SCIENTIFIC COMMITTEE
TWENTIETH REGULAR SESSION

Manila, Philippines
14 – 21 August 2024

Analysis of Longline Size Frequency Data for the 2024 Southwest Pacific Striped Marlin and South Pacific Albacore Assessments

WCPFC-SC20-2024/SA-IP-03
25 July 2024

Joanne Potts[1], Claudio Castilo-Jordan [1], Jemery Day[1], Paul Hamer[1], Thomas Teears[1]

[1] Oceanic Fisheries Programme (OFP), Pacific Community (SPC), Nouméa, New Caledonia

Executive Summary

The primary objective of this analysis was to aggregate standardised eye orbital to fork length frequency (LF) measurements for longline caught south Pacific albacore (ALB) and striped marlin (MLS) across the southern WCPFC convention area. This analysis was undertaken to identify broad spatial regions exhibiting homogenous LF characteristics. Identification of such regions could be valuable for informing the spatial structure of the stock assessments for south Pacific ALB and southwestern Pacific MLS.

The R package 'FishFreqTree' (v. 3.3.2, Xu and Lennert-Cody, 2023) was used to fit spatial regression trees to analyse these LF data. The analyses were conducted in R (v. 4.4.0, R Core Team 2024). A range of one to ten regression tree splits (i.e., 2 to 11 regions) were explored. Percentage variance explained and visual inspection of LF histograms within each region were used to identify notable LF histogram features (e.g., bimodality, skewness, etc.), that suggest further splits might be required to ensure homogeneity of LF within each region.

The results highlighted significant challenges in identifying reliable regions for MLS due to the temporal and spatial sparsity of the data, with many grid cells lacking sufficient observations to accurately reflect underlying LF distributions. Data between 2011 and 2022 were included for the analysis, a period when spatial coverage was relatively even. Attempts to delineate regions yielded low percentage variance explained, with only 12% variance explained with as many as ten splits (i.e., 11 regions). Consequently, the results for MLS were deemed unreliable for robust inferences on spatial structure that might be useful for the stock assessment.

Conversely, the ALB data exhibited better temporal and spatial coverage, and a longer period was analysed (1990 to 2022). In addition to the full data set, different combinations of flags were subset and analysed (i.e., PICT, DWFN and AU+NZ). Regardless of flag-combination, smaller fish were consistently identified in the southern region, with the first split typically occurring between 25°S and 30°S, explaining 10-15% of the variance in the data. Subsequent splits improved the percentage variance explained, more than doubling within four to six splits to over 20%. Although compared to the MLS analysis, percentage variance explained was higher, it remained overall low (up to 30%, depending on the number of splits). While these results offer some insight into the spatial structure of ALB size composition in the regions longline catches, they may not provide sufficient evidence to warrant a revision of the current ALB stock assessment structure.

Aim

Length-frequency (LF) measurements sampled from longline catches (standardised eye orbital to fork measurements) were provided for the two species assessed in 2024 (South Pacific albacore (ALB) and striped marlin (MLS)) at a 5-degree spatial resolution across the south of the WCPFC convention area between 1976 and 2024. The aim was to aggregate these LF distributions to identify broad spatial regions that exhibit similar characteristics. Identifying such regions could be useful for informing the spatial structure of the south Pacific ALB and southwestern Pacific MLS stock assessments.

Methods

Rules for Data Inclusion

Only longline fishing data were analysed. Some very small (minimum length: MLS = 9 cm, ALB = 30 cm) and very large (maximum length: MLS = 400 cm, ALB = 140 cm) fish were recorded in the data. Although these may possibly be legitimate records, a decision was made to omit fish with lengths < 60 cm (n = 30) and > 300 cm (n = 28) for MLS and < 50 cm (n = 1,459) and > 300 cm (n = 285) for ALB.

Temporal and spatial coverage was sparse over the entire data set, particularly for the early periods. Consequently, for MLS the retained data was for the period 2011-2022, and for ALB was 1990-2022. Spatial coverage was highly variable depending on flag. Most flags had little to no contribution (e.g., Guam (GU) had 13 records, and Marshall Islands (MH) had 180 records for MLS). For MLS, Japan (JP) and Taiwan (TW) had the most comprehensive spatial and temporal coverage (5,378 and 6,915 records, respectively) and therefore were used in the final analysis. Data from all flags were retained for ALB (but see Methods).

Since the analysis requires a spatially continuous grid, observations with longitude < 140 were omitted, as were grid cells with few LF records (i.e., if a grid cell had fewer than 10 or 50 LF records it was excluded for MLS and ALB, respectively).

Data analysis

The methods applied follow Lennert-Cody et al. (2010), Lennert-Cody et al. (2013), Xu and Lennert-Cody (2023) and Hamer et al. (2023). That is, the R package 'FishFreqTree' (v. 3.3.2) was used to fit regression trees to the spatial LF data. LF data were binned into 5 cm length classes. The influence of Year and Quarter were explored, but since the resulting regions were not spatially contiguous, the results are not presented here. Between 1 and 10 regression tree splits (i.e., 2 and 11 regions) were explored. The percentage variance explained for each split was calculated, and the corresponding LF histograms within each region were visually inspected for 'features' (i.e., the presence of bimodality, skewness, and other features that might suggest further splitting is required). All analyses were undertaken in R (v. 4.4.0, R Core Team 2024).

For MLS, data from only JP and TW flags were used as these two flags (1) had the most comprehensive spatial and temporal coverage (see 'Rules for Data Inclusion' for reasoning) and (2) are the two flags used in the CPUE analysis for the current assessment (Castillo-Jordan et al. 2024). Results for one to five splits are presented.

For ALB, results for one to five splits are presented for the following spatial LF analyses:

(1) ALL, i.e., all data from all flags;

(2) DWFN, distance water fishing nations included data from Belize (BZ), China (CN), Chinese Taipei (TW), France (FR), Japan (JP), Panama (PA), Republic of Korea (KR), United States of America (US), and Vanuatu (VU);

(3) PICT, included data from Cook Islands (CK), Fiji (FJ), Federated States of Micronesia (FM), Indonesia (ID), Kiribati (KI), New Caledonia (NC), Niue (NU), French Polynesia (PF), Papua New Guinea (PG), Philippines (PH), Solomon Islands (SB), Tonga (TO), Tuvalu (TV) and Samoa (WS), and

(4) AUNZ, included data from Australia (AU) and New Zealand (NZ) only.

MLS

Background

The current spatial structure of the MLS stock assessment is shown in Figure 1. The LF of JP and TW records for each of these four cells is shown in Figure 2. There is evidence of 'heaping' in all spatial regions, and region 4 appears to be bimodal (i.e., evidence the LF measurements within this spatial region aren't aggregated well).

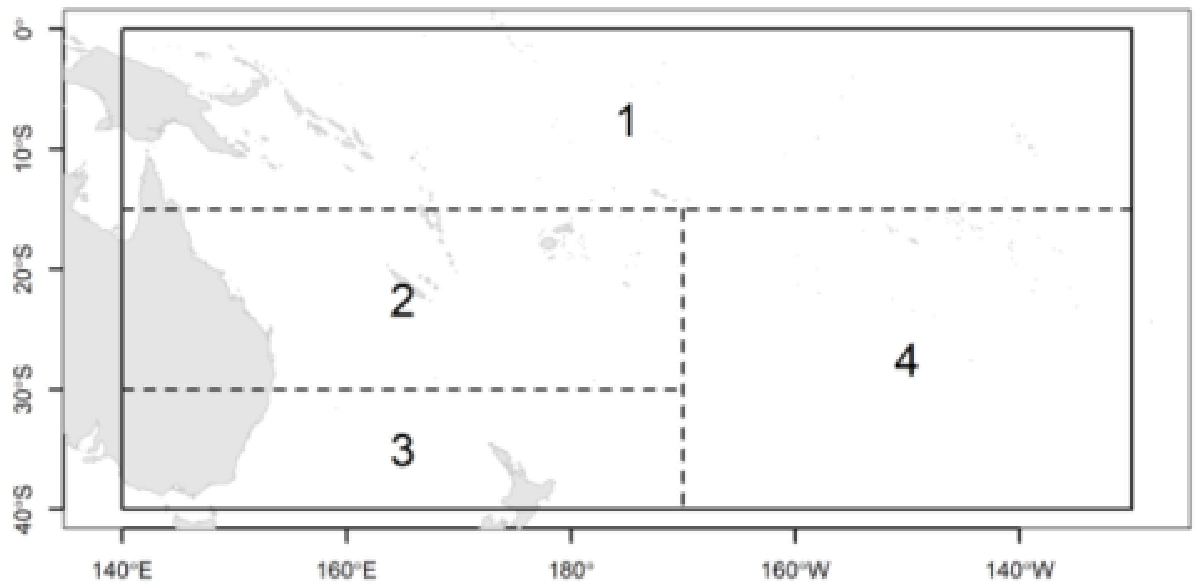


Figure 1. Spatial structure of the previous and proposed current stock assessment for MLS. Note this represents a spatial structure for defining fleets as areas (dashed lines), however the model is a single region model (solid line) with no movement.

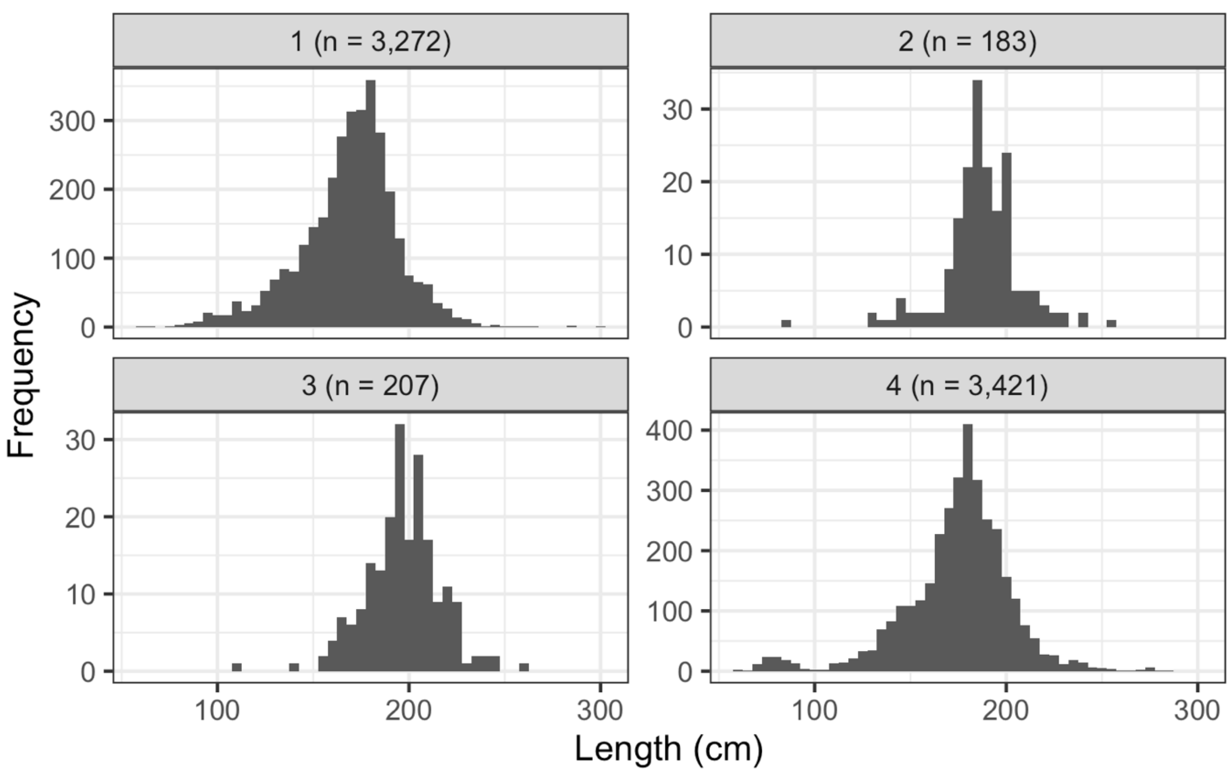


Figure 2. Histogram of JP and TW eye orbital to fork LF (records for each of the spatial (fleets) areas proposed in the current MLS stock assessment.

Results: JP and TW only

Spatial location of JP and TW LF records for each year are shown Figure 3. The length distributions and the mean length throughout the region are shown in Figure 4 and Figure 5, respectively. Although attempts can be made to delineate regions (see Figure 6 and Figure 7), the percentage variance explained was exceptionally low (less than 12% even with as many as 10 splits, Figure 8), rendering the results unreliable for making robust decisions on spatial fleet structure. In particular, the sample size of LF measurements in each grid cell was typically low (minimum size was set to N=10, Figure 4) and their exclusion was influential on the extent of the input data and resulting splits. Due to temporal and spatial sparsity in the data, meaningful identification of regions for this species was not feasible.

