NPFC-2024-SSC PS14-WP13

**Exploratory analysis of spatio-temporal patterns in Pacific saury size composition data**

Yi-Jay Chang, Jhen Hsu, Zi-Wei Yeh, Kuan-Chun Teng, Chi-Yi Chung

Institute of oceanography, National Taiwan University

**Abstract**

This study analyzed Pacific saury size composition data from Japan and Chinese Taipei using multivariate regression tree analysis to identify spatial patterns and propose candidate fleet definitions for stock assessment models. Length frequency and commercial category analyses revealed distinct spatial boundaries at 155.5°E for Japanese fleets and 160.3°E for Chinese Taipei fleets. Japanese data showed a mixed length distribution west of 155.5°E and unimodal patterns to the east, while Chinese Taipei data exhibited uniform length frequency patterns across space but heterogeneous commercial size distributions. Both fleets demonstrated consistent seasonal patterns, with smaller fish distributed in southwestern areas in the fourth season (October-December), and between 163.3°E-170°E. Although additional spatial splits increased explained variance, we recommend using primary boundaries (155.5°E and 160.3°E for Japanese and Chinese Taipei fleets, respectively) to maintain analytical simplicity while effectively capturing size composition heterogeneity for both fleets. These findings suggested the candidate spatially distinct fleet definitions may better account for heterogeneity in catch-at-size data in stock assessment modeling.

**1. Introduction**

Catch-at-size data often exhibits complex distributions, characterized by skewness and multimodality, stemming from both biological and fishing-related factors including size-dependent habitat preferences, seasonal migration patterns, and spatiotemporal variation in fishing activities. Such patterns of spatiotemporal heterogeneity in catch composition have also been observed in the Pacific saury fisheries (Suyama et al., 2006; Chang et al., 2019). In this study, we applied a multivariate regression tree approach to analyze Pacific saury size composition data (length frequency and commercial category data) from Japanese and Chinese Taipei fleets, which collectively provide comprehensive spatial coverage of saury in the Northwestern Pacific Ocean. The objectives were to: (i) examine spatiotemporal patterns in size composition; (ii) identify areas with consistent size composition patterns; and (iii) propose candidate fleet definitions for the development of age/size-structured stock assessment models.

**2. Material and methods**

*2.1 Size composition data*

The size composition data for this analysis were obtained from Japan and Chinese Taipei through the NPFC SSC PS. Although other members (Korea, Russia, China, and Vanuatu) also provided catch-at-size data, this exploratory analysis specifically focused on Japanese and Chinese Taipei fleets due to their extensive spatiotemporal coverage (Japan: 1994-2023; Chinese Taipei: 2007-2023) and larger data amount. The dataset consists of monthly aggregated length frequency data at a spatial resolution of 1° × 1° grid cells. It includes information on year, month, location (latitude and longitude), and estimated fish numbers by age class (Age-0 and Age-1) and by 1-cm length intervals (14-35 cm knob length, KnL). The detailed estimation procedures for catch-at-length/age for each fleet are described in Fuji et al. (2019) and Huang et al. (2020; 2021). Additionally, Chinese Taipei provided commercial category data classified by weight into six categories (Ex\_large, No\_1, No\_2, No\_3, No\_4, and No\_5). The weight ranges for each commercial category are generally defined as follows: Ex\_large: > 155 grams; No\_1: 130 - 155 *g*; No\_2: 100 - 130 *g*; No\_3: 80 - 100 *g*; No\_4: 70 - 80 *g*, and No\_5: 60 - 70 *g*.

*2.2 Multivariate regression tree analysis*

We applied a multivariate regression tree method to the binned length data and the commercial category data for Pacific saury, following the approach of Lennert-Cody et al. (2010). The length data were binned into 22 intervals from ≤14 cm to ≥35 cm in 1 cm increments. This method employs binary decision rules to recursively partition data into increasingly homogeneous groups, with node heterogeneity measured by the Kullback-Leibler divergence (KLD; Wang et al., 2005). The method intentionally omits a pruning step as its primary objective is large-scale data exploration and development of concise tree models, rather than optimization of predictive accuracy (Breiman et al., 1984).

To examine spatiotemporal variations in Pacific saury size composition, we incorporated four variables in the regression tree analysis: year, season (2nd season: May - June; 3rd season: July - September; 4th: October - December), longitude, and latitude. We evaluated several fishery definitions by varying the number of groups (*k* = 2 - 4), assessing their explanatory power to determine optimal grouping configurations and boundaries. All analyses were conducted using R version 4.2.2 (R Core Team, 2022), with the multivariate regression tree analysis implemented through the “FishFreqTree” R package (Xu and Lennart-Cody, 2022).

**3. Results and discussions**

*3.1 Spatial distribution of size composition data by seasons*

Spatial distributions of length-frequency data were analyzed at a 1° × 1° grid resolution for the Japanese fleet (May-December, 1994-2023; **Figure 1**) and Chinese Taipei fleet (May-December, 2007-2023; **Figures 2 and 3**). The Japanese fleet data exhibited a distinct bimodal length-frequency distribution in regions west of 155°E. The Chinese Taipei fleet data showed a more uniform pattern, characterized by a single modal peak in length-frequency distributions across all spatial grids. However, the spatial distribution of commercial categories composition showed slightly large heterogeneity across the spatial grids compared to its length-frequency distributions for the same fleet.

