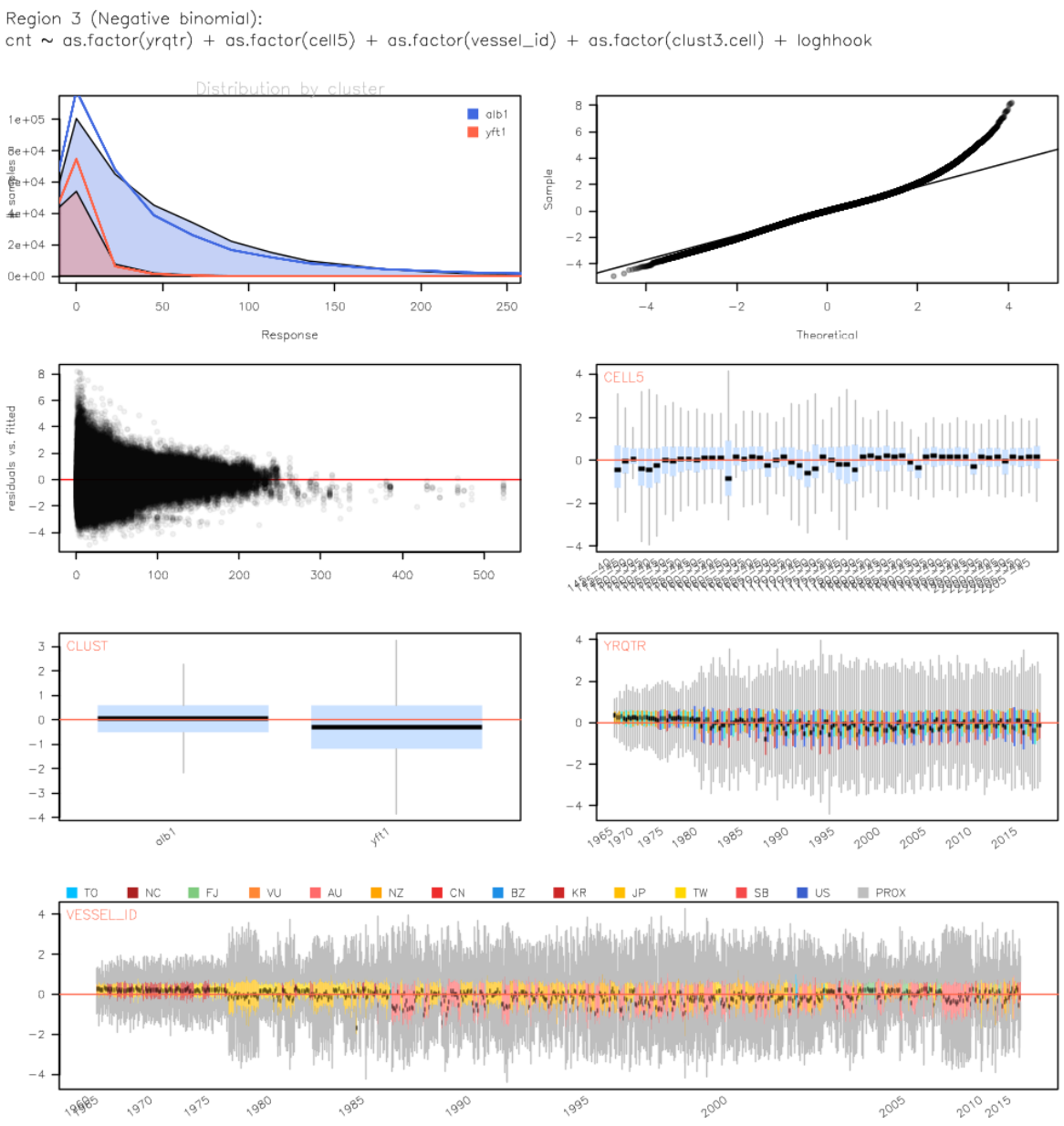

Figure 50: Diagnostic plots of fitted GLM models for region 3 (negative binomial distribution) showing characteristics of the model residuals and comparisons between observed and simulated data.

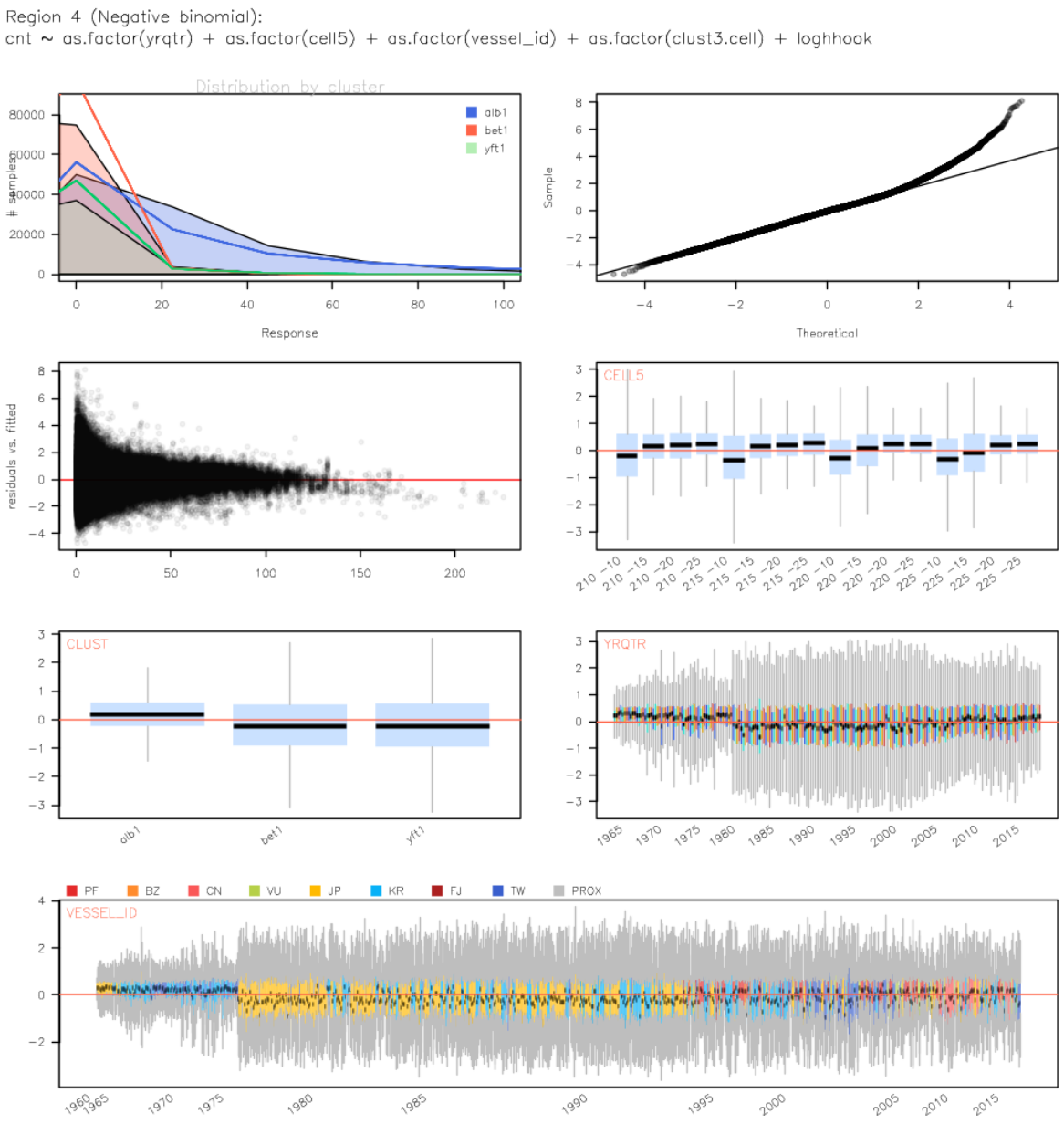

Figure 51: Diagnostic plots of fitted GLM models for region 4 (negative binomial distribution) showing characteristics of the model residuals and comparisons between observed and simulated data.

