DRAFT

INTER-AMERICAN TROPICAL TUNA COMMISSION

EXTERNAL REVIEW OF IATTC YELLOWFIN TUNA ASSESSMENT

La Jolla, California (USA)
15-19 October 2012

DOCUMENT YFT-01-02

EXPLORING LARGE-SCALE PATTERNS IN YELLOWFIN TUNA DATA FROM DOLPHIN SETS IN THE EASTERN PACIFIC OCEAN PURSE-SEINE FISHERY

Cleridy E. Lennert-Cody, Mark N. Maunder, Alexandre Aires-da-Silva

CONTENTS

1. BACKGROUND

Stratification is used in stock assessment to address differences in stock and fishery dynamics. In general, fisheries data (catch, catch per unit of effort (CPUE), and age/size-composition data) are stratified after data collection to support the assumption that fishery-related parameters, catchability and selectivity, are constant over time. Presently, stock assessments for all tuna species in the eastern Pacific Ocean (EPO) (*e.g.*, Aires-da-Silva and Maunder 2010; Aires-da-Silva and Maunder 2012a; Aires-da-Silva and Maunder 2012b) use large areas formed by aggregating the spatial strata of the IATTC port-sampling program (*e.g.*, Figure 1; Suter 2010; Tomlinson 2004). Although these sampling strata were refined in the 1990s (Suter 2010), they were primarily developed in the late 1960s (Suter 2010, and references therein) when the purse-seine fishery was more coastal (Watters 1999). This document describes progress on the analysis of large-scale patterns in yellowfin fishery data from sets on dolphin-associated yellowfin by large (≥ 364 metric tons (t) fish-carrying capacity) purse-seine vessels for the purpose of developing other options for defining areas for stock assessment strata.

2. DATA AND DATA PROCESSING

The general analytic approach taken in this work is to try to identify similar large-scale structures in yellowfin length-frequency distributions and CPUE trends. This requires that these distributions and trends be estimated on a fine-scale spatial-temporal grid throughout the EPO. The spatial-temporal grid selected was 5º latitude by 5º longitude by quarter (January-March; April-June; July-September; October-December), based on the spatial-temporal resolution of the IATTC port sampling data (see below) and the time step used in the stock assessment model

(quarterly). Each unique 5º latitude by 5º longitude by quarter-of-the-year will be referred to as a "grid cell."

Length-frequency data

Data on the species and size composition of the catches of tuna by purse-seine vessels are collected when vessels arrive in port to unload (Tomlinson 2004; Suter 2010). Samples are collected according to a 'two-stage' approach, where the wells of a vessel are the first stage, and the fish within a well are the second stage. Due to logistic constraints vessel wells to be sampled are selected opportunistically. However, a well is sampled only if all the catch it contains is from the same area (Figure 1), month and fishing mode. For yellowfin tuna in dolphin sets, sampling since 2000 has generally been proportional to the level of fishing effort and catch (Lennert-Cody *et al*. 2012). Once a well of a vessel has been selected to be sampled, individual fish are sampled from the well as the catch is unloaded. A number of fish of each species (typically 50) are measured for length (tip of snout to fork of tail) to the nearest millimeter. The 1-mm length measurements are then grouped into 1-cm intervals. Depending on the port of unloading, catches may be sorted by species and weight category during unloading before the fish are accessible to IATTC staff for sampling ('sorted' unloadings). Further details of the port-sampling data collection procedures can be found in the appendix of Suter (2010). Additional information on the fishing location (5º latitude and 5º longitude) and date of fishing (month, year) of the sets that went into the sampled well, and the total well catch (all three tuna species combined) is also available.

Samples used in this analysis were limited to those from purse-seine sets on tunas associated with dolphins ('dolphin' sets) made by large vessels (vessels with $\geq$ 364 mt fish-carrying capacity) for years 2000-2011. There were a total 2611 such samples, with an average of just over 200 samples per year (range: 122-321 samples per year).