Mean length distributions of Pacific saury were calculated at a 1° × 1° grid resolution for both Japanese (**Figure 4**) and Chinese Taipei fleets (**Figure 5**). Seasonal patterns in mean length distribution were also examined for both fleets. For the Japanese fleet, larger fish were mainly found west of 150°E during the third season (July-September; **Figure 4c**), while smaller fish were distributed in southern areas in the fourth season (October-December; **Figure 4d**). The Chinese Taipei fleet data indicated that larger fish were mainly distributed east of 155°E between the second (May-June; **Figure 5b**) and third (July-September; **Figure 5c**) seasons. Similar to the Japanese fleet’s observations, smaller fish were more prevalent in southern areas in the fourth season (**Figure 5d**).

*3.2 Results for multivariate regression tree analysis*

(1) Japan

Multiple regression tree analysis of Japanese length composition data identified optimal boundaries beginning with 155.5°E (*k* = 1; **Figure 6a**), explaining 4.2% of variance (**Table 1**). This boundary separated a mixed modal distribution in the west from a unimodal distribution in the eastern high seas. Additional splits (*k* = 2 - 4) incorporated seasonal effects and identified boundaries at 148.5°E and 170.5°E, increasing explained variance to 7.2%, 8.7%, and 10%, respectively (**Figures 6b - 6d**). Overall, the analysis revealed two distinct regions: waters west of 155.5°E (near Japanese coastal waters) were characterized by a mixed modal length-frequency distribution, with a higher proportion of smaller fish dominated in southern waters during the 4th season, while the high seas region exhibited a unimodal length-frequency distribution dominated by larger fish, except for areas east of 170.5°E where smaller fish were more prevalent (**Figure 6d**).

(2) Chinese Taipei

Results of multiple regression tree analysis on Chinese Taipei’s length composition data indicated that seasonal effects best explained the spatial heterogeneity in length frequencies (*k* = 1; 1.65%; **Table 2**). However, the length-frequency patterns were similar between regions, with only slightly higher proportions of smaller fish observed in the high seas area (**Figure 7a**). Additional splits (*k* = 2 and 3) incorporated longitudinal boundaries at 163.3°E and 149.3°E, increasing explained variance to 3.3% and 4.1%, respectively (**Figures 7b** and **7c**). The fourth split (*k* = 4) used 40.3°N as a boundary with explained variance of 5%, suggesting higher proportions of smaller fish in southern waters in the fourth season (**Figure 7d**). Overall, length-frequency distributions were relatively consistent across regions, with the main distinction being a higher proportion of smaller fish in southern waters during the fourth season.

Analysis of commercial categories composition identified 160.3°E as the boundary (*k* = 1; 6.11% explained variance; **Table 3**), separating predominantly large fish (Ex\_large and No\_1) in western waters from moderate fish (No\_2 and No\_3) in eastern waters (**Figure 8a**). Additional splits revealed seasonal effects in western waters (*k* = 2; 13%), with dominated by size of moderate fish (No\_2, No\_3 and No\_4) in the 4th season (**Figure 8b**), and further established boundaries at 163.3°E and 158.3°E (*k* = 3 and 4; 14.2% and 15.1%) (**Figures 8c** and **8d**). The overall pattern showed a gradual transition from larger fish in western waters to medium-sized fish in central regions, with smaller fish (No\_4 and No\_5) becoming more prevalent east of 163.3°E.

*3.3 Summary*

Our analysis found distinct spatial and size composition patterns of Pacific saury across both fleets. The analysis also identified a consistent seasonal pattern, with smaller fish mainly distributed in southwestern areas in the fourth season (October-December). Small fish were also observed in the eastern region between 163.3°E and 170°E.

Although additional splits could increase the explained variance, this may lead to complex spatial definitions for fleets without providing substantial analytical benefits. Based on the primary spatial boundaries that effectively differentiate size compositions for both fleets, we recommend distinct spatial units for the Japanese and Chinese Taipei fleets at 155.5°E and 160.3°E, respectively. This spatial separation would create fleet units with more homogeneous size compositions, may better reflecting the observed spatial heterogeneity in catch-at-size data.

**References**

Breiman, L., and R. Ihaka. (1984). Nonlinear discriminant analysis via scaling and ACE. Davis One Shields Avenue Davis, CA, USA: Department of Statistics, University of California.

Chang, Y. J., Lan, K. W., Walsh, W. A., Hsu, J., and Hsieh, C. H. (2019). Modelling the impacts of environmental variation on habitat suitability for Pacific saury in the Northwestern Pacific Ocean. Fisheries Oceanography, 28(3), 291-304.

Cope, J.M., and Punt, A.E. (2009). Drawing the lines: resolving fishery management units with simple fisheries data. Can. J. Fish. Aquat. Sci. 66, 1256-1273.

Fuji, T., Nakayama, S.I., Hashimoto, M., Oshima, K., and Suyama, S. (2019). Estimation of catch at size (CAS) for Japanese stick-held dip net fishery. NPFC-2019-SSC PS05-WP14.

Huang W-B., Chang, Y. J., and Hsieh C. H. (2020). Body length distributions and age compositions of the Pacific saury caught by the Chinese Taipei saury fishery in 2007-2018. NPFC-2020-SSC PS06-WP18

Huang W-B., Chang, Y-J., and Hsieh C-H (2021). Estimation of catch-at-size/age data of Pacific saury using stratified random sampling with proportional allocation. NPFC-2021-SSC PS07-WP15

Huang W-B., Chang, Y. J., and Hsieh C. H. (2021). Outline of the protocol for catch at size (CAS) data for the Chinese Taipei stick-held dip net fisher. NPFC-2021-SSC PS07-WP17

Lennert-Cody, C. E., M. Minami, P. K. Tomlinson, and M. N. Maunder. (2010). Exploratory analysis of spatial-temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research 102(3), 323-326

Quinn II., T.J, and Deriso, R.B. (1999). Quantitative Fish Dynamics. Oxford University Press, New York.