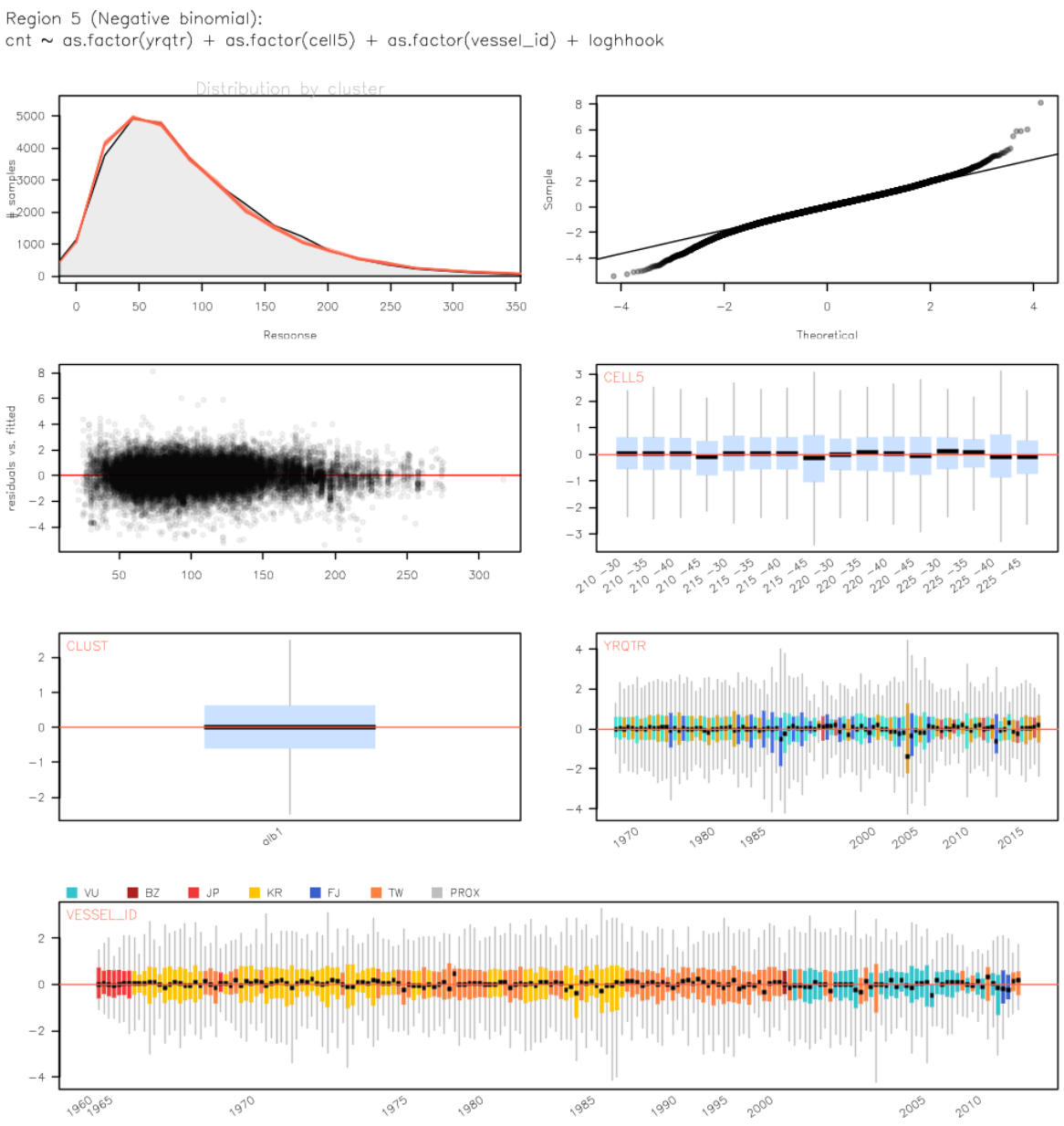

Figure 52: Diagnostic plots of fitted GLM models for region 5 (negative binomial distribution) showing characteristics of the model residuals and comparisons between observed and simulated data.

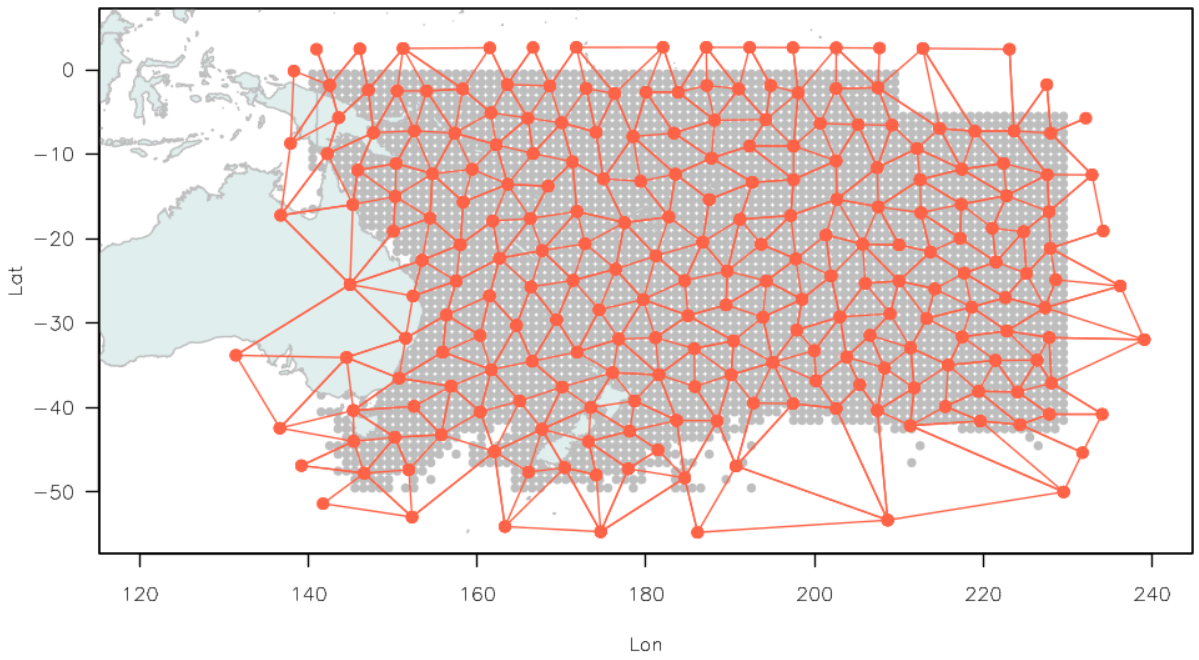

Figure 53: Mesh used to fit the geostatistical model. An effect is estimated for each of the 200 core knots (red).

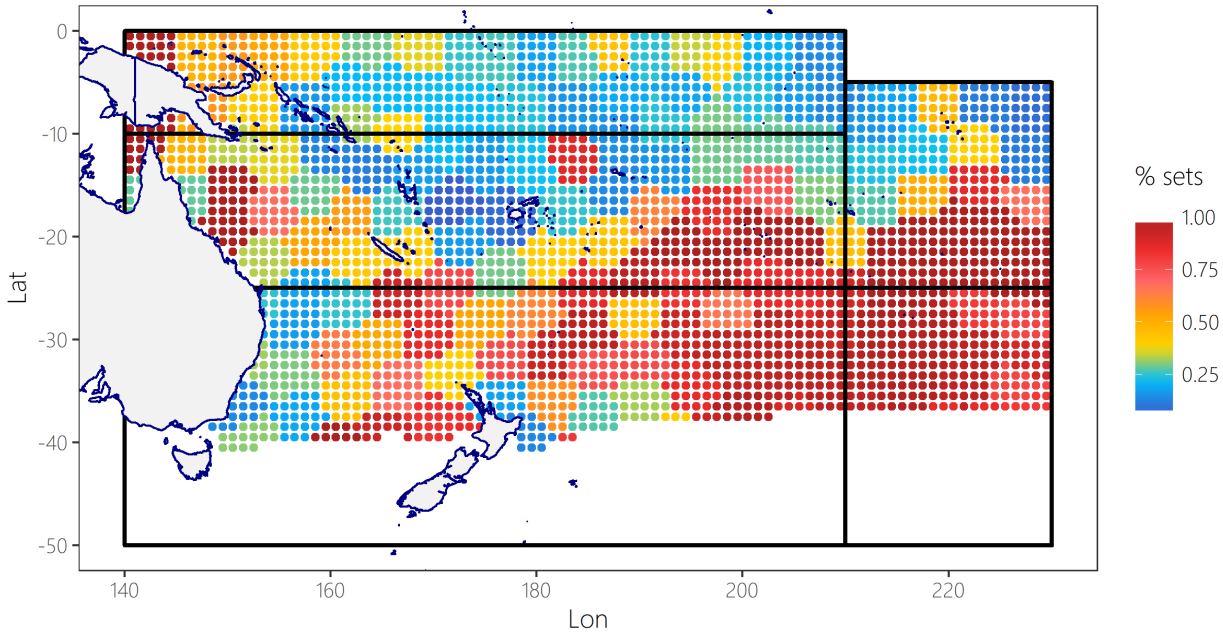

Figure 54: Proportion of sets retained by the subsampling procedure.