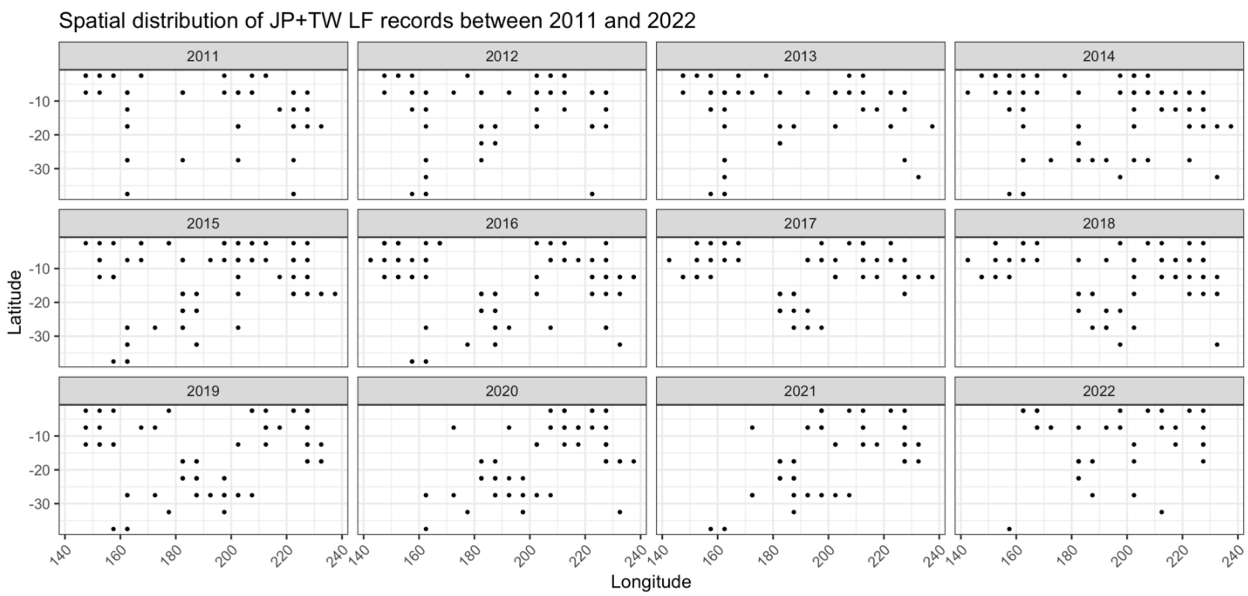


Figure 3. Spatial location of JP and TW eye orbital to fork LF MLS records, by year.

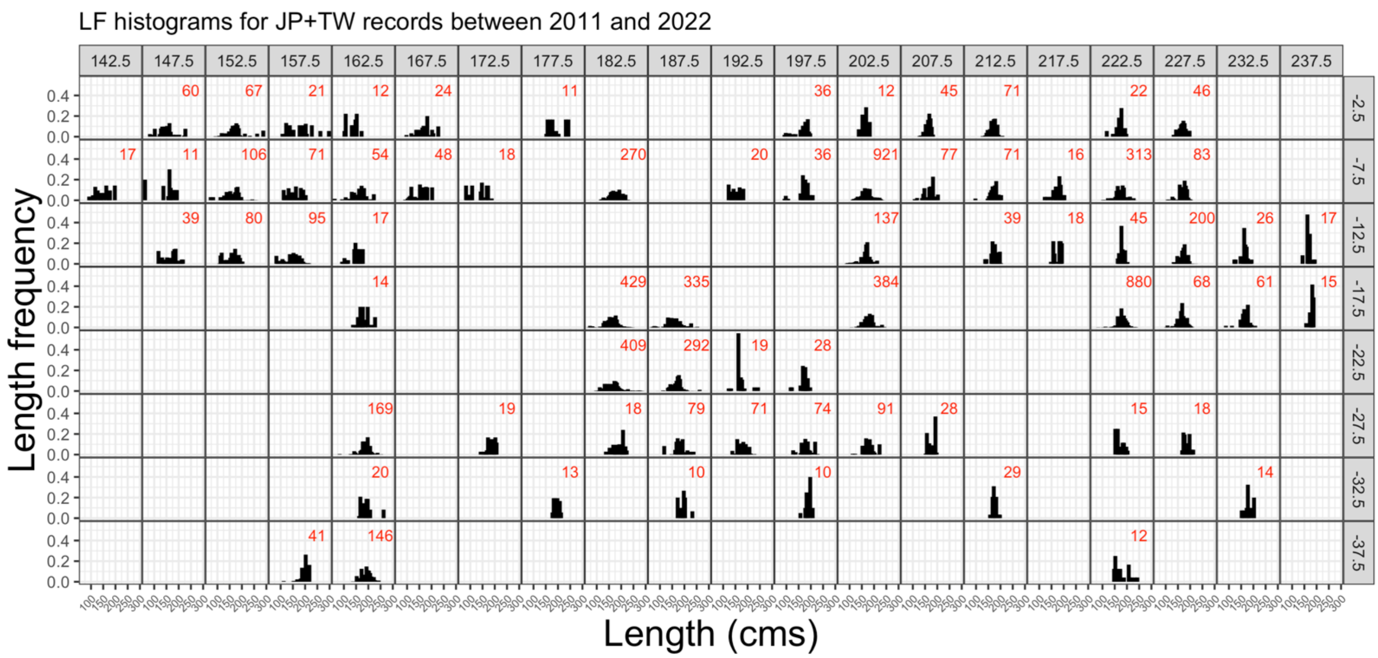


Figure 4. LF histograms for MLS, with sample size included in top righthand corner (red).

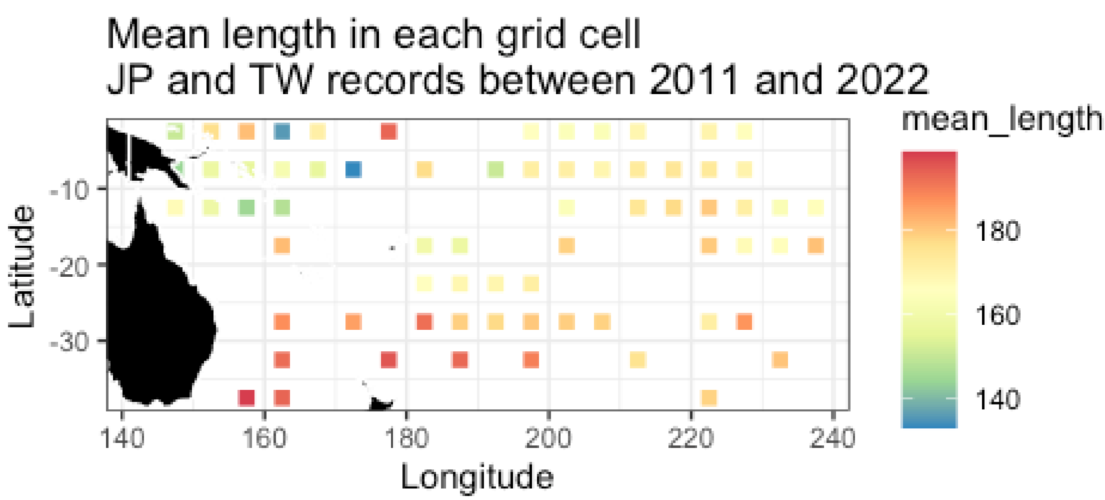


Figure 5. Mean length of MLS caught between 2011 and 2022 on JP and TW flags.

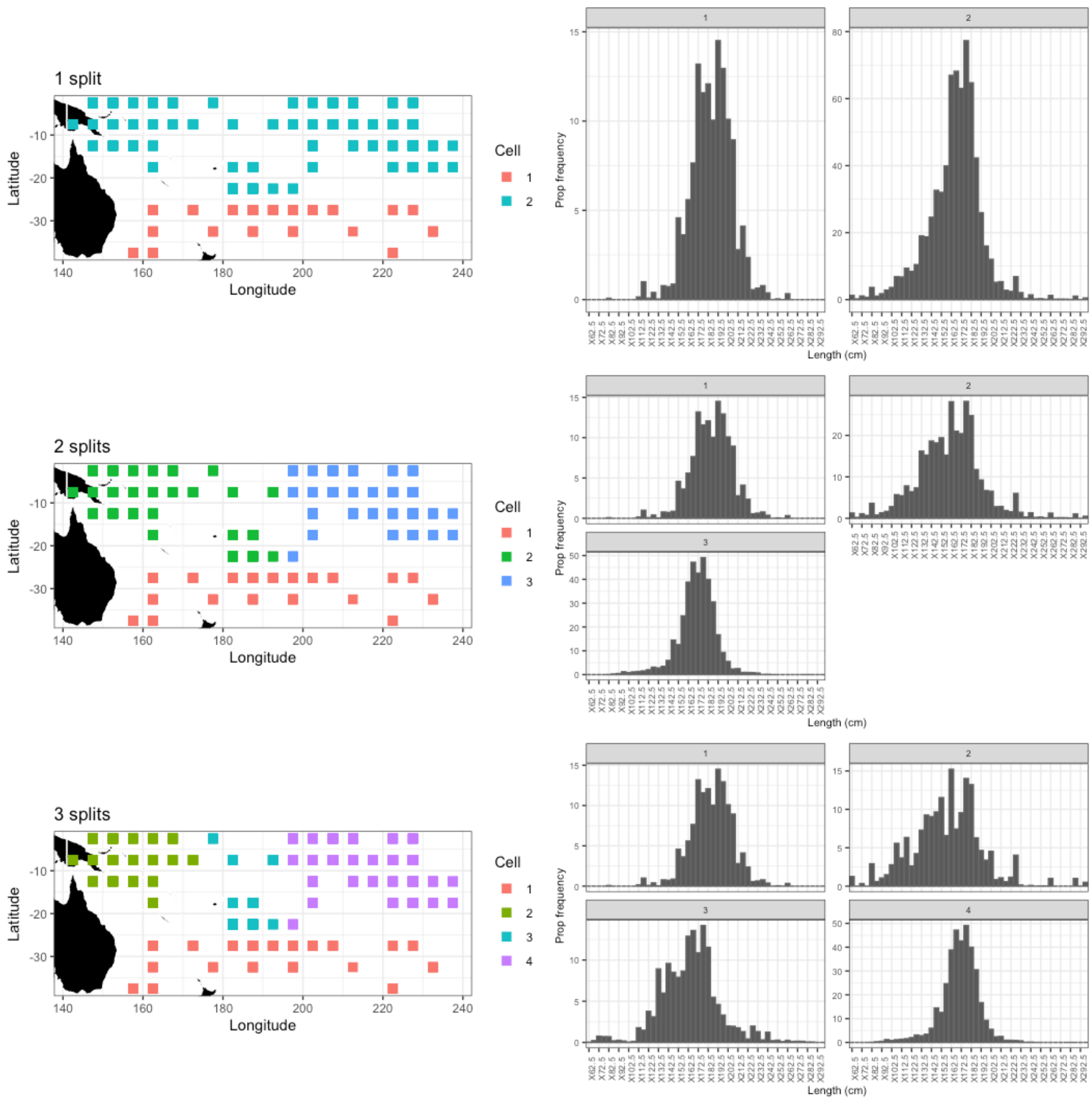


Figure 6. Spatial LF results for MLS using 1 (first row), 2 (middle row) and 3 (third row) splits (i.e., 2-, 3- and 4-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 7 for the results of 4 and 5 splits.

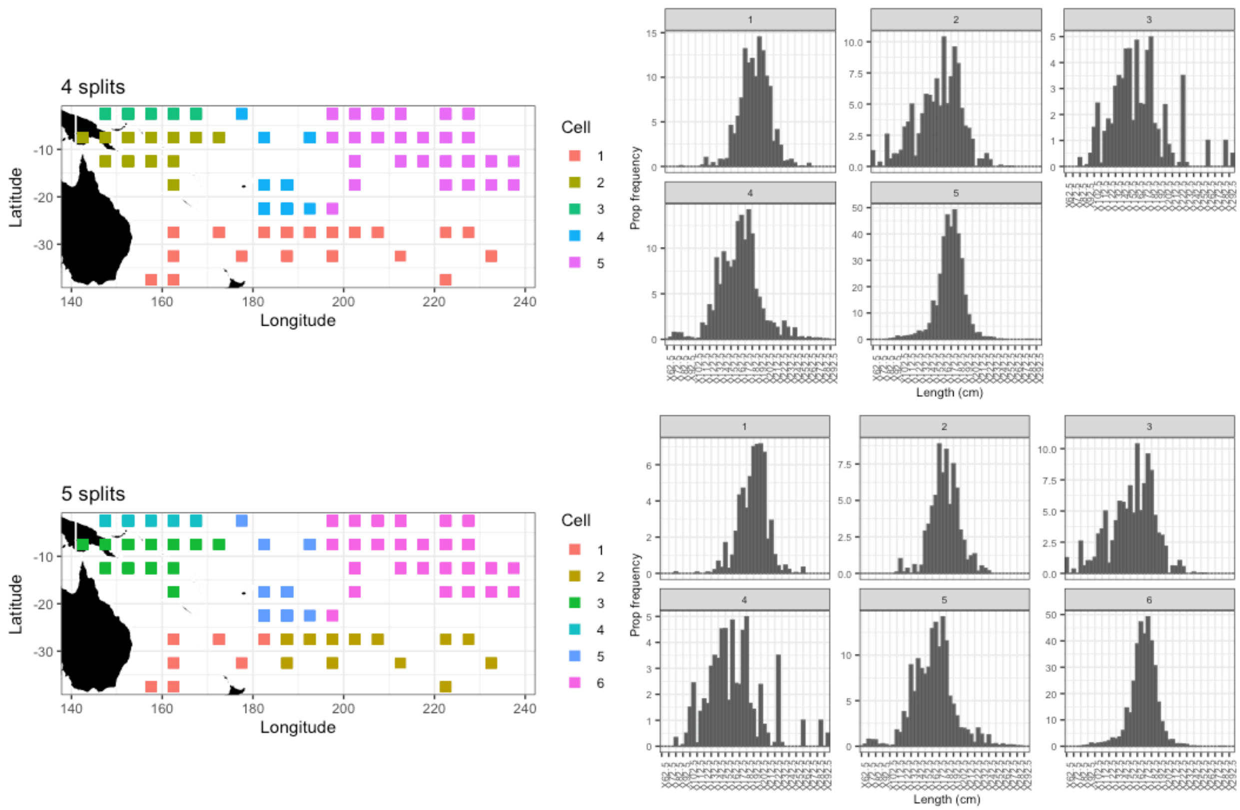


Figure 7. Spatial LF results for MLS using 4 (first row) and 5 (last row) splits (i.e., 5- and 6-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 6 for the results of between 1 and 3 splits.