Suyama, S., Kurita, Y., and Ueno, Y. (2006). Age structure of Pacific saury *Cololabis saira* based on observations of the hyaline zones in the otolith and length frequency distributions. Fisheries Science, 72, 742-749.

Team, R. C. (2022). R: A Language and Environment for Statistical Computing, Reference Index, R Foundation for Statistical Computing.

Wang, Q., Kulkarni, S.R., Verdú, S. (2005). Divergence estimation of continuous distributions based on data-dependent partitions. IEEE Trans. Inform. Theory 51, 3064-3074.

Xu, H., and Lennart-Cody, C. E. (2022). FishFreqTree: IATTC’s regression tree R package for fisheries’ size frequency data. R package version 3.3.1.

Table 1. The best four-split combination selected by multivariate regression tree analysis based on length frequency data from Japanese stick-held dip net fishery during 1994 – 2023. The last column shows the percentage of variance in the length frequency data being explained.

|  |  |  |  |
| --- | --- | --- | --- |
| Split | Key | Value | Var\_explained |
| Split1 | Longitude | 155.5oE | 4.2% |
| Split2 | Season | Q2 and 3; Q4 | 7.2% |
| Split3 | Longitude | 148.5oE | 8.7% |
| Split4 | Longitude | 170.5oE | 9.9% |

Table 2. The best four-split combination selected by multivariate regression tree analysis based on length frequency data from Chinese Taipei stick-held dip net fishery during 2007 – 2023. The last column shows the percentage of variance in the length frequency data being explained.

|  |  |  |  |
| --- | --- | --- | --- |
| Split | Key | Value | Var\_explained |
| Split1 | Season | Q2 and 3; Q4 | 1.6% |
| Split2 | Longitude | 163.3oE | 3.3% |
| Split3 | Longitude | 149.3oE | 4.1% |
| Split4 | Latitude | 39.3oN | 4.9% |

Table 3. The best four-split combination selected by multivariate regression tree analysis based on commercial categories composition data from Chinese Taipei stick-held dip net fishery during 2007 – 2023. The last column shows the percentage of variance in the commercial size composition data being explained.

|  |  |  |  |
| --- | --- | --- | --- |
| Split | Key | Value | Var\_explained |
| Split1 | Longitude | 160.3oE | 6.1% |
| Split2 | Season | Q2 and 3; Q4 | 13.0% |
| Split3 | Longitude | 163.3oE | 14.2% |
| Split4 | Longitude | 158.3oE | 15.0% |