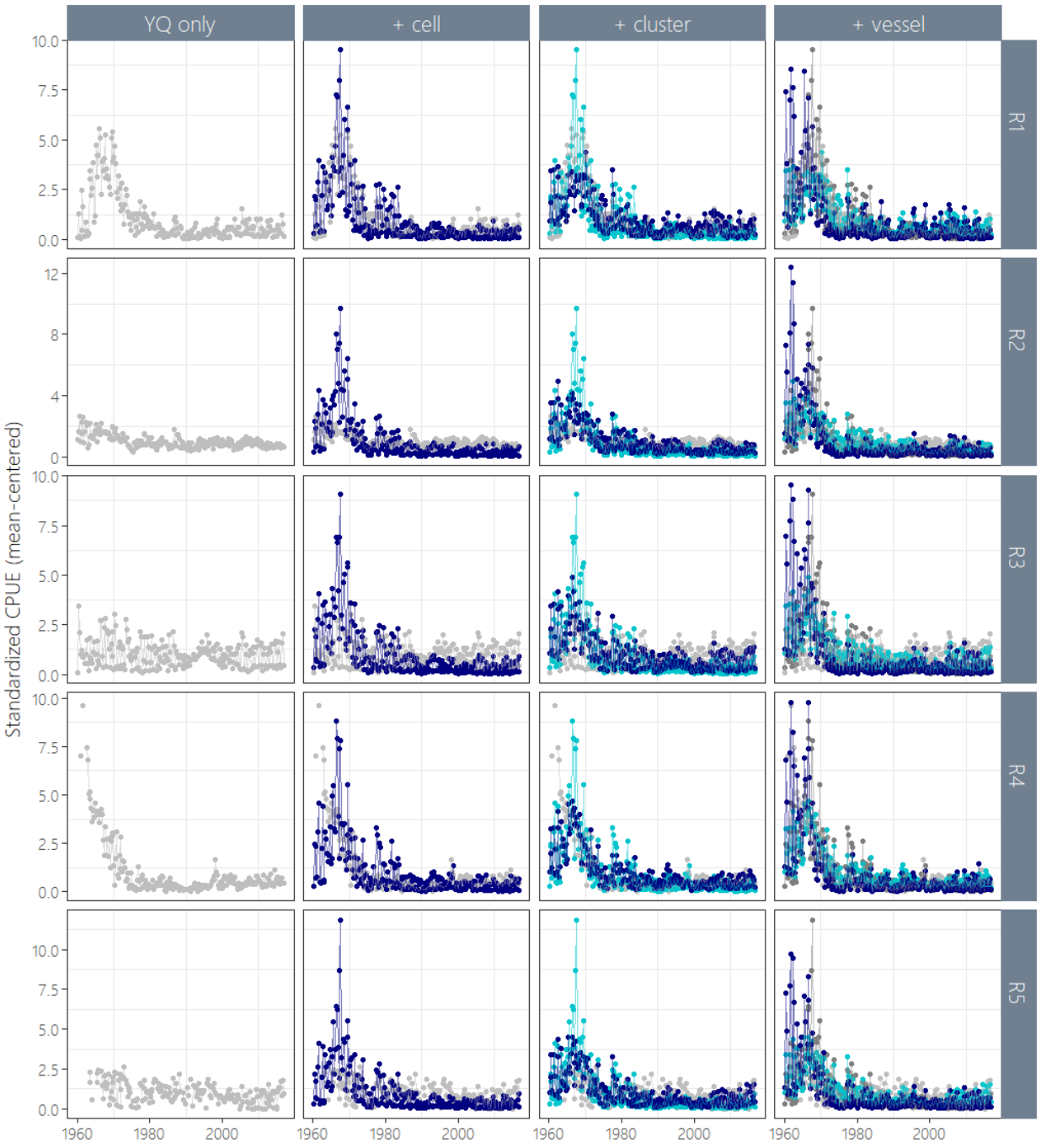

Figure 55: Step plots showing the effects of removing individual factors from the GLM model with respect to the estimated CPUE indices produced by standardisations for the diagnostic case model of ALB using the geostatistical approach. The left hand panel shows the nominal CPUE, the panel second from the left shows the model with the cell factor added (dark blue line; grey line is the year-quarter-only model), the next panel shows the model with the cluster factor added (dark blue line; light blue line is the cell + year-quarter model, etc.) and finally the right panel shows a model with the vessel factor added (dark blue line; the light blue line is the previous model - cluster + cell + year-quarter, etc.)

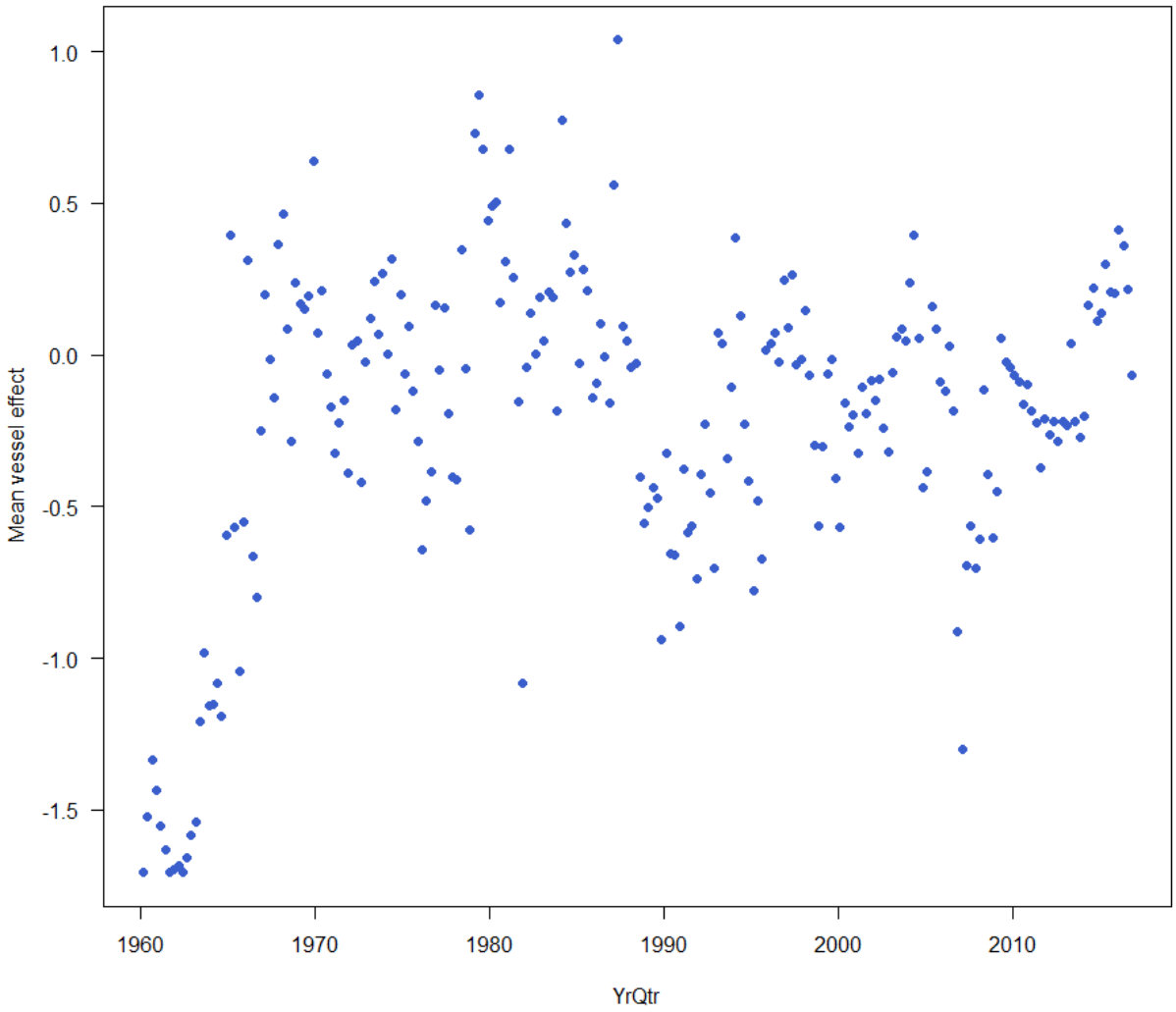

Figure 56: Effort-weighted average vessel effect per year-quarter for the binomial component of the geostatistical model when vessel ID is included as a covariate.