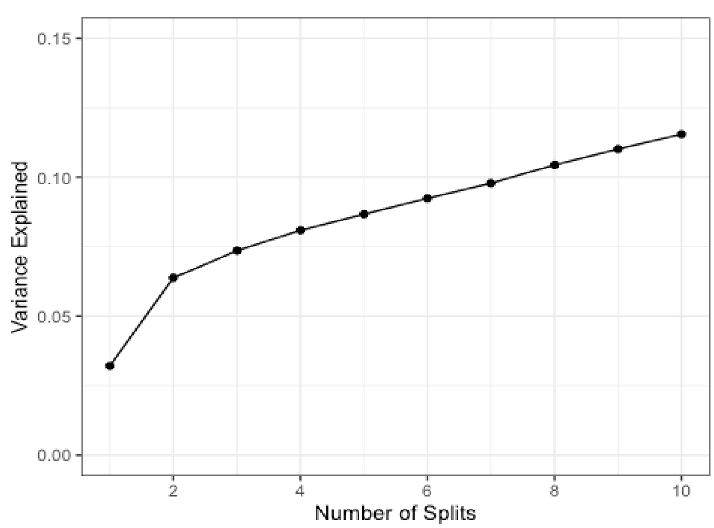


Figure 8. Percentage variance explained with increasing the number of splits in the spatial LF analysis for MLS.

ALB

Background

The current spatial structure of the ALB stock assessment is shown in Figure 9. The LF of all records for each of the six areas in region 1 are shown in

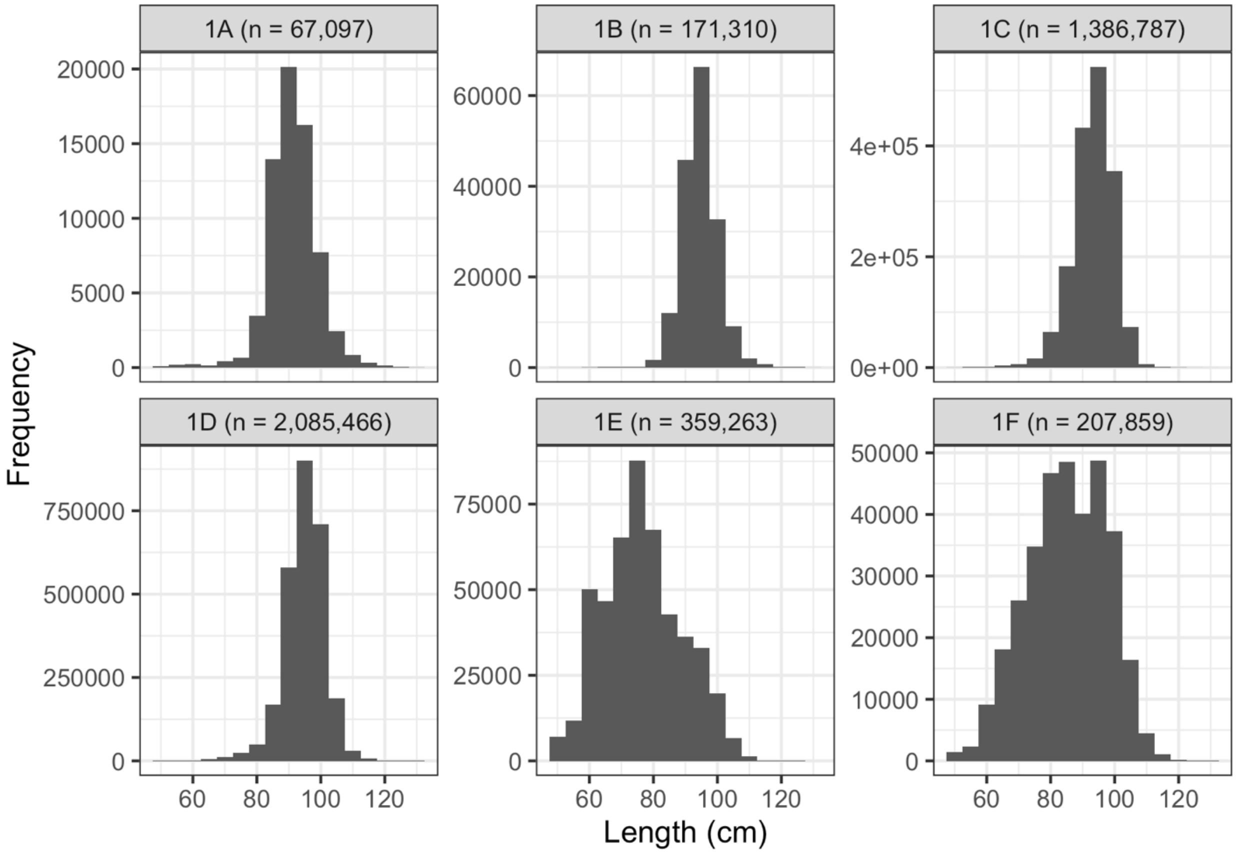


Figure 10. There is evidence of 'heaping' in all spatial regions, and regions 1E and 1F appear to be non-normally distributed (skewed, i.e., evidence the LF measurements within these spatial regions aren't aggregated well).

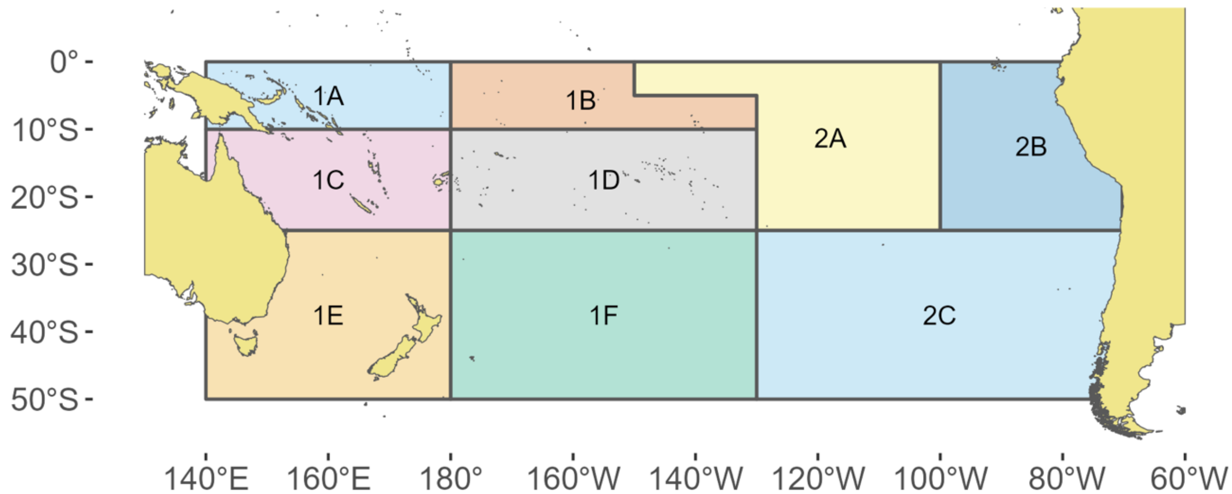


Figure 9. Spatial structure of the current stock assessment for ALB. Regions 1 and 2 indicate separate model regions, the letters i.e. 1A etc., indicate fleet areas within the regions. Region 1 represents the WCPFC including the overlap region with IATTC, region 2 represents the IATTC region (or EPO).

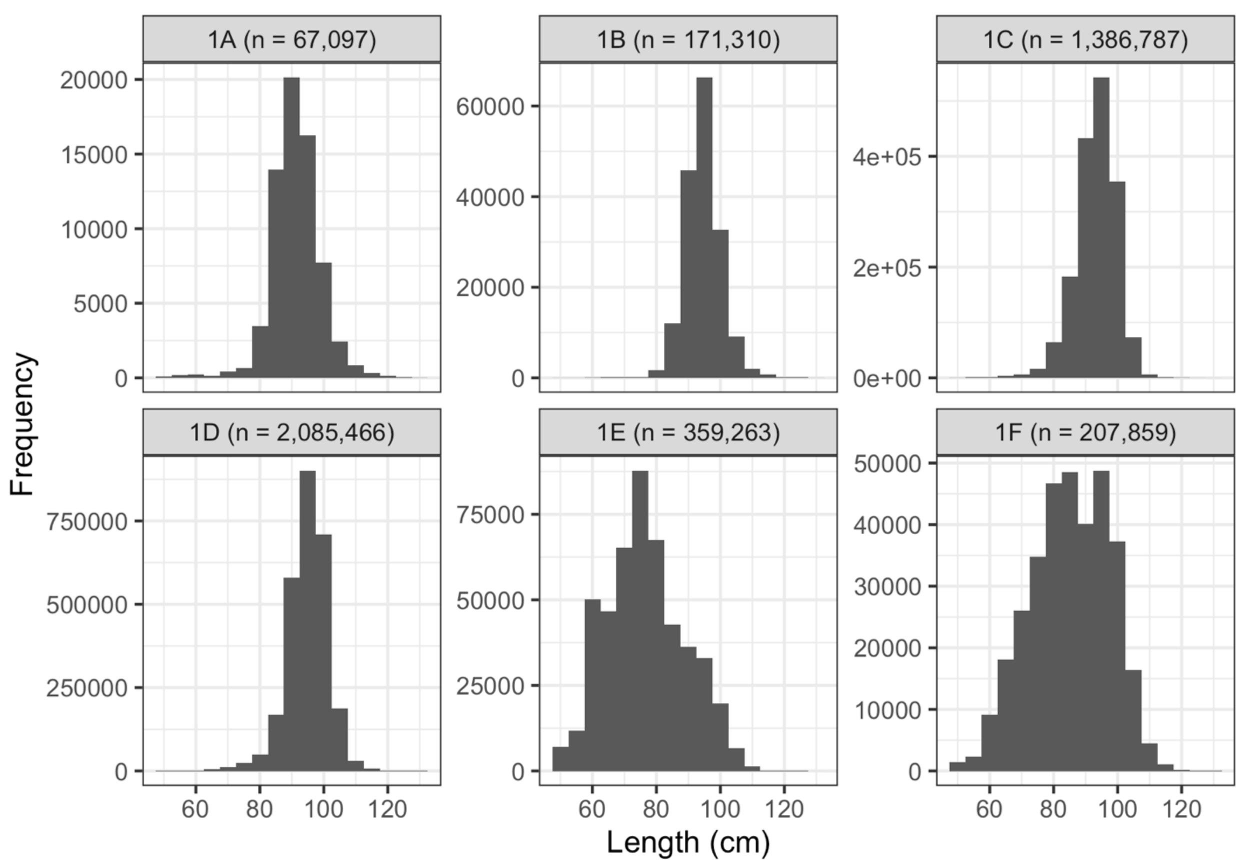


Figure 10. Histogram of all ALB eye orbital to fork LF records for each of the fishing areas proposed in region 1 of the current ALB stock assessment.

Results: All data

The spatial and temporal coverage of all ALB LF records between 1990 and 2022 was extensive (Figure 11). The length distributions and the mean length throughout the region are shown in Figure 12 and Figure 13A, respectively. The number of length measurements in each grid cell was large, most grid cells had many thousand LF records (Figure 12). Larger length fish were observed in the northern part of the region (Figure 13A). The first split delineated this region at 30°S (Figure 14) and this split explained over 15% of the variance in the data (Figure 16). Further spatial delineations are shown in Figure 14 and Figure 15. The percentage variance explained increased relatively quickly up to four splits (nearly 25% variance explained), but plateaued thereafter and at 10 splits (i.e., 11 regions), with the percentage variance explained remaining just under 28% (Figure 16).