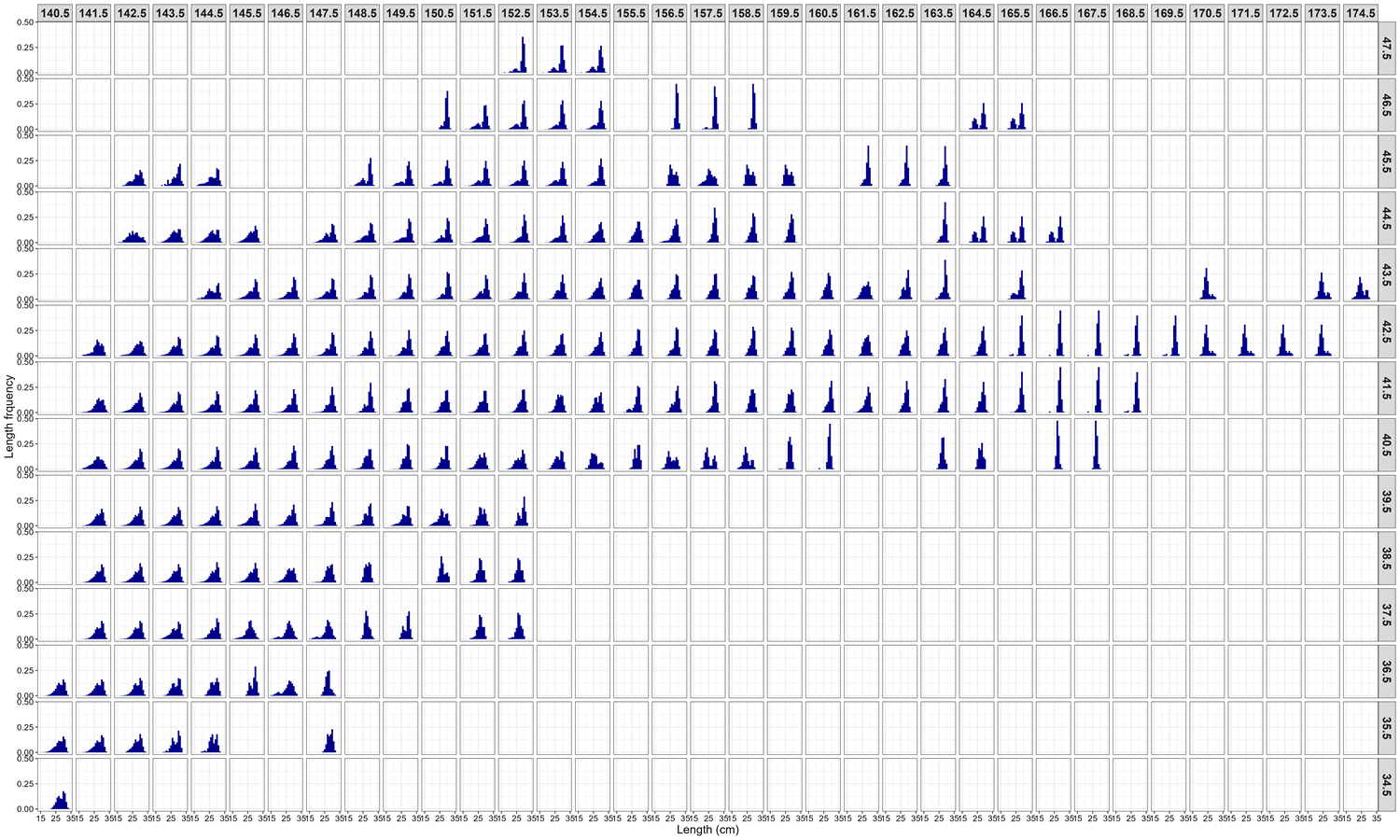


Figure 1. Spatial distribution of mean length composition of Pacific saury based on the Japanese stick-held dip net fishery during 1994-2023.

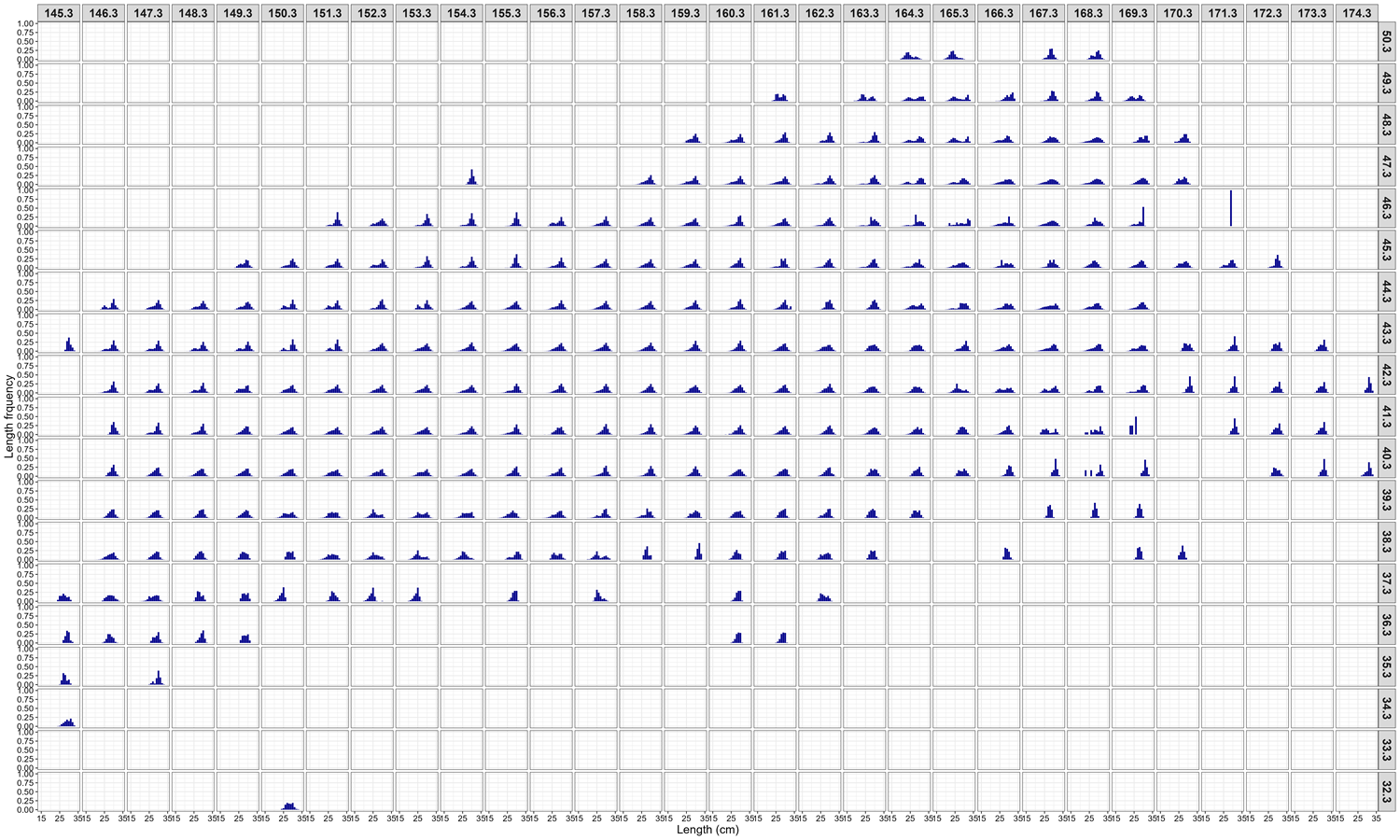


Figure 2. Spatial distribution of mean length composition of Pacific saury based on the Chinese Taipei stick-held dip net fishery during 2007-2023.