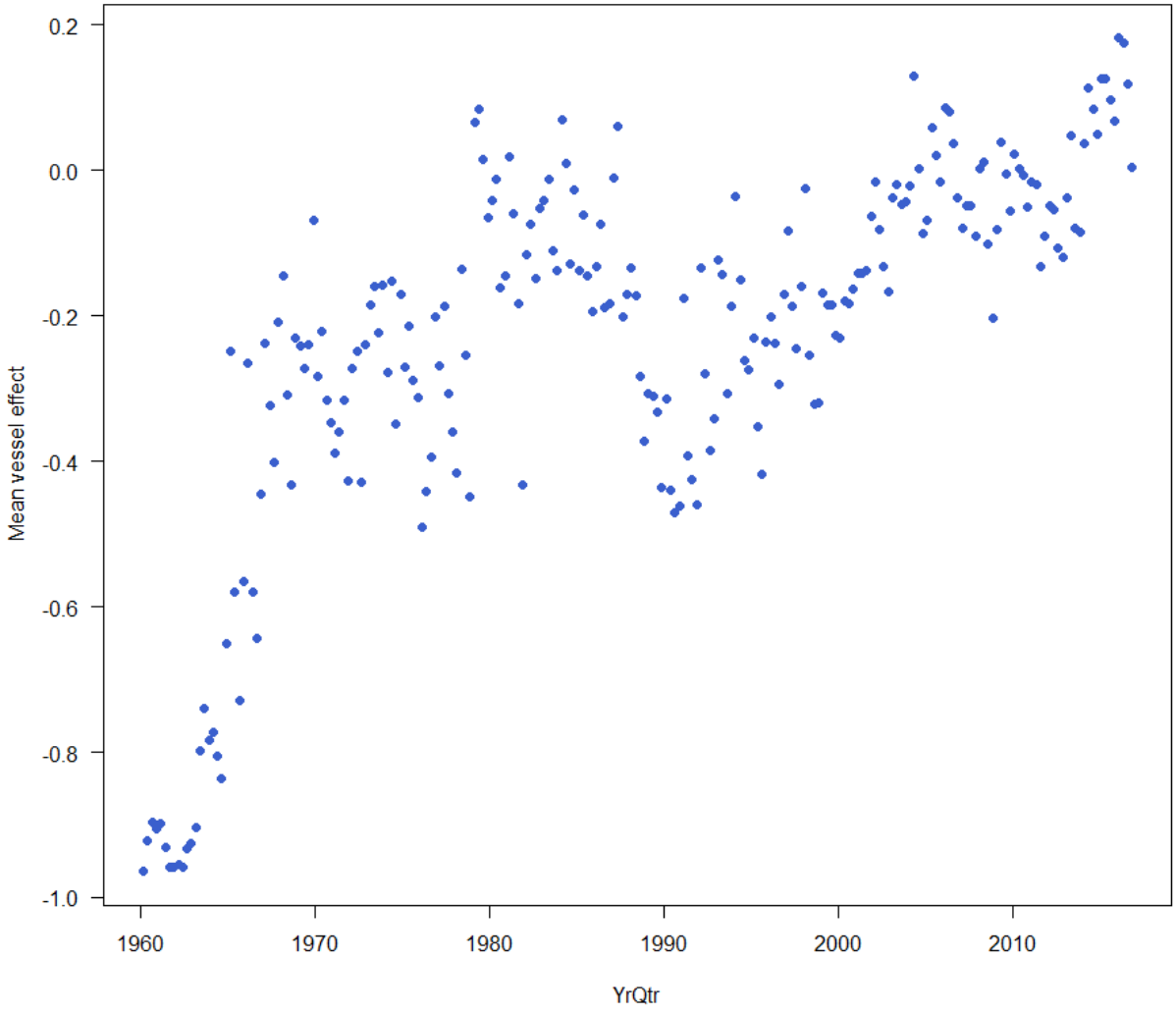

Figure 57: Effort-weighted average vessel effect per year-quarter for the lognormal component of the geostatistical model when vessel ID is included as a covariate.

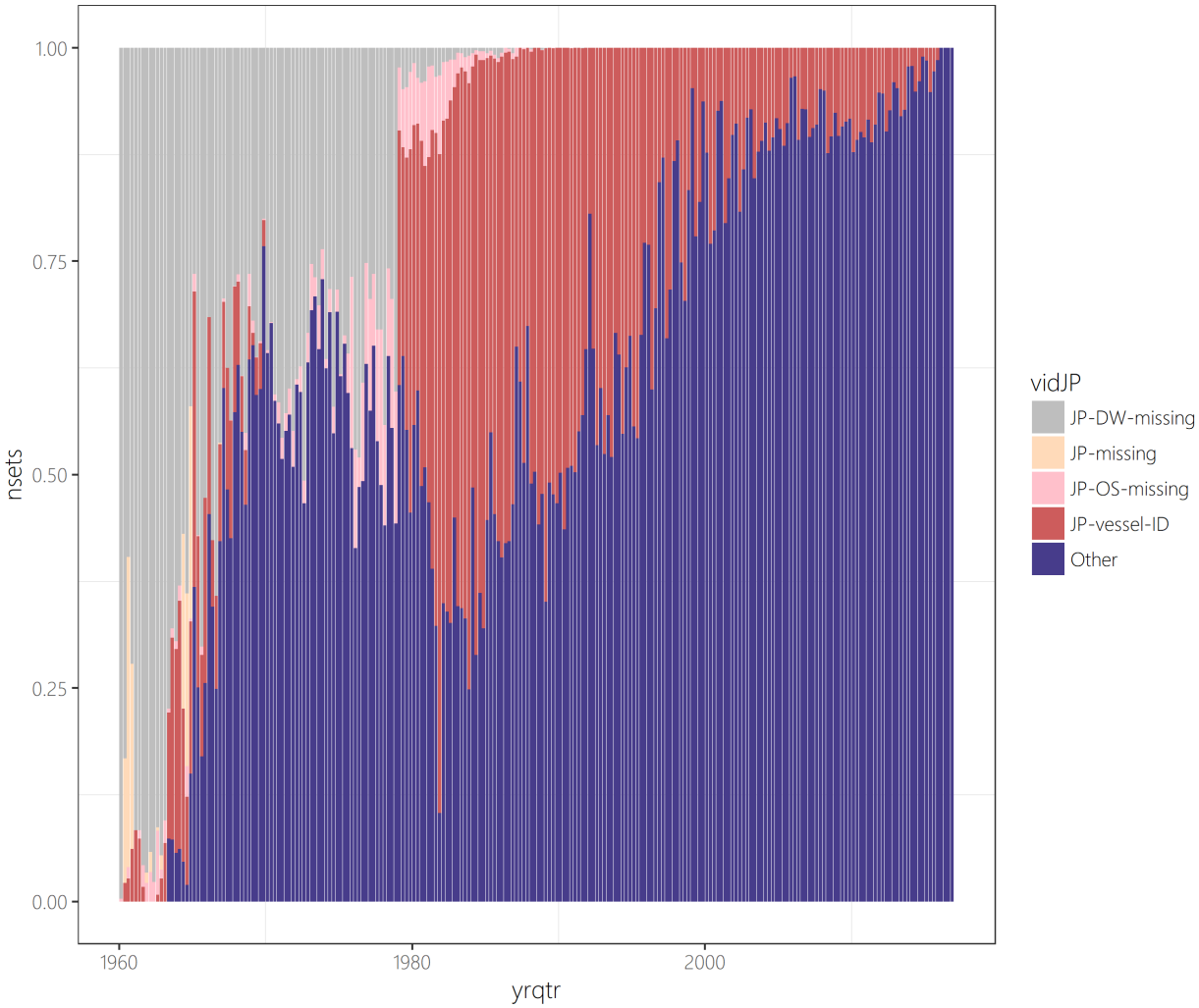

Figure 58: Relative proportion of vessel category per year-quarter.