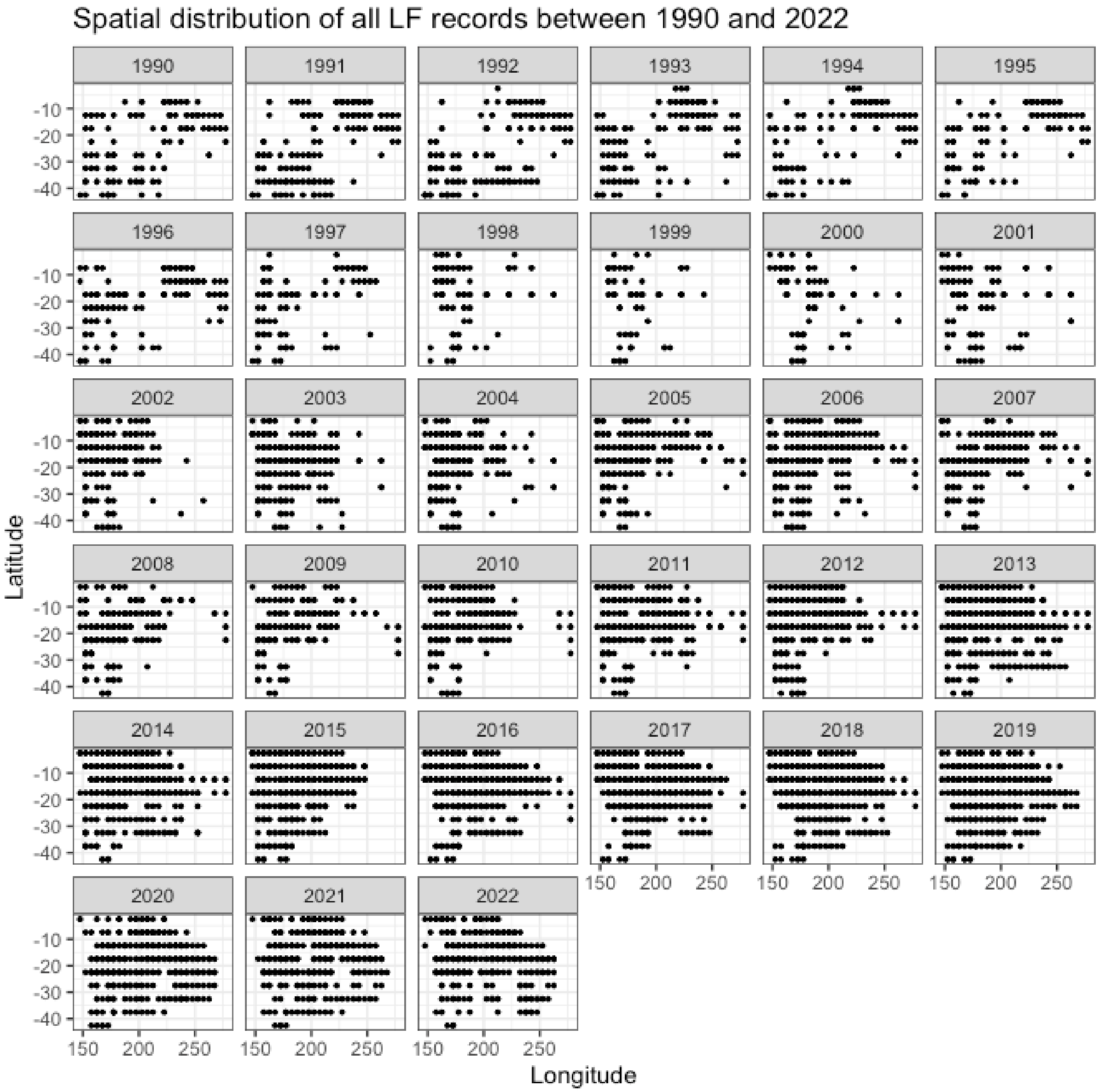


Figure 11. Spatial location of all LF ALB records, by year.

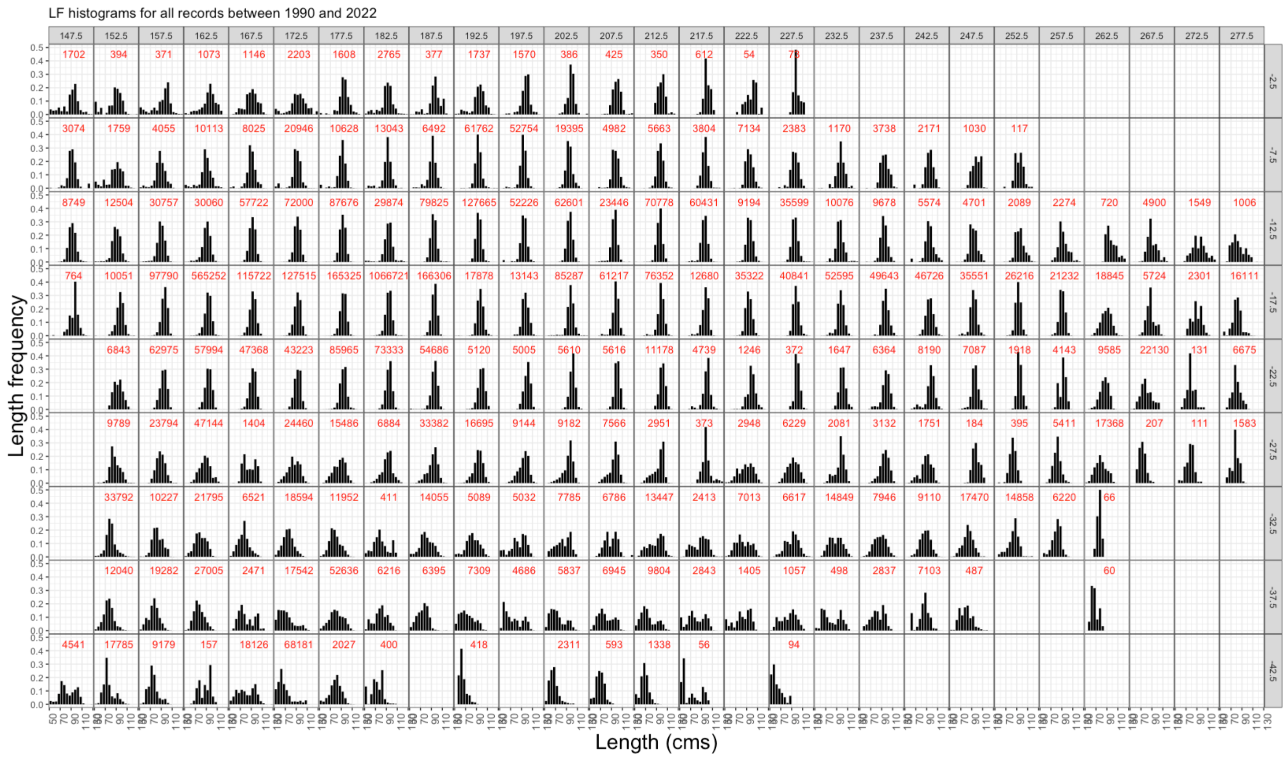
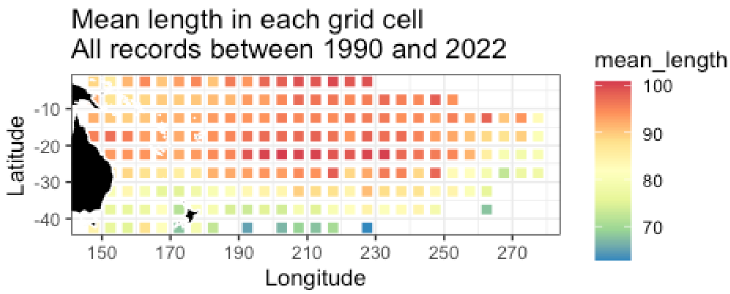
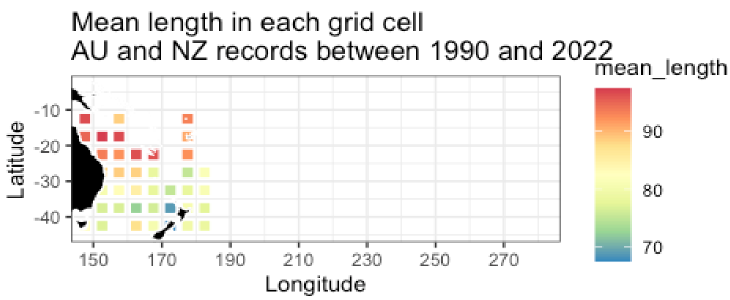


Figure 12. LF histograms for all ALB, with sample size included at top of grid cell in red.

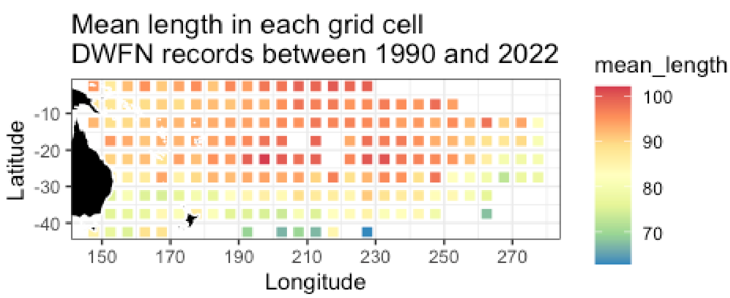
(a) All data



(b) AU and NZ only



(c) DWFN only



(d) PICT only

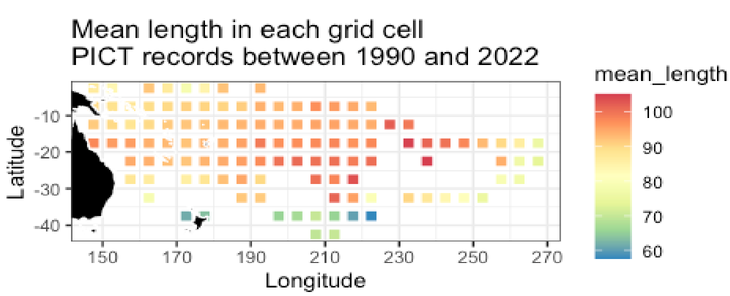


Figure 13. Mean length of all ALB caught between 1990 and 2022 for (a) all flags, (b) AU and NZ flags only, (c) DWFN only and (d) PICT only.

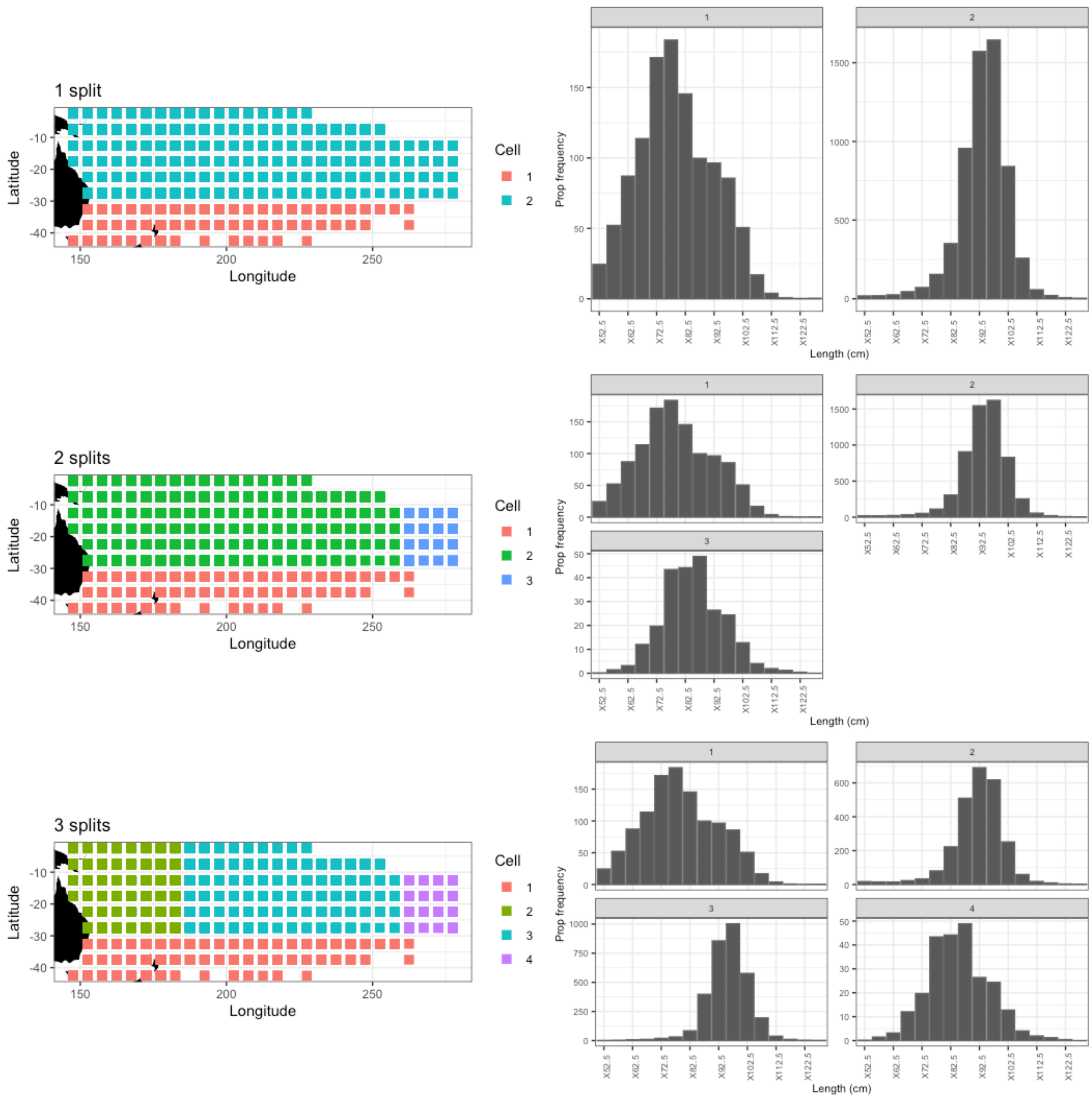


Figure 14. Spatial LF results for all ALB using 1 (first row), 2 (second row) and 3 (last row) splits (i.e., 2-, 3- and 4-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 15 for the results of 4 and 5 splits.