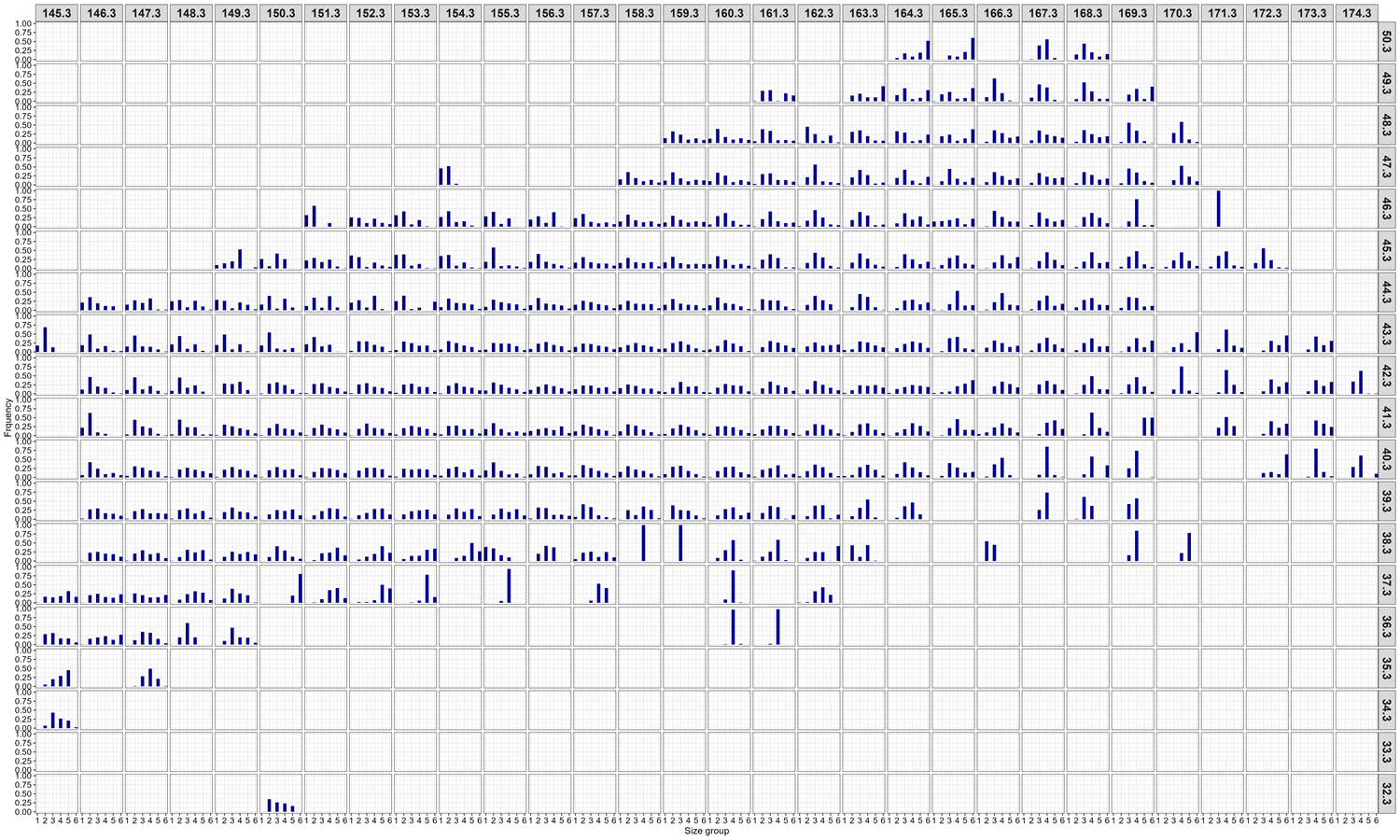


Figure 3. Spatial distribution of mean commercial size group composition of Pacific saury based on the Chinese Taipei stick-held dip net fishery during 2007-2023.