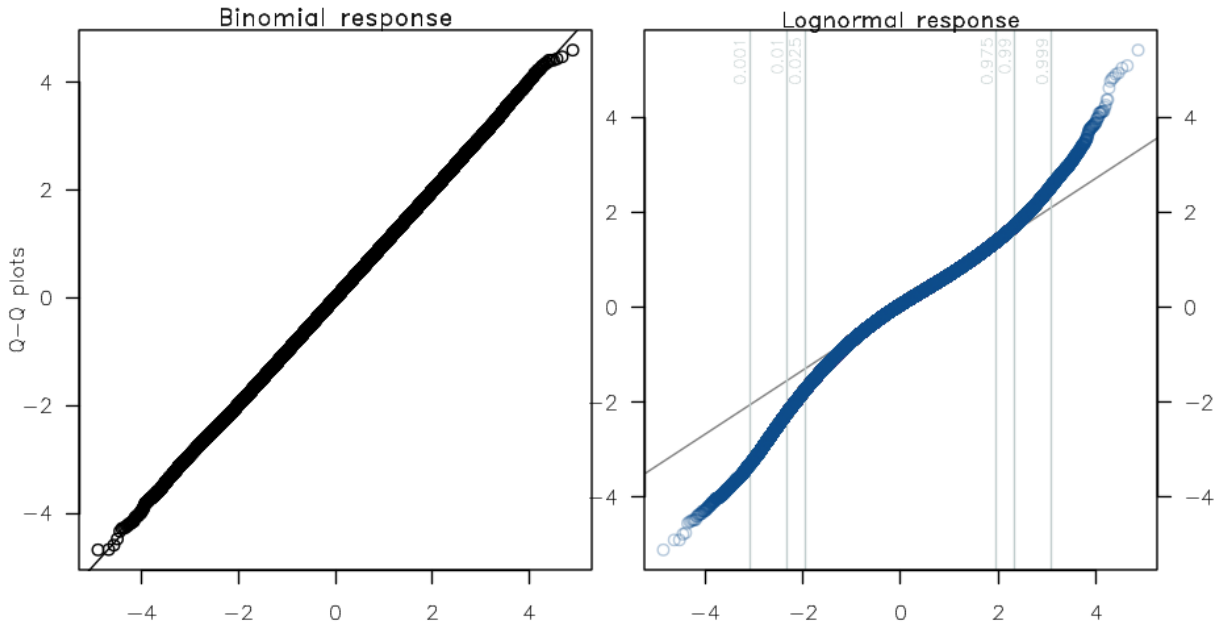

Figure 59: QQ plot for the binomial (left) and the lognormal (right) components of the geostatistical model used to produce final standardized indices by region. The quantiles of the data are indicated with the vertical grey lines.

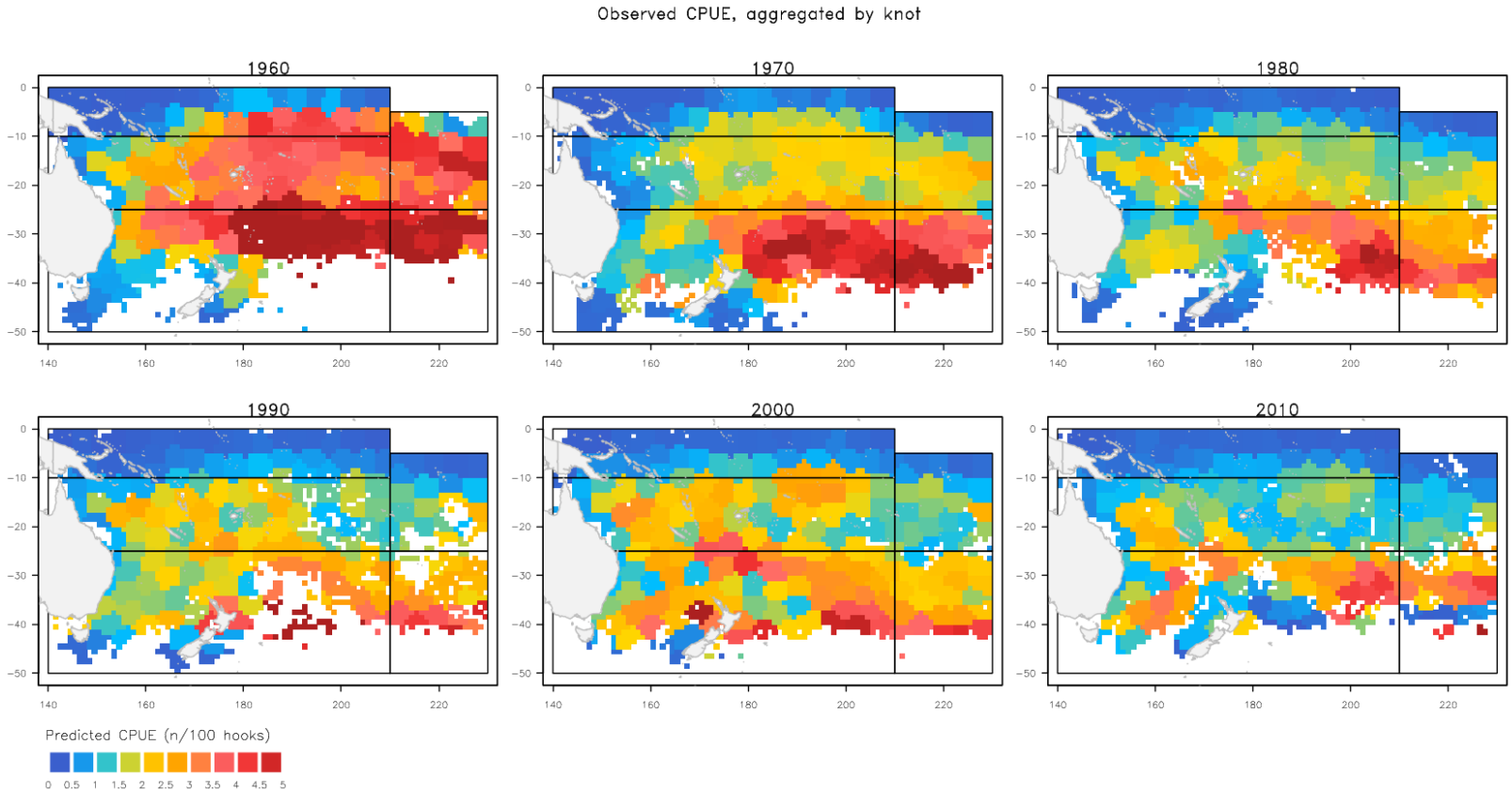

Figure 60: Observed CPUE by decade in individuals per hundred hooks from the operational longline set used to fit the geostatistical model, aggregated by knot but shown at the 1° cell for ease of comparison with the geostatistical predictions.