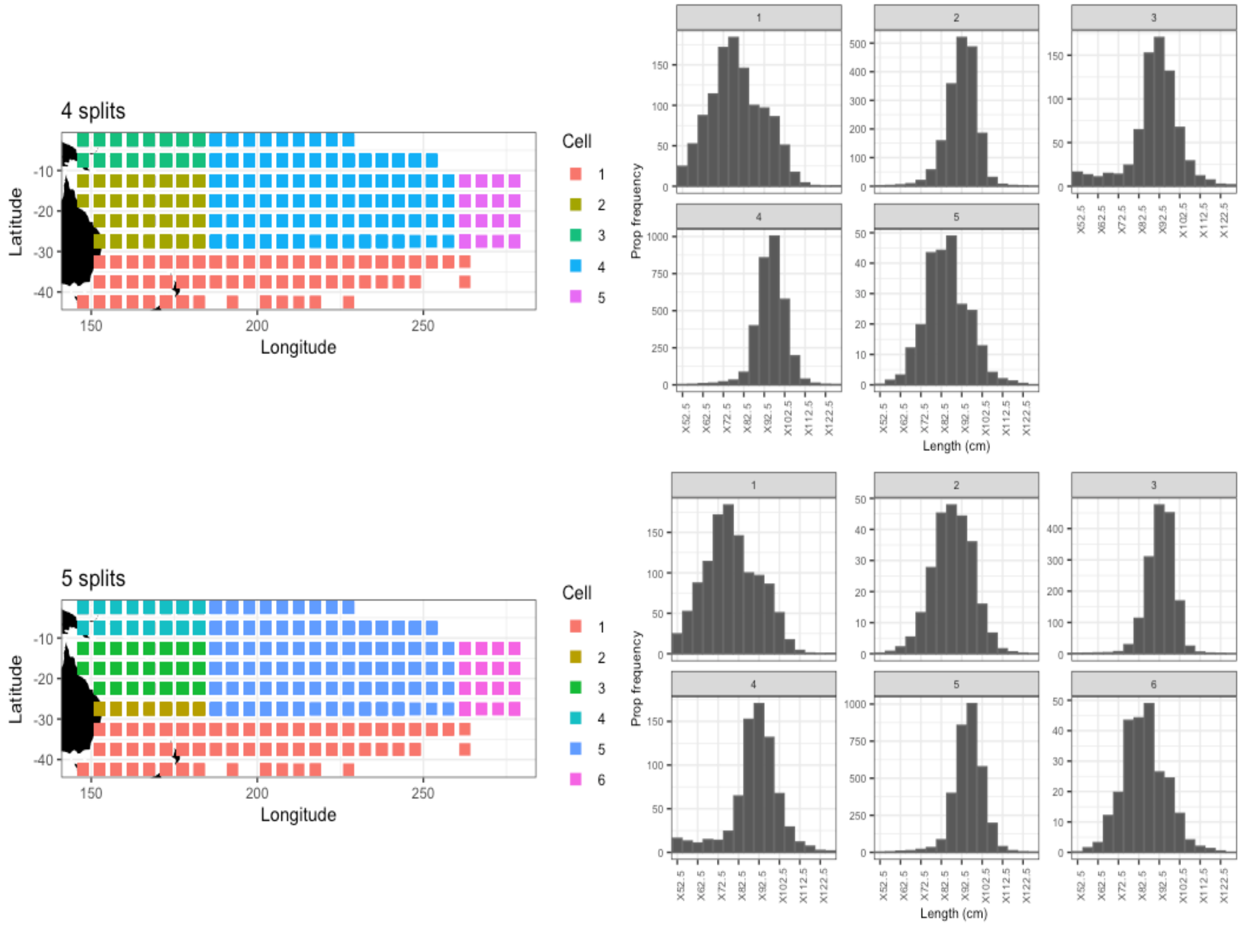


Figure 15. Spatial LF results for all ALB using 4 (first row) and 5 (second row) splits (i.e., 5- and 6-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 14 for the results of between 1 and 3 splits.

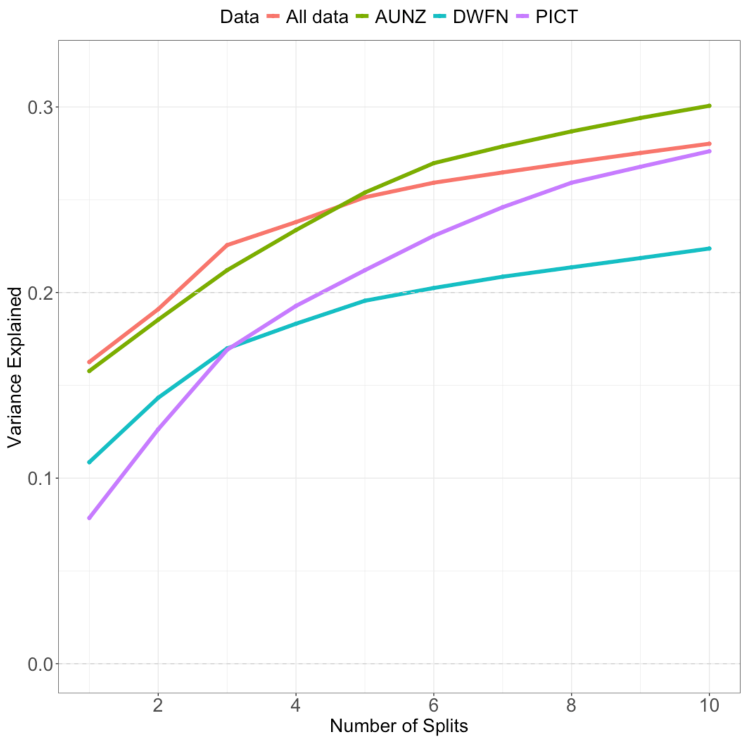


Figure 16. Percentage variance explained with increasing the number of splits in the ALB spatial LF analysis for all data (red line), AU and NZ data only (green line), DWFN only (blue line) and PICT only (purple line).

Results: AU and NZ only

The spatial and temporal coverage of AU and NZ only ALB LF records between 1990 and 2022 was patchy, with geographical sparseness in the earlier part of this time series (prior to 2002, Figure 17) and in the later part of the time series (post-2013, Figure 17). The length distributions and the mean length throughout the region are shown in Figure 18 and Figure 13B, respectively. The number of length measurements in each grid cell was variable, some grid cells had thousands of records and others had < 100 (Figure 18). Larger length fish were observed in the northern region (Figure 13B), and – similar to the analysis for All Data – the first split delineated this region at 30°S (Figure 19). This split explained over 15% of the variance in the data (Figure 16). Further spatial delineations are shown in Figure 19 and Figure 20. The percentage variance increased steadily until six splits (i.e., a seven region model explained 27% of the variance in the data), but plateaued thereafter and at 10 splits (i.e., 11 regions) the percentage variance explained was approximately 30% (Figure 16).

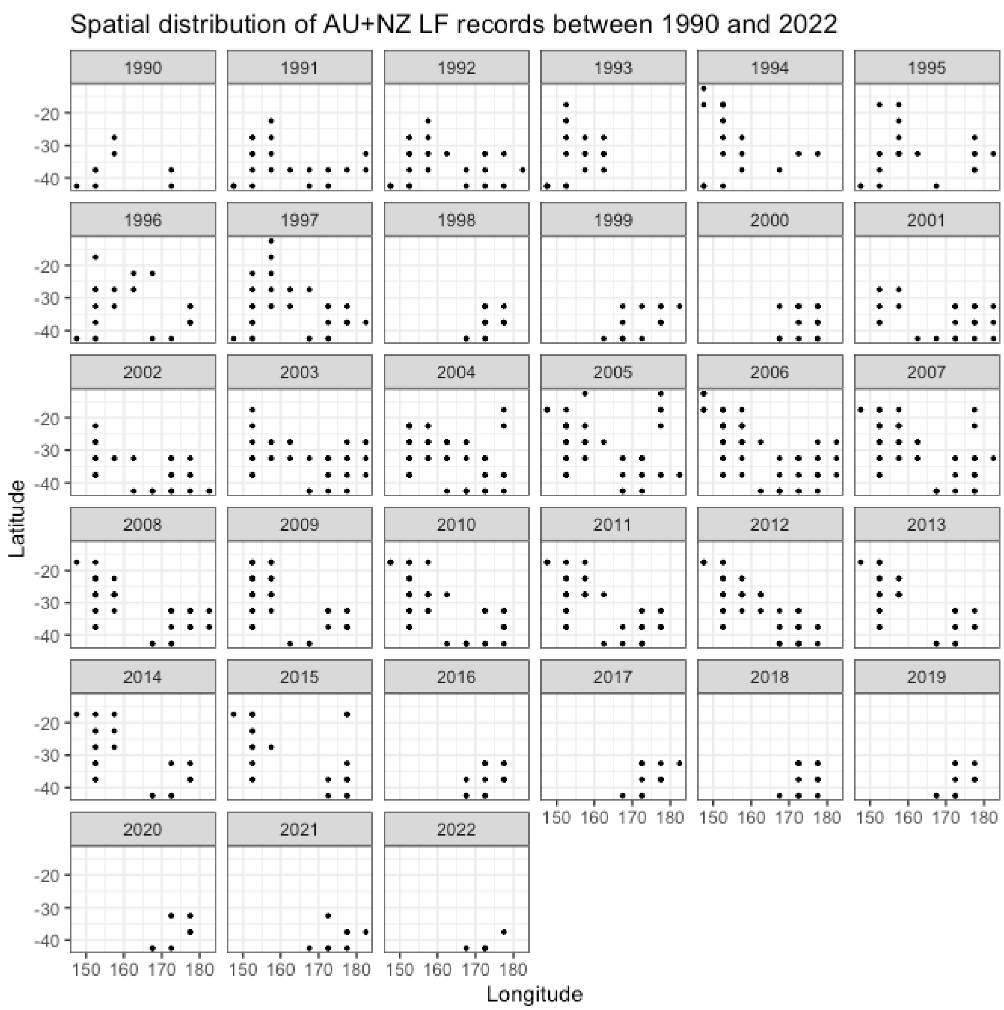


Figure 17. Spatial location of AU and NZ LF ALB records, by year.

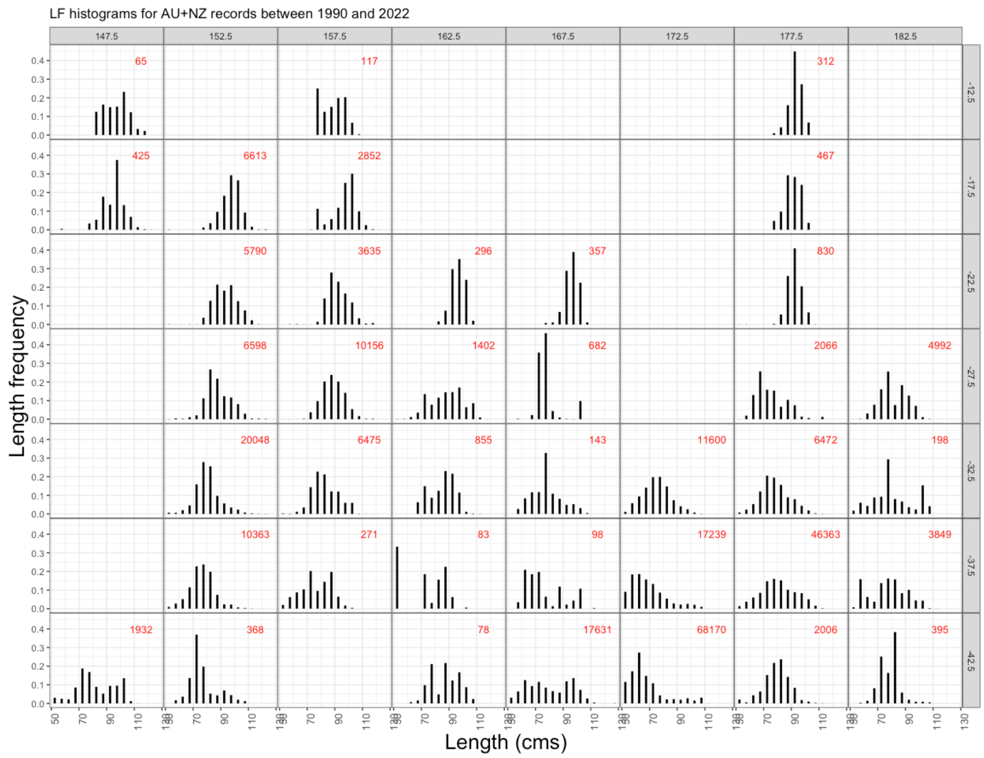


Figure 18. LF histograms for AU and NZ ALB, with sample size included at top of grid cell in red.