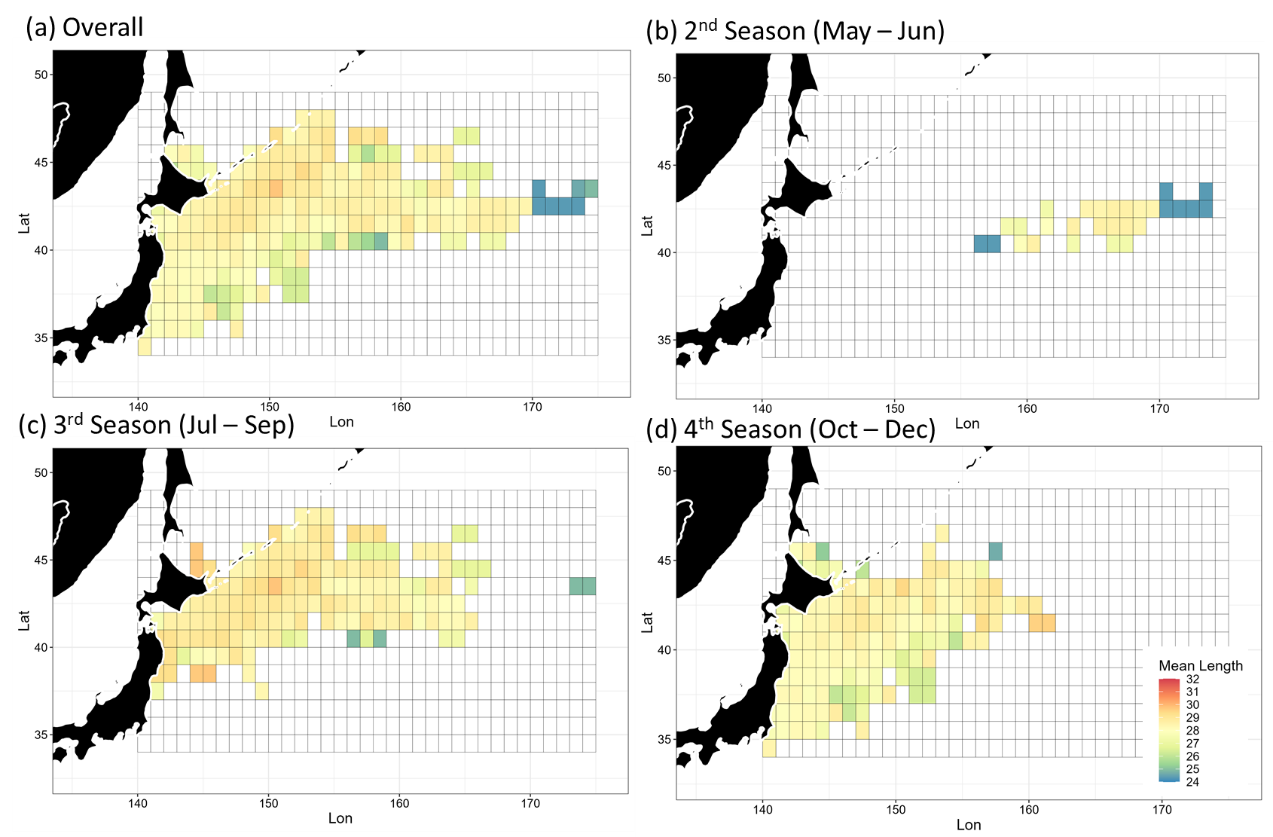


Figure 4. Spatial distribution of mean lengths of Pacific saury from the Japanese stick-held dip net fishery during 1994-2023 for (a) overall and (b-d) seasonal distributions.

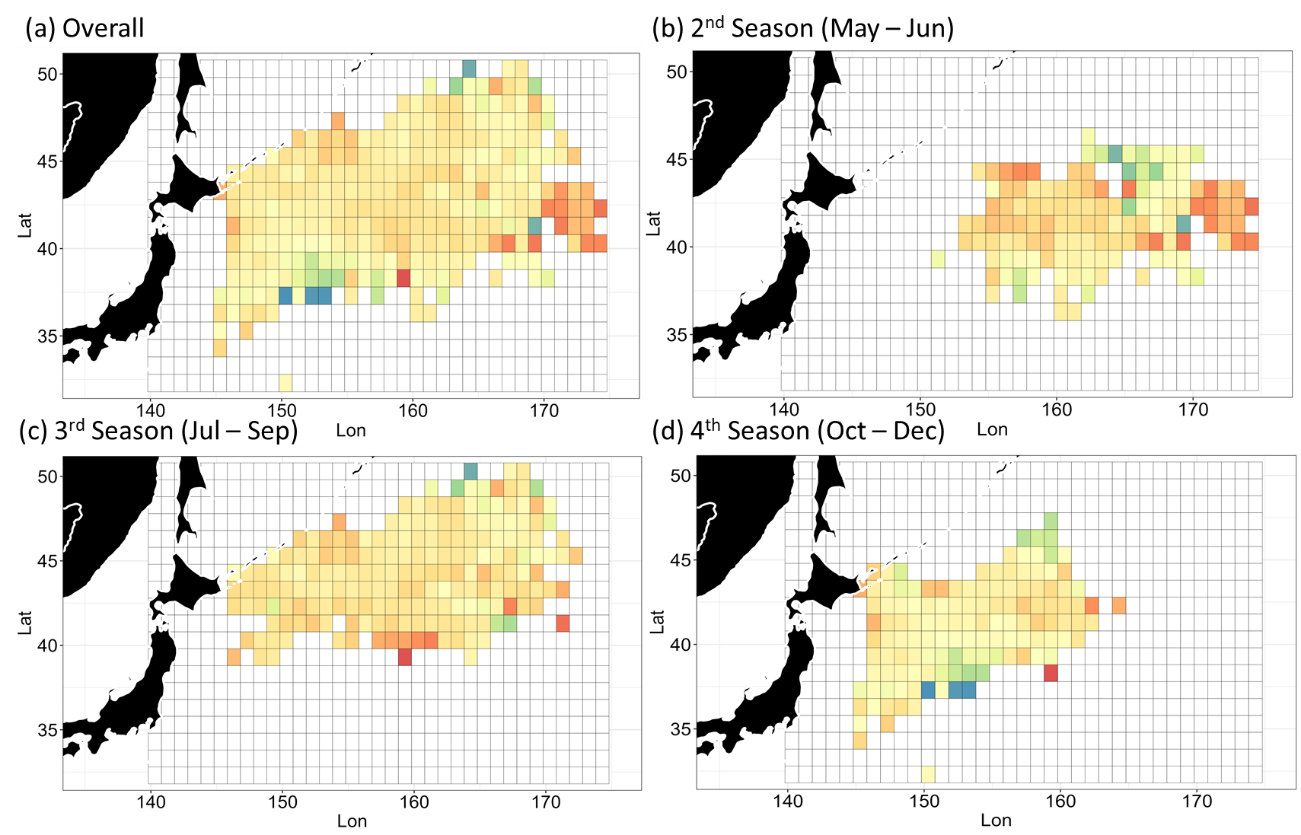


Figure 5. Spatial distribution of mean lengths of Pacific saury from the Chinese Taipei stick-held dip net fishery during 2007-2023 for (a) overall distribution and (b-d) seasonal distributions.