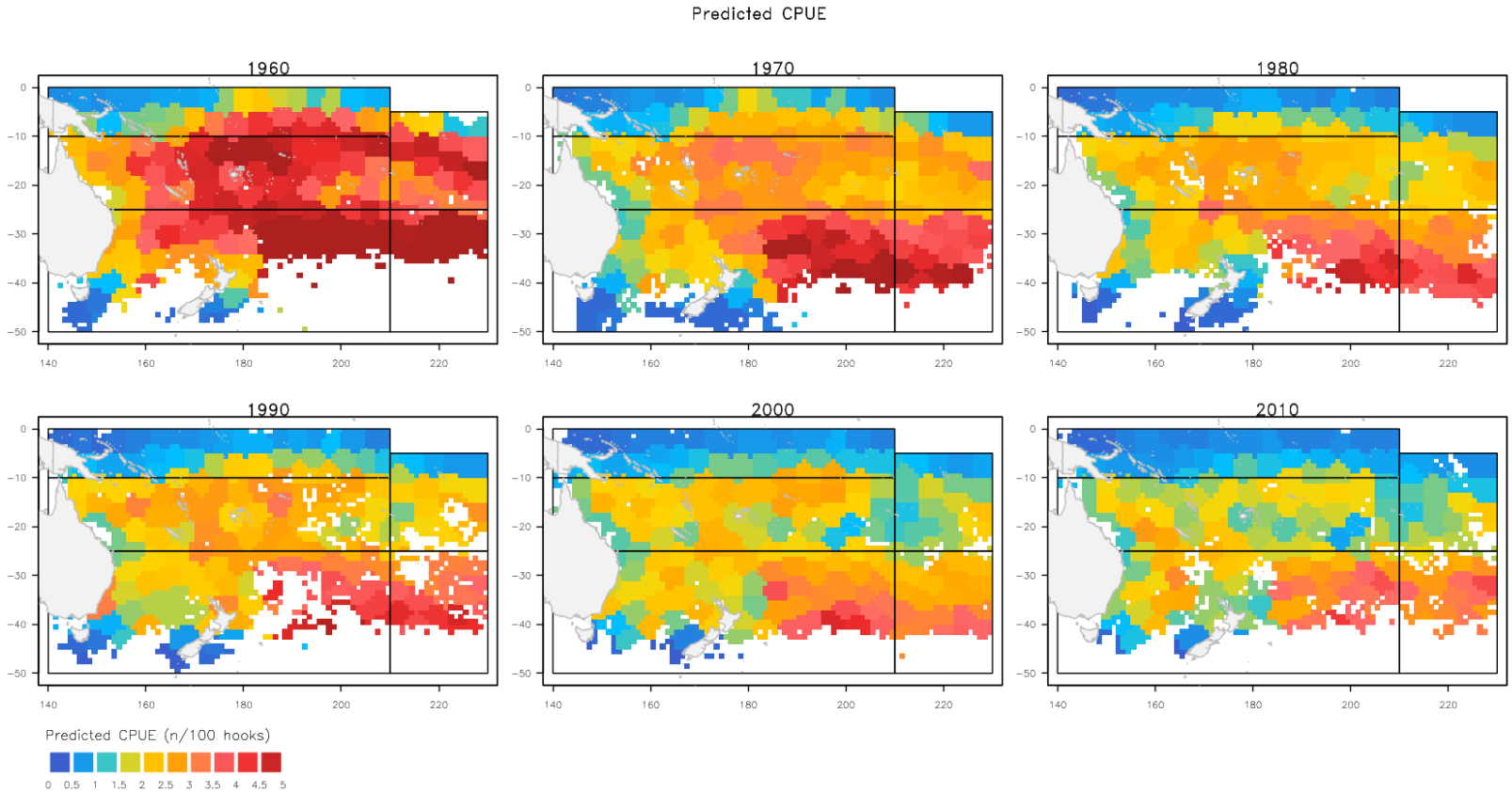

Figure 61: Predicted CPUE by decade in individuals per hundred hooks from the geostatistical model fitted to the operational longline set, shown by 1° cell.

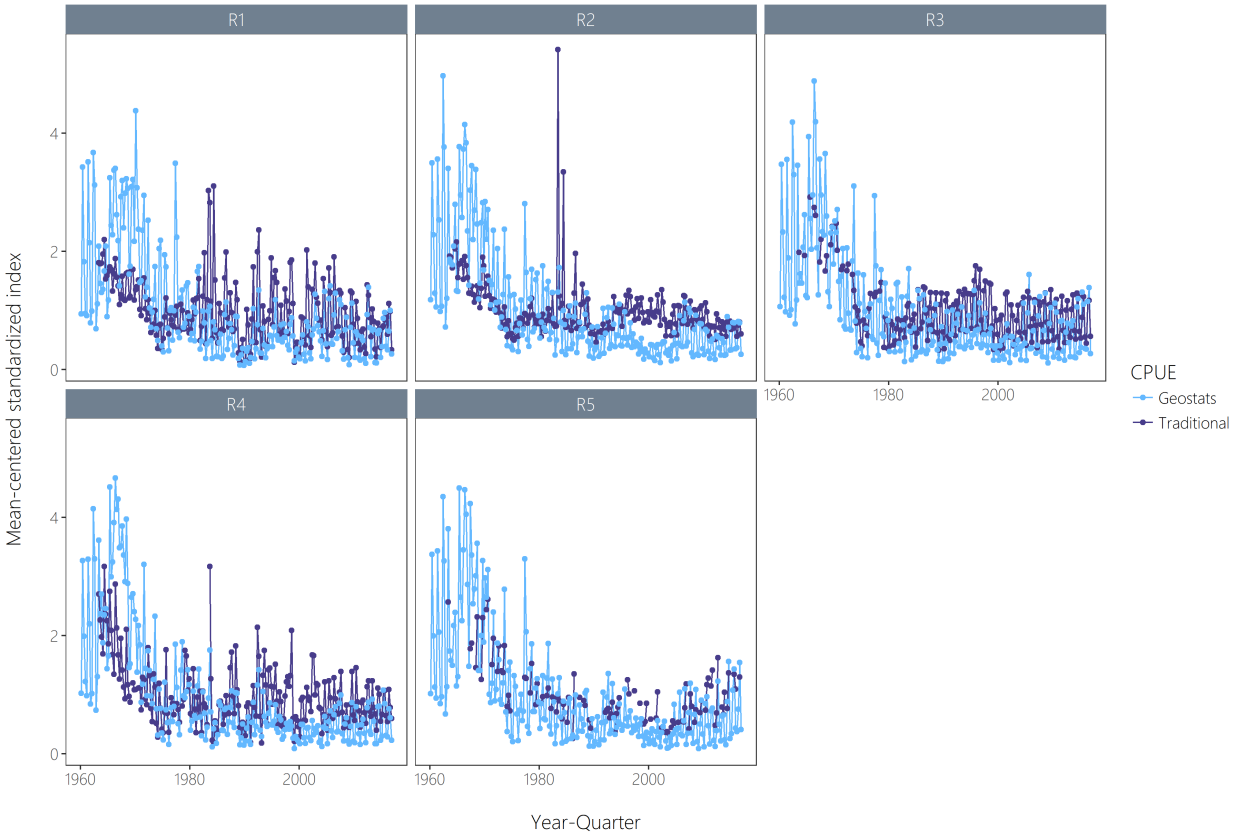

Figure 62: Mean-centered 'traditional' *vs.* geostatistical standardized CPUE indices by year-quarter used in the 2018 South Pacific albacore assessment.