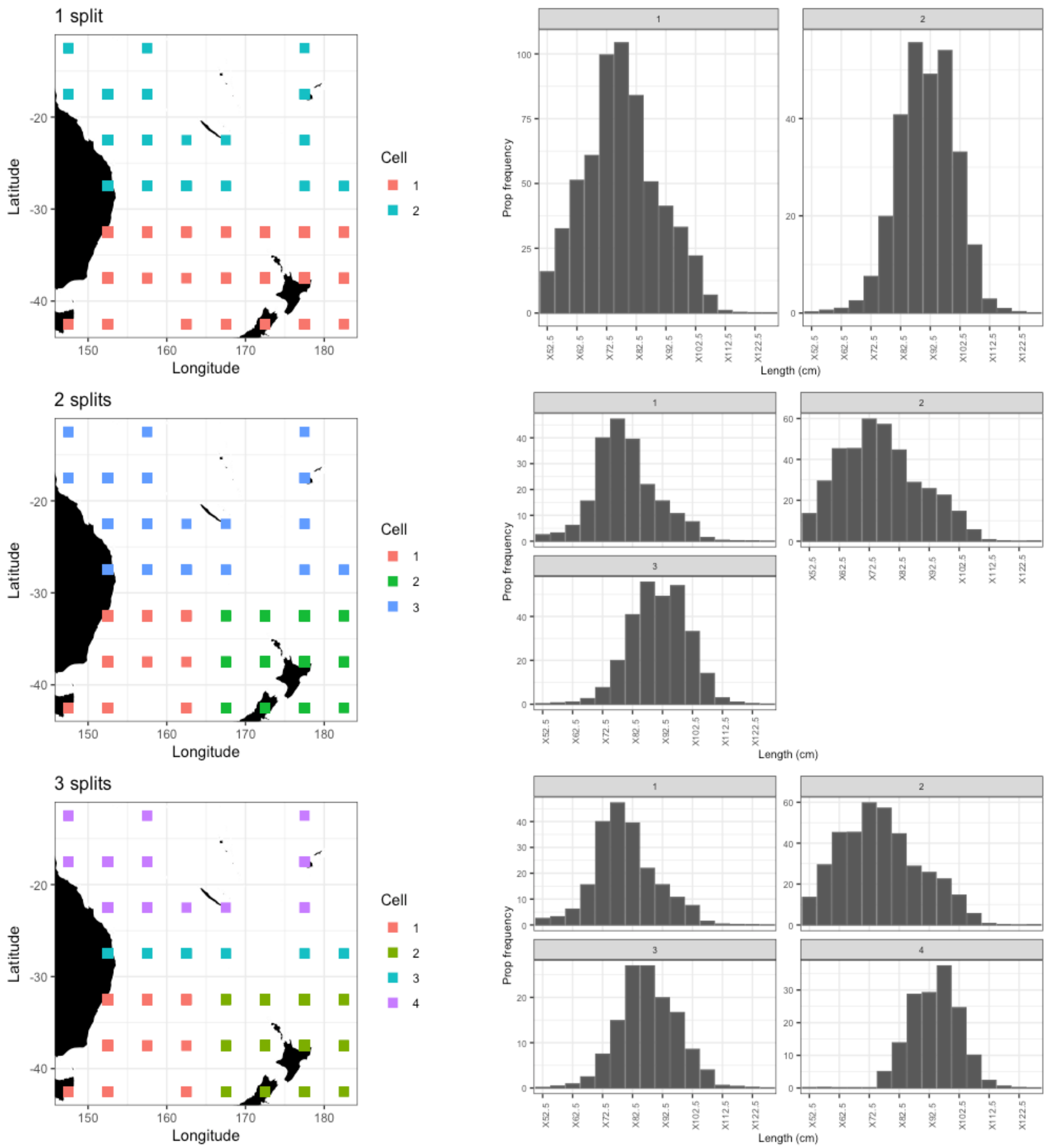


Figure 19. Spatial LF results for AU and NZ ALB using 1 (first row), 2 (second row) and 3 (last row) splits (i.e., 2-, 3- and 4-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 20 for the results of 4 and 5 splits.

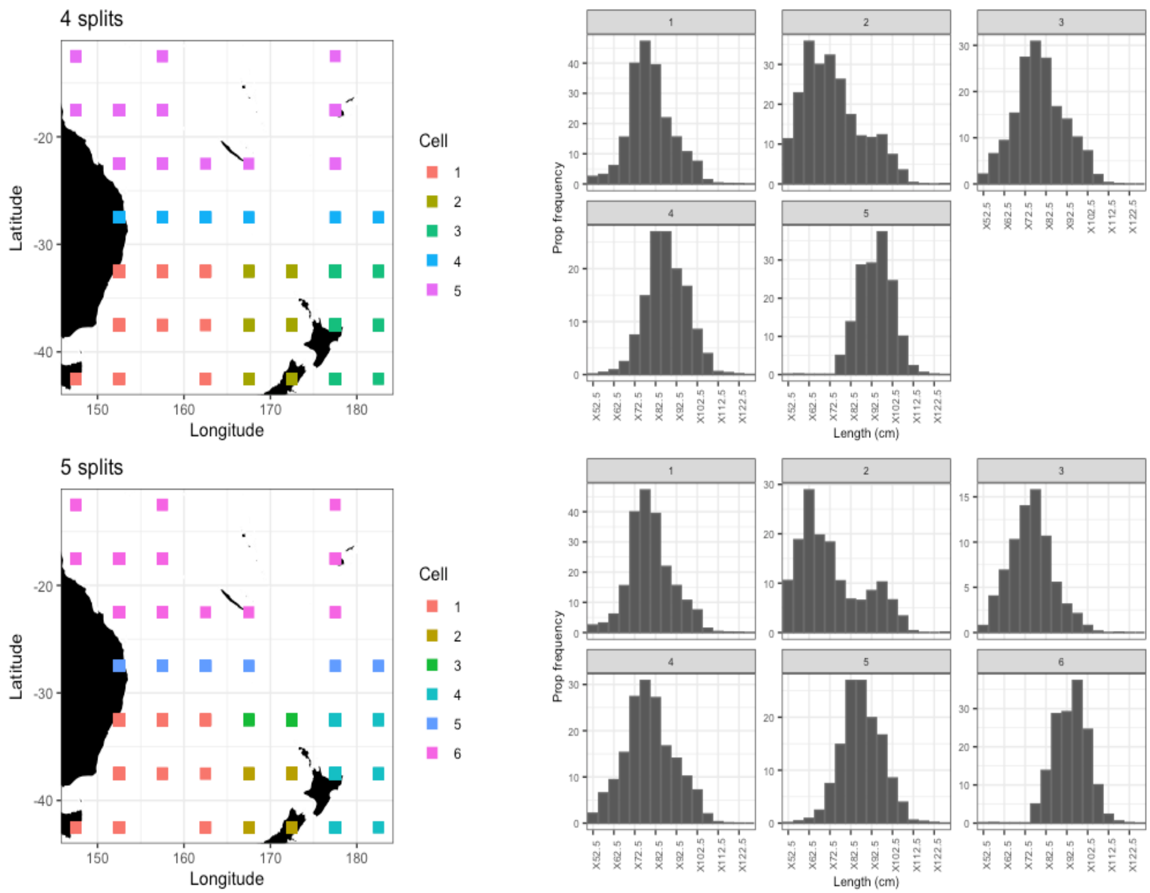


Figure 20. Spatial LF results for AU and NZ ALB using 4 (first row) and 5 (last row) splits (i.e., 5- and 6-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 19 for the results of between 1 and 3 splits.

Results: DWFN only

The spatial and temporal coverage of DWFN ALB LF records between 1990 and 2022 was patchy, with geographical sparseness in the middle part of this time series (between 1996 and 2010, Figure 21). The length distributions and the mean length throughout the region are shown in Figure 22 and Figure 13C, respectively. The number of length measurements in each grid cell was variable, but most grid cells had hundreds, if not thousands, of LF records (Figure 22). Larger length fish were observed in the northern region (Figure 13C), and the first split delineated this region at 25°S (Figure 23) – slightly further north than the All Data and AU and NZ only analyses. This split explained approximately 11% of the variance in the data (Figure 16). Further spatial delineations are shown in Figure 23 and Figure 24. The percentage variance explained was lower than other analyses (just over 20% after 10 splits, Figure 16).

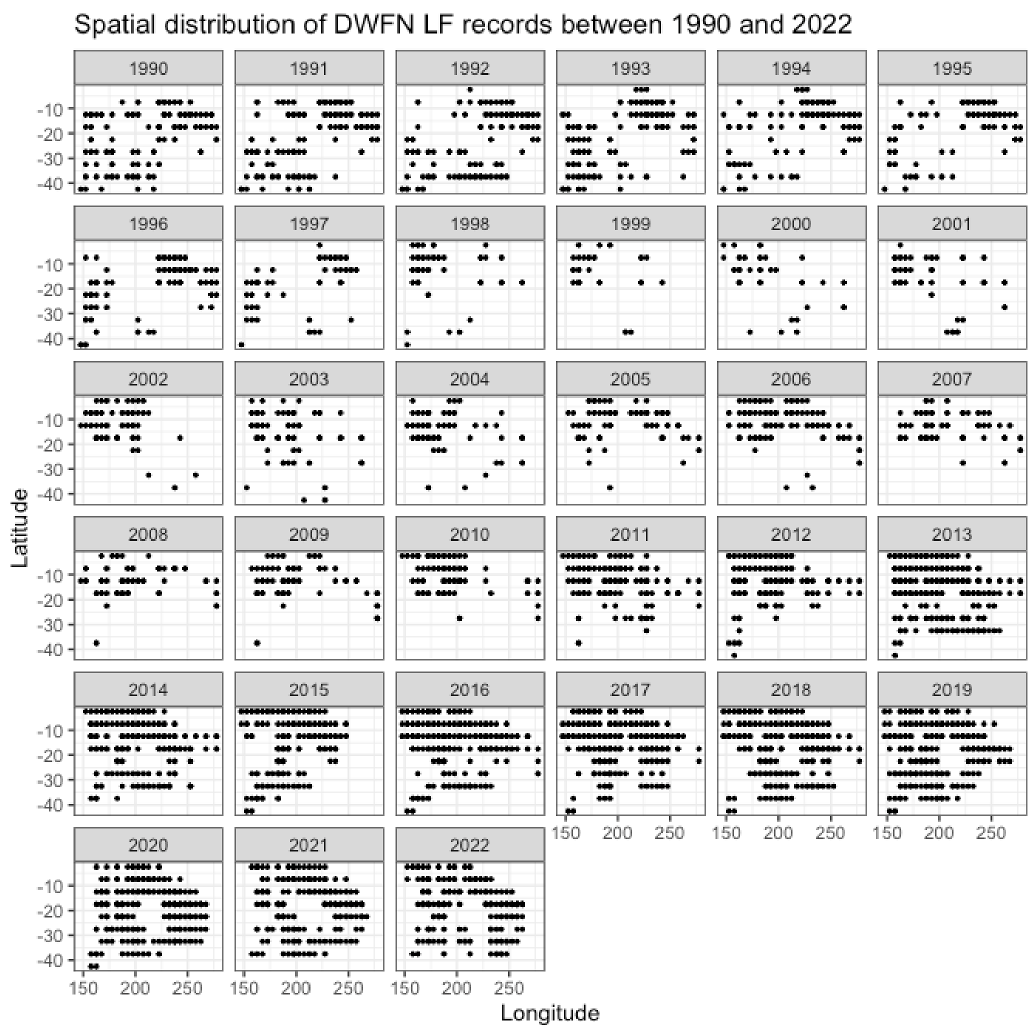


Figure 21. Spatial location of DWFN LF ALB records, by year.

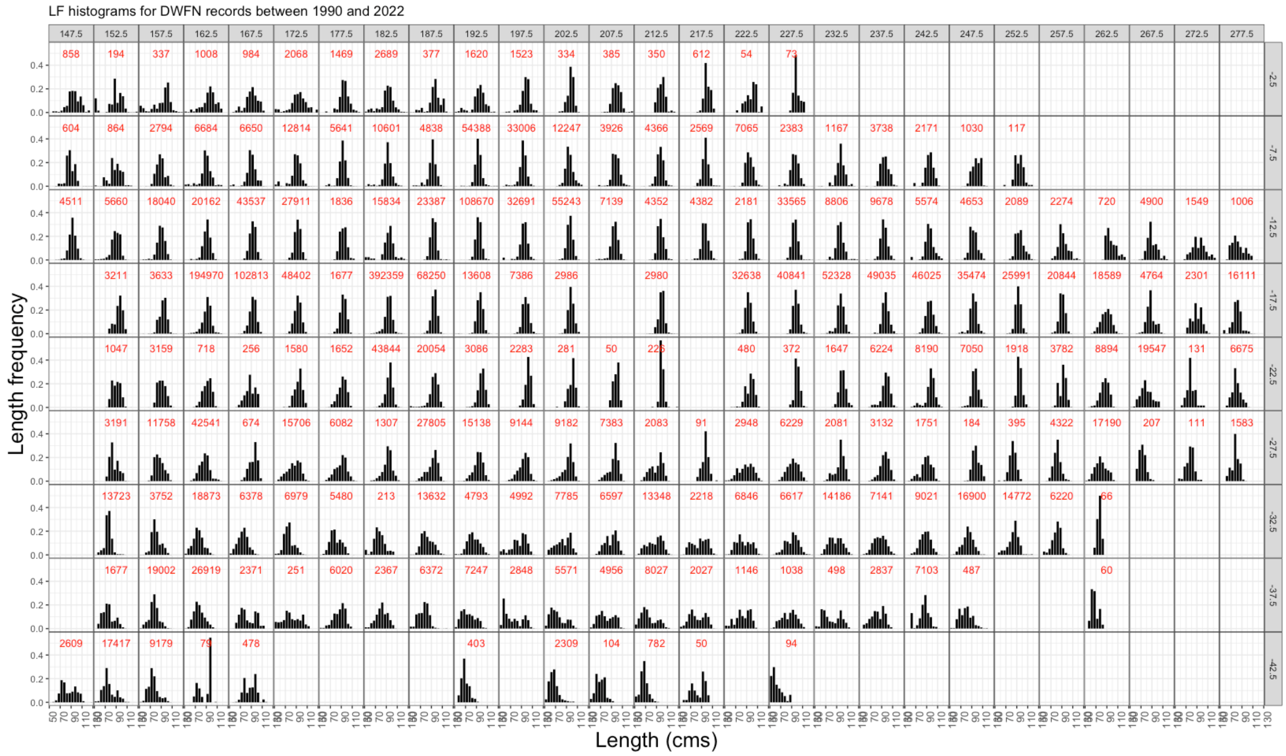


Figure 22. LF histograms for DWFN ALB, with sample size included at top of grid cell in red.