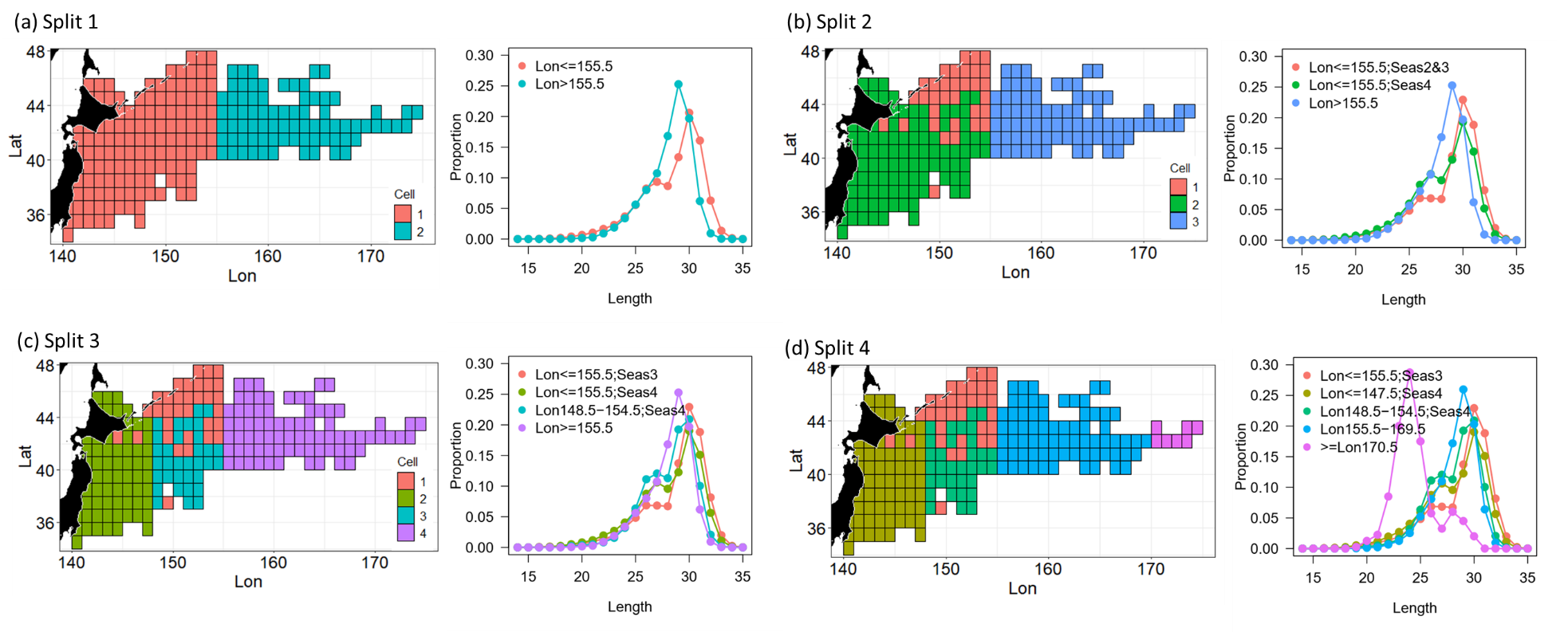


Figure 6. Spatial distribution of fisheries classified using multivariate regression tree analysis from one to four splits (a-d), based on length frequency data from Japanese stick-held dip net fishery during 1994-2023. Mean length frequency distributions for each classified region are shown in the right columns.