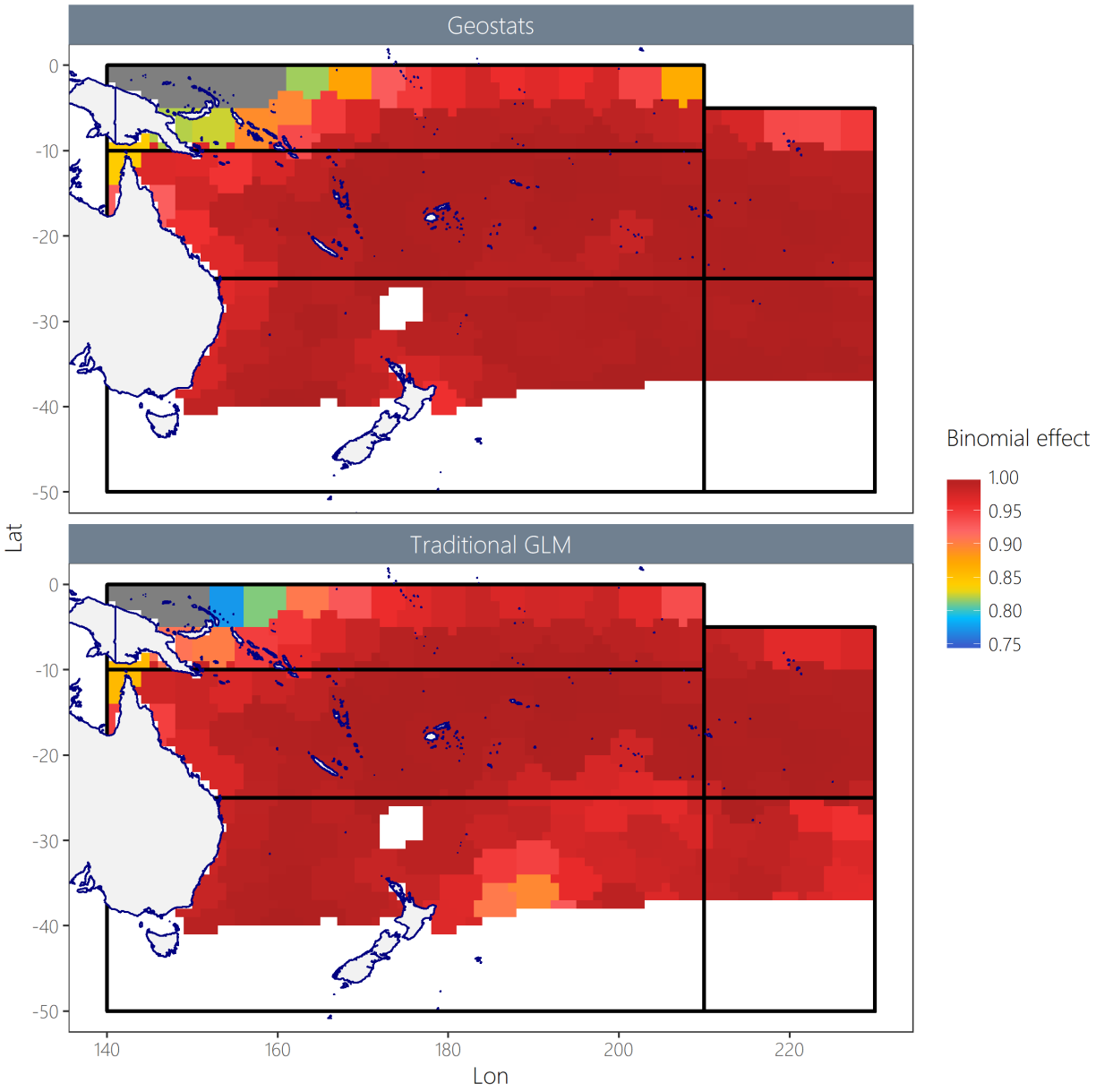

Figure 63: Comparison of the knot effects estimated with categorical levels ('traditional') *vs.* with a geostatistical surface for a binomial model of the probability of positive albacore catch in a given knot. Note the colour scale starts at 0.75 to allow contrast in the shown effects across the regions. Cells in grey had predicted values below 0.75.

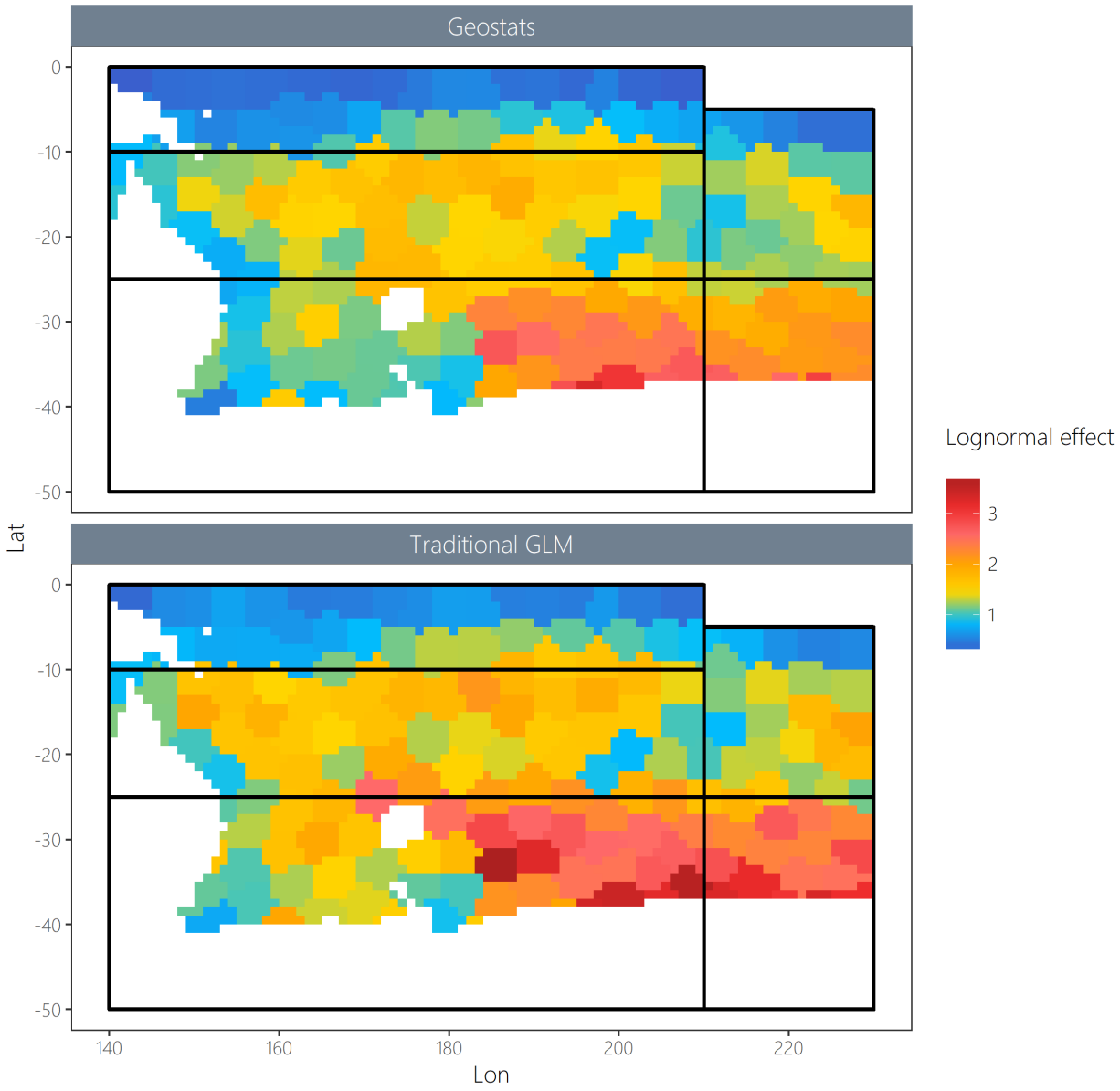

Figure 64: Comparison of the knot effects estimated with categorical levels ('traditional') *vs.* with a geostatistical surface for a positive model of albacore catch rates in a given knot.