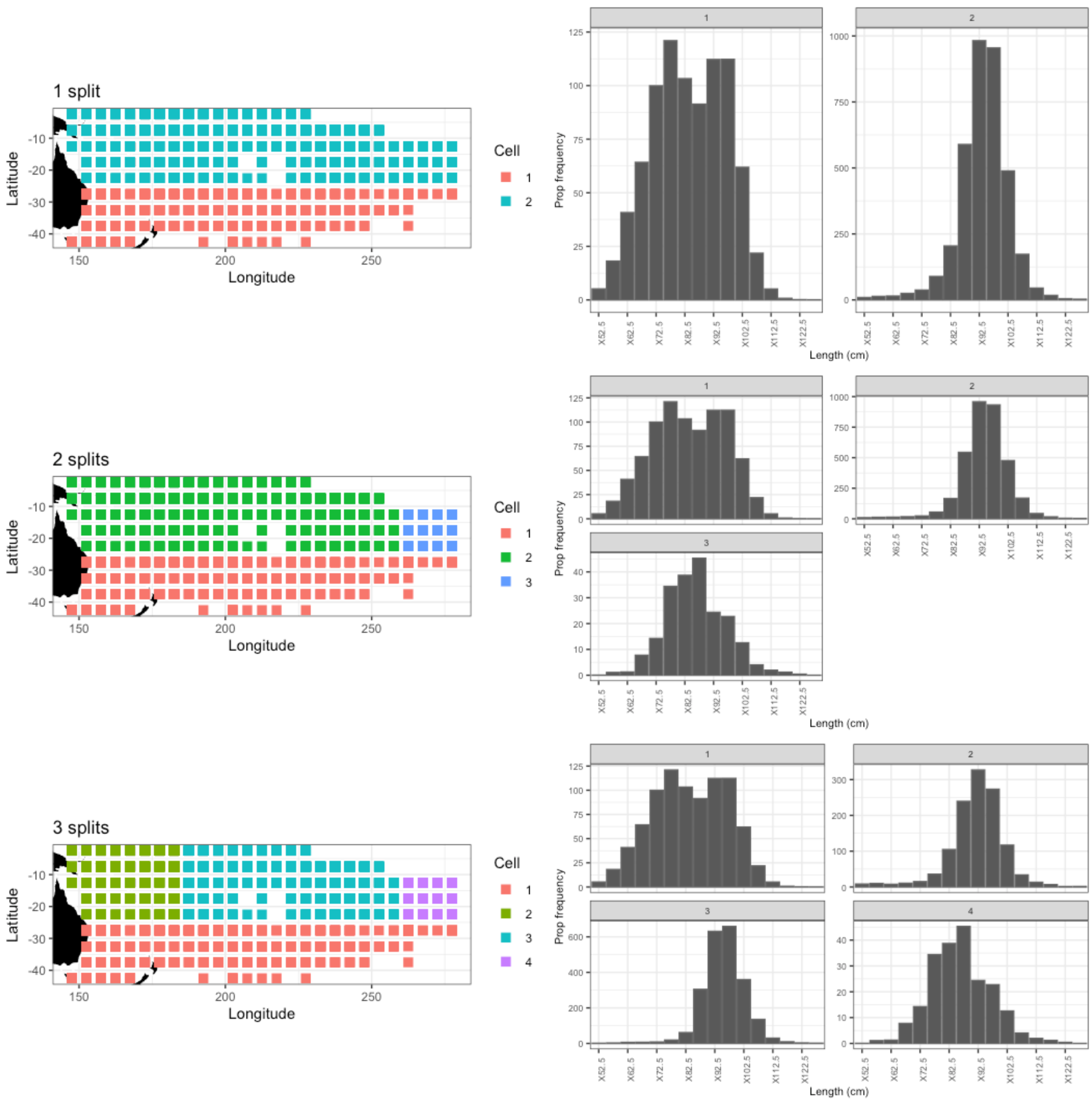


Figure 23. Spatial LF results for DWFN ALB using 1 (first row), 2 (second row) and 3 (last row) splits (i.e., 2-, 3- and 4-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 24 for the results of 4 and 5 splits.

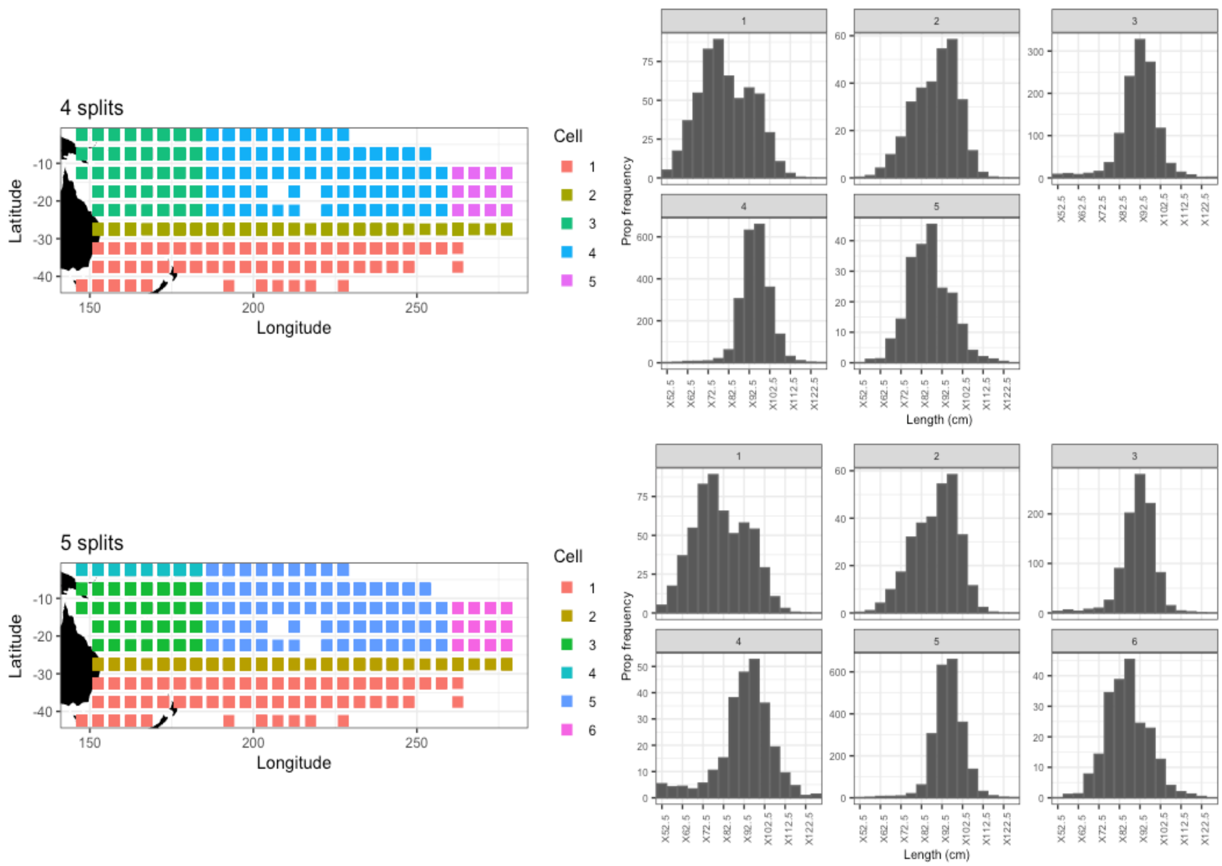


Figure 24. Spatial LF results for DWFN ALB using 4 (first row) and 5 (last row) splits (i.e., 5- and 6-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 23 for the results of between 1 and 3 splits.

Results: PICT only

The spatial and temporal coverage of PICT ALB LF records between 1990 and 2022 was patchy, with geographical sparseness in the beginning of time series (up until 2002, Figure 25). The length distributions and the mean length throughout the region are shown in Figure 26 and Figure 13D, respectively. The number of length measurements in each grid cell was variable, many grid cells had hundreds, if not thousands, of LF records of others had < 100 records (Figure 26). Smaller length fish were observed in the southern region below 35°S (Figure 13D) and east of 260°. This did not concord well with the first splits (Figure 27) and this split explained less than 10% of the variance in the data (Figure 16). Further spatial delineations are shown in Figure 27 and Figure 28. The percentage variance explained was lower than other explorations initially, however it doubled by 4 splits (i.e., 5 region model) to approximately 20% variance explained, and at 10 splits was nearly 27% (Figure 16).

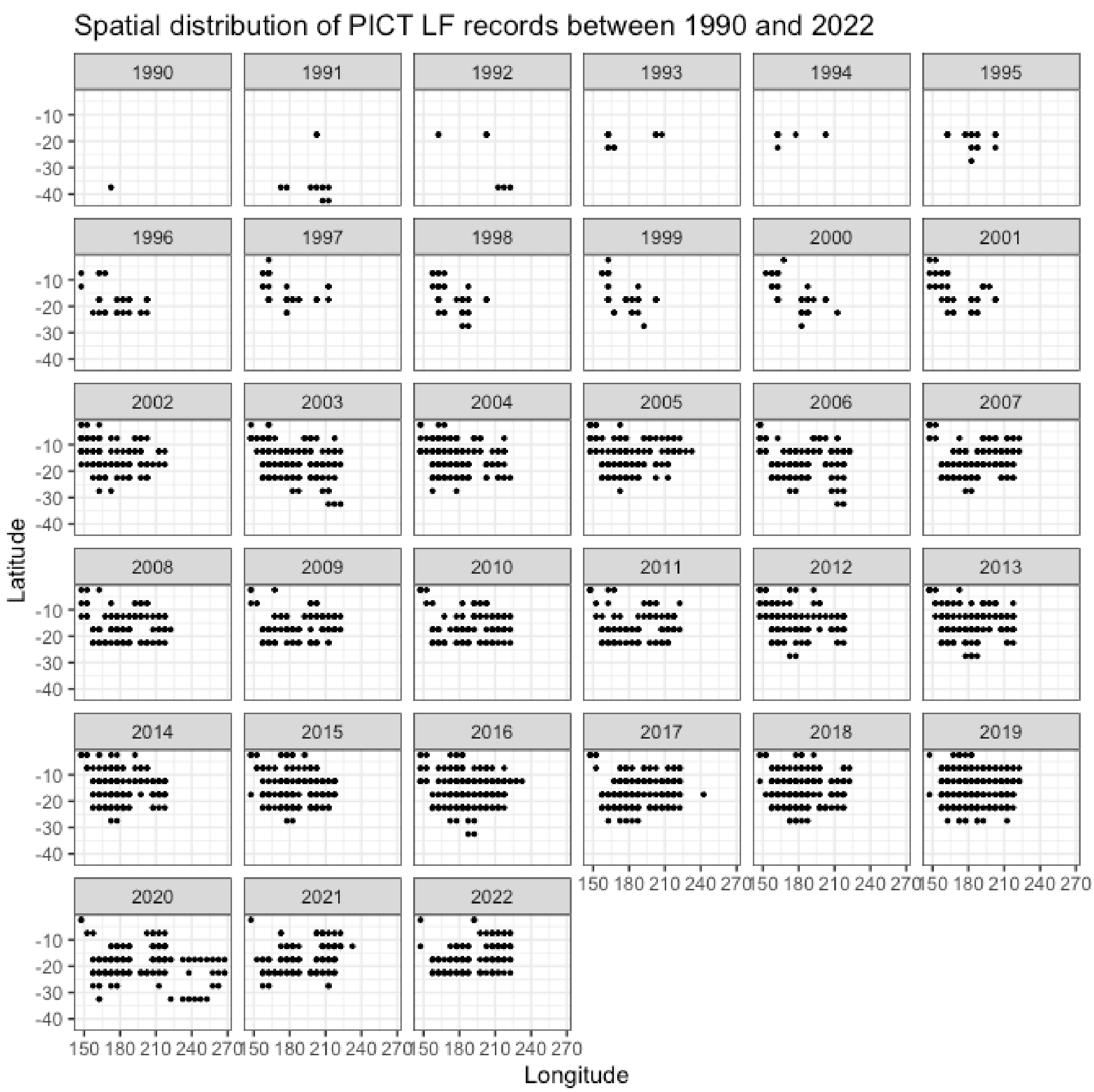


Figure 25. Spatial location of PICT LF ALB records, by year.