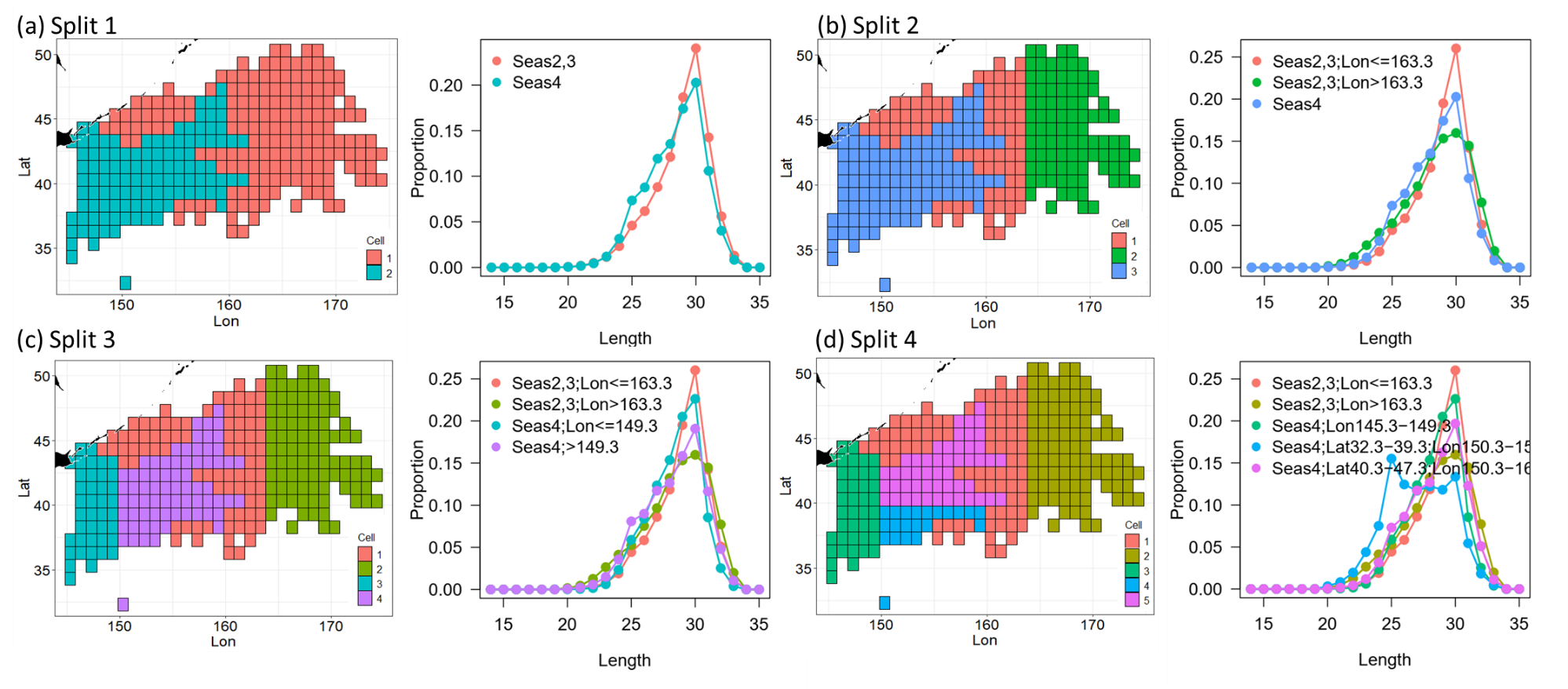


Figure 7. Spatial distribution of fisheries classified using multivariate regression tree analysis from one to four splits (a-d), based on length frequency data from Chinese Taipei stick-held dip net fishery during 2007-2023. Mean length frequency distributions for each classified region are shown in the right columns.

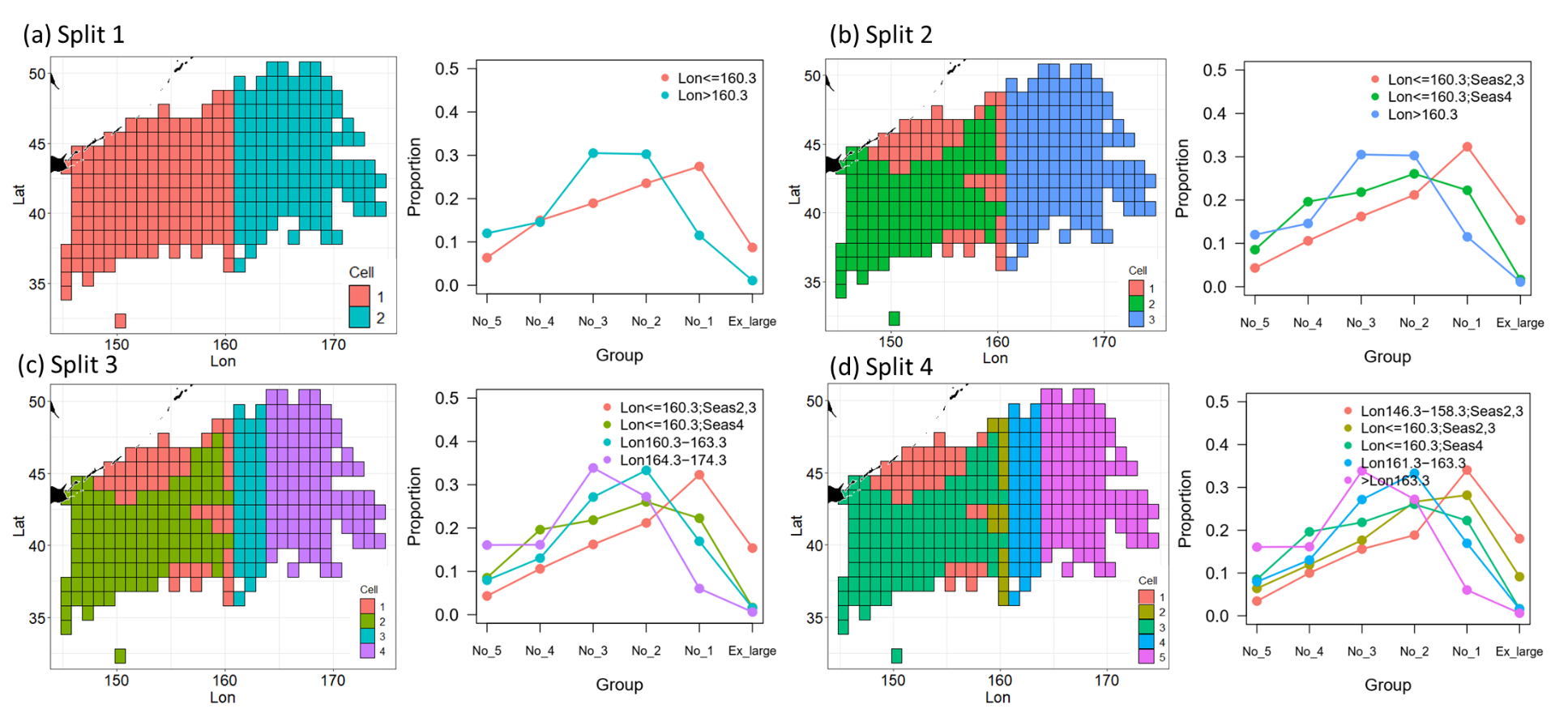


Figure 8. Spatial distribution of fisheries classified using multivariate regression tree analysis from one to four splits (a-d), based on commercial size group composition data from Chinese Taipei stick-held dip net fishery during 2007-2023. Mean size frequency distributions for each classified region are shown in the right columns.