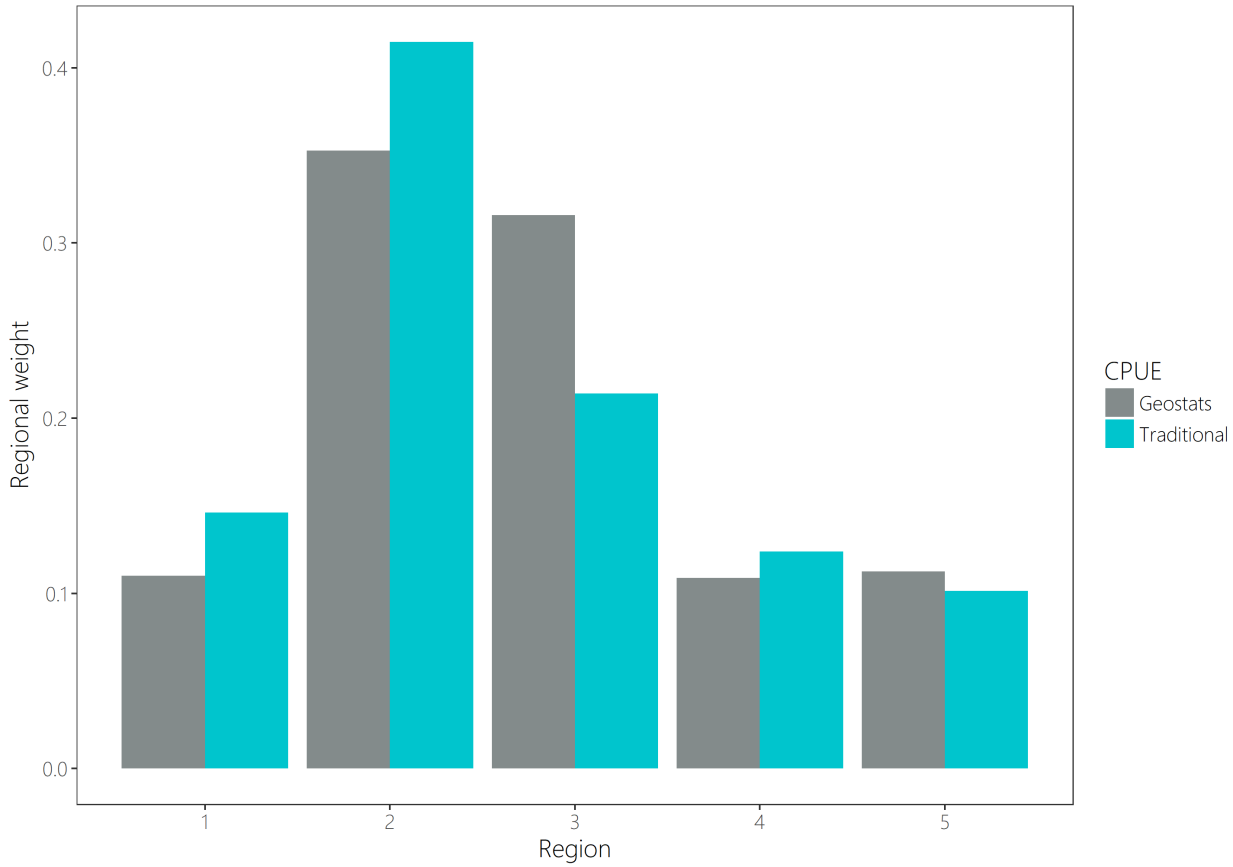

Figure 65: Comparison of the regional weights estimated under the 2015 'traditional' method (described in Tremblay-Boyer et al., 2015a) and the updated weights under the 2018 geostatistical approach. All regional weights sum up to one under both methods.

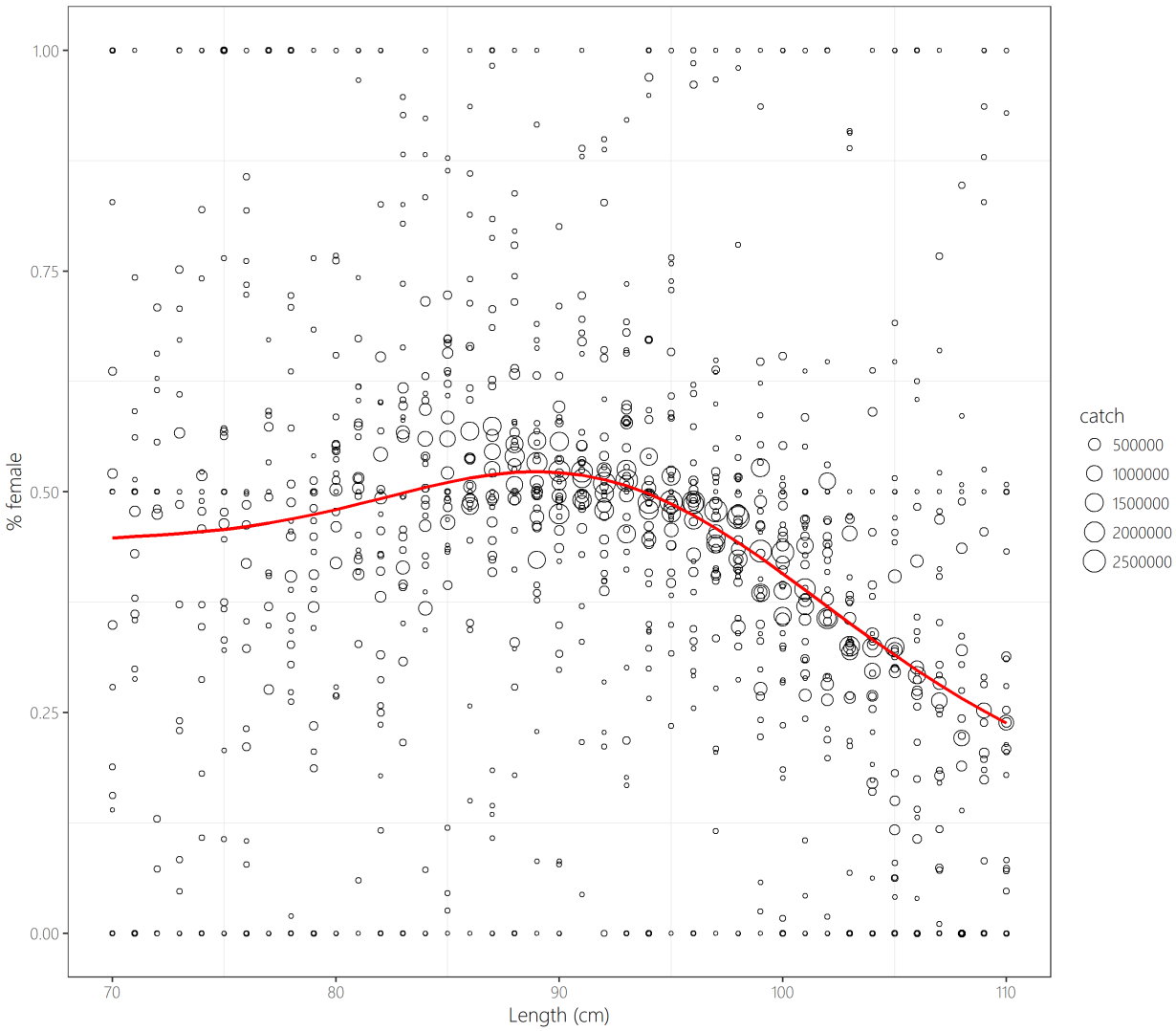

Figure 66: Sex ratio predicted from longline observers' data for the South Pacific.

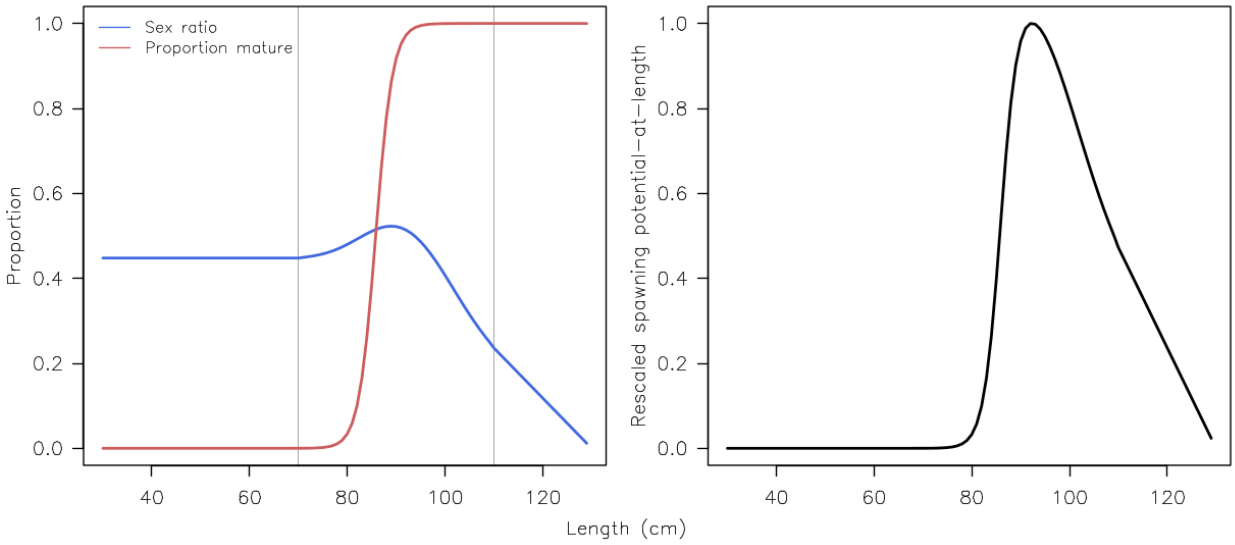

Figure 67: Components of the spawning potential-at-length relationship (left panel: sex-ratio, blue line, and proportion mature-at-length, red line, from Farley et al. (2014), and the resulting spawning-potential-at-length rescaled to spawn 0 to 1.