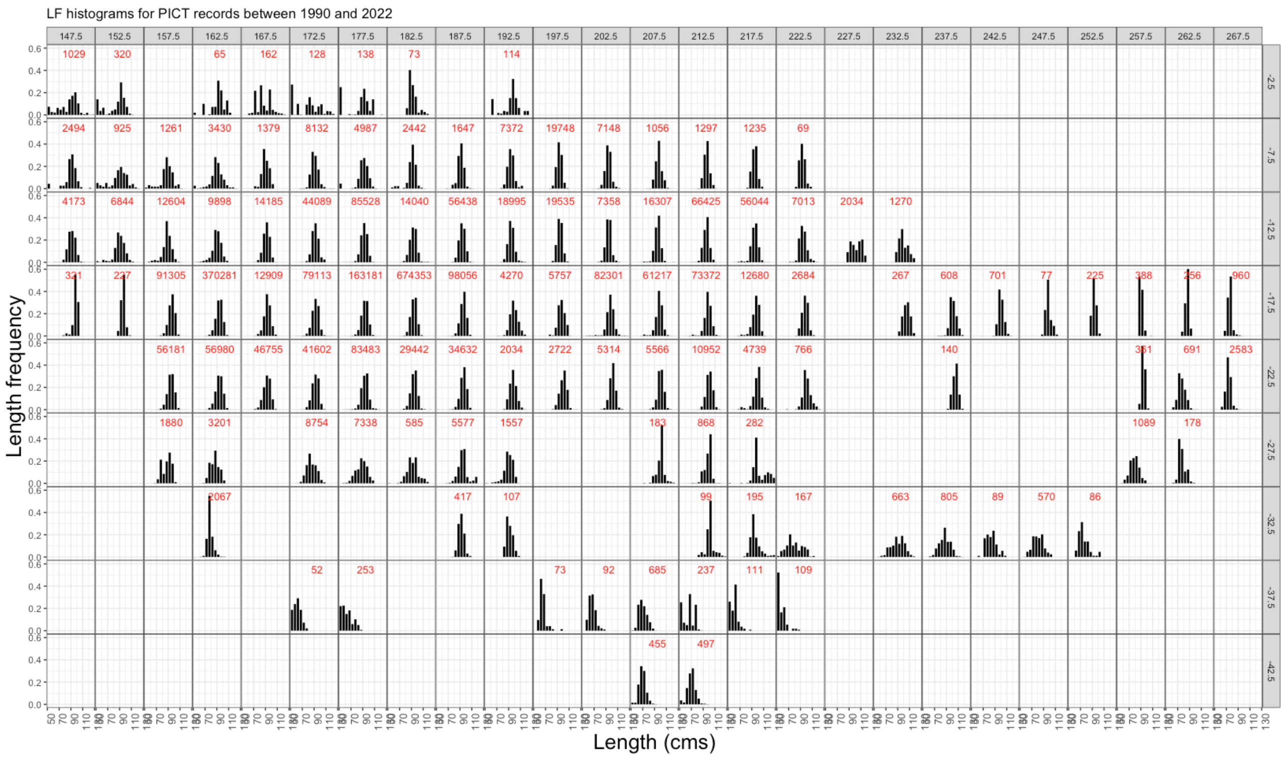


Figure 26. LF histograms for PICT ALB, with sample size included at top of grid cell in red.

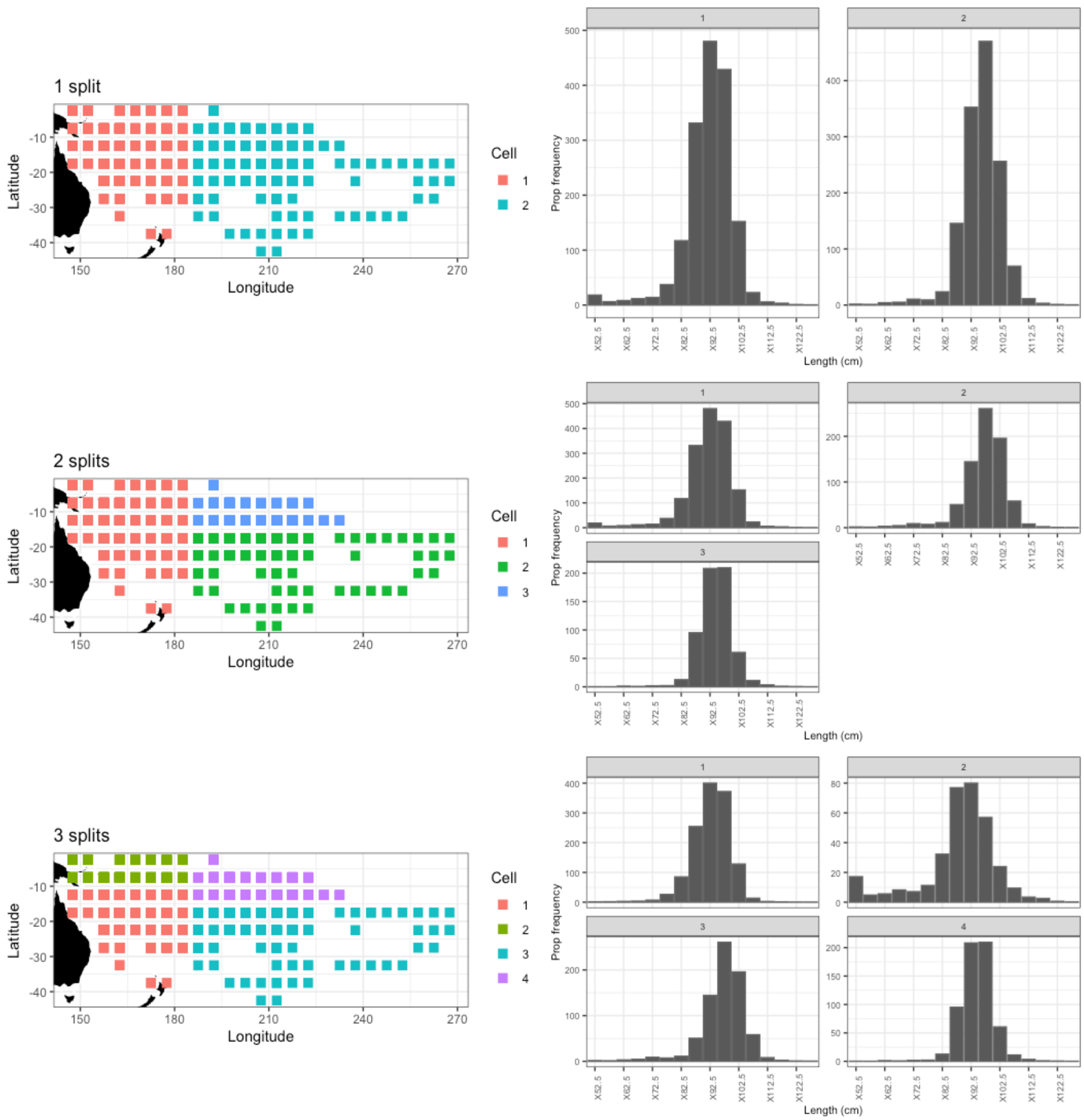


Figure 27. Spatial LF results for PICT ALB using 1 (first row), 2 (second row) and 3 (last row) splits (i.e., 2-, 3- and 4-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 28 for the results of 4 and 5 splits.

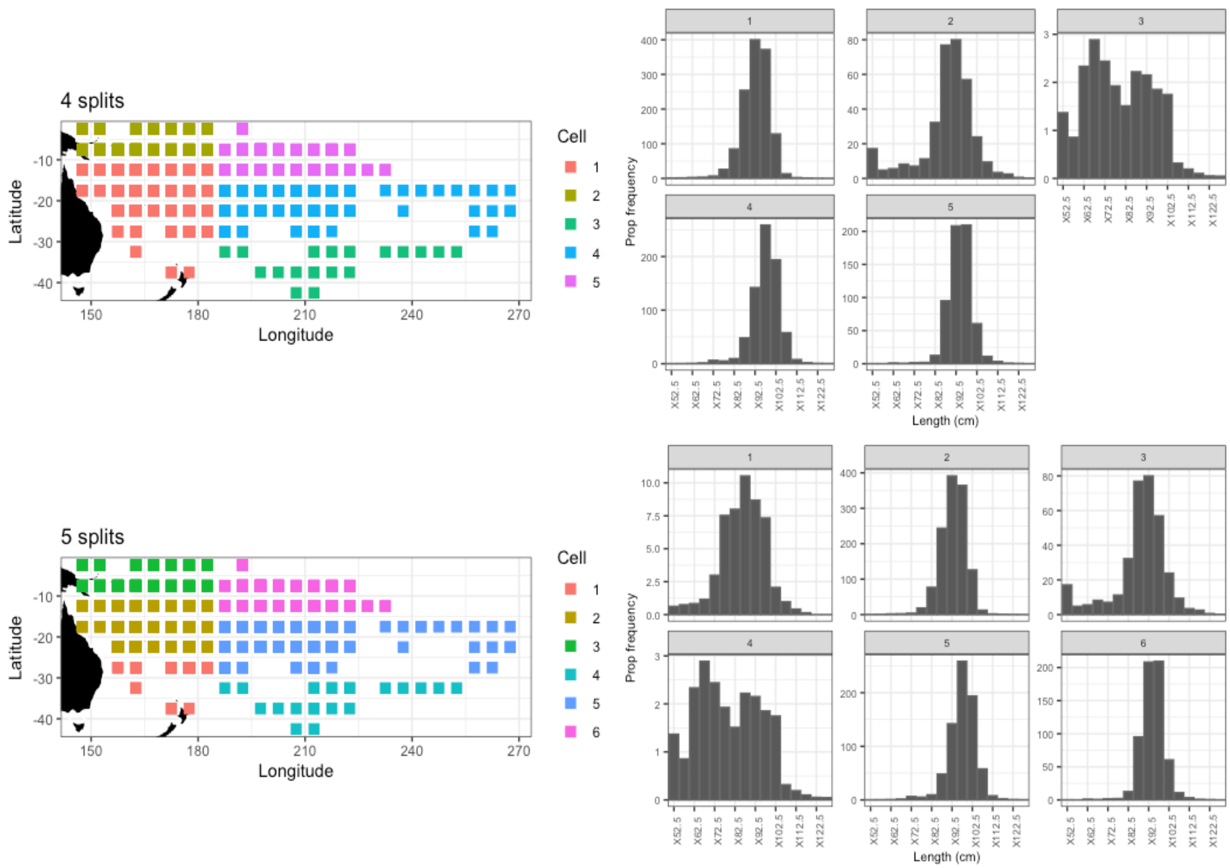


Figure 28. Spatial LF results for PICT ALB using 4 (first row) and 5 (last row) splits (i.e., 5- and 6-regions). Lefthand column contains the spatial aggregation of grid cells in the same delineated region, and the righthand column is the LF of data from each region. See Figure 27 for the results of between 1 and 3 splits.

Conclusions

Due to temporal and spatial sparsity in the MLS data, identification of regions for this species was not considered reliable. Specifically, (1) many grid cells had too few observations to reliably reflect the underlying LF distribution and were omitted from the analysis, and (2) although attempts were made to delineate regions, the percentage variance explained was exceptionally low (less than 12% even with as many as 10 splits, Figure 8), rendering the results unreliable for making robust decisions of spatial fleet delineations.

The temporal and spatial coverage of ALB data was significantly better than MLS. Regardless of which flags were included in each analysis (i.e., All Data, AU and NZ, DFWN, or PICT), smaller fish were observed in the southern region, and except for the PICT-only analysis, this region was always identified in the first split. The exact location of the split varied between 25°S and 30°S, depending on which flags were included. The percentage variance explained of the first split was between 10-15%, except for the PICT-only analysis that was < 10%. However, regardless of which flags were included in the analysis, the increase in percentage variance explained more than doubled within 4 to 6 splits (to over 20%).

Recommendations

The spatial and temporal coverage of length-frequency data for the MLS was not sufficient to ensure reliable spatial fleet delineations based on the spatial regression tree analysis presented here. For ALB, regardless of flag-combination analysed, smaller fish were consistently identified in the southern region, with the first split typically occurring between 25°S and 30°S. While these results offer some insight into the spatial structure of ALB in the region, they may not provide sufficient evidence to warrant a revision of the current ALB stock assessment structure, as percentage variance explained plateaued at approximately 30% with as many as 10 splits (i.e., 11 regions). As spatial coverage improves with future data collection, and accuracy of collected data improves (i.e., rounding length observations as evident by heaping in the LF histograms), these analyses should be repeated.

References

Castillo-Jordan, C. et al. (2024) Background Analyses for the 2024 Stock Assessment of Striped Marlin in the Southwest Pacific Ocean. SC20-SA-IP-06. August, 2024.

Hamer, P., J. Macdonald, J. Potts, T. Vidal, T. Teears & I. Senina (2023). Review and analyses to inform conceptual models of population structure and spatial stratification of bigeye and yellowfin tuna assessments in the Western and Central Pacific Ocean. WCPFC-SC19-2023/SA-WP-02. July, 2023.

Lennert-Cody, C. E., M., Minami, P. K., Tomlinson & M. N., Maunder (2010). Exploratory analysis of spatial–temporal patterns in length-frequency data: an example of distributional regression trees. *Fish. Res.* 102, 323–326.

Lennert-Cody, C. E., M. N., Maunder, A., Aires-de-Silva, & M., Minami (2013). Defining population spatial units: Simultaneous analysis of frequency distributions and time series. *Fish. Res.* 139, 85–92.

Xu, H. & C. E. Lennert-Cody (2023). FishFreqTree: IATTC's regression tree R package for fisheries' size frequency data. R package v. 3.3.2