Length-frequency distributions

The yellowfin tuna length-frequncy sample data were processed in the following manner prior to analysis. First, each yellowfin tuna sample was first used to estimate the length composition of the total well catch of yellowfin tuna (following methods of Tomlinson 2004), and all subsequent analyses were done using these 'raised' samples. In this way, samples from sorted and unsorted unloadings could be treated similarly. For each sorted sample, the proportion of yellowfin in each 1 cm length interval was computed and multiplied by the estimated number of fish in the sort. Estimated numbers of fish by sort were then summed across sorts to obtain an estimate of the total number of yellowfin tuna in the well in each 1 cm interval. For unsorted samples, the proportion of yellowfin tuna in each 1 cm interval was multiplied by the estimate of the total number of yellowfin tuna in the well.

Second, to be consistent with the yellowfin tuna stock assessment model, which has a quarterly time step, the 1 cm length intervals were 'grown' or 'shrunk' to the middle month of each quarter-of-the-year by adding or subtracting a monthly length increment, where applicable (the middle month of each quarter requires no adjustment). It was assumed that, from year to year for the same quarter, the length composition remained stable, however, within quarters the length adjustment was necessary because length-frequency samples taken from the same population but in different months of the same quarter could appear to represent different populations due solely to growth. The monthly length increments used to grow or shrink fish were obtained from the

Gompertz growth model of Wild (1986).

Finally, for each sample, counts of fish in 1 cm length intervals were grouped into the following 11 larger intervals: ≤ 58 cm, 59-69 cm, …, 136-146 cm, 147-159 cm, and ≥ 160 cm. Gaps created in the length-frequency distributions by growing/shrinking the length intervals were taken into consideration when constructing these intervals. Within each grid cell, monthly samples were treated as repeat observations for the quarter of the grid cell. Quarterly summaries of the processed length-frequency data are shown in Figures 2-5. Larger fish tended to be caught more frequently in the southern and offshore areas of the EPO.

Catch and effort data

Catch and effort data for 1975-2011were obtained from observer and logbook data bases,. The observer and logbook data bases also have information on the date and location of fishing. The species catches in these data bases are those estimated by observers or navigators and fishing captains. Effort, estimated in number of days fishing, was computed using the method of Maunder *et al*. (2010). The number of days fished by set type was estimated from a multiple regression of total days fished against the number of sets by set type. Nominal catch per day fishing ("CPD") was then estimated for each 5º square area by month as the sum of catches divided by the sum of days fished. Within each grid cell, monthly CPD values of the same year were treated as repeat observations for the quarter of the grid cell. The CPD data within each grid cell are shown in Figures 6-9. There is a considerable amount of variability in CPD within most grid cells.

Trends in CPD within each grid cell were summarized following the method outlined in the appendix of Lennert-Cody *et al*. (*In press*). Trends were only estimated for grid cells with sufficient data, defined as those grid cells with at least 50 data points over 25 years and at least 0.01% of the total yellowfin catch in the data base. There are 137 grid cells that met this requirement. For each grid cell with sufficient data, the temporal CPD trend over years was estimated using penalized cubic regression splines (Wood 2006). Spline smooths for all grid cells were based on the following set-up: 6 basis functions, knots at years 1975, 1982, 1989, 1997, 2004 and 2011, and a smoothing parameter value of 2.64. A square root transformation was first applied to the CPD values to try to help the data conform to the assumption of equal variance. Smooth trends are shown in Figures 6-9. Along the main east-west axis of the fishery there is sometimes an indication that trends in CPD peaked in the middle part of the 37-year time series and that it has increased towards the end of the period in some inshore areas. However, there is considerable variability about the estimated trends. The fitted smooth model of each grid cell was used to predict a time series of annual CPD values (on the scale of the square root) within the grid cell. The first-differenced vector of these annual CPD estimates was used to summarize the trend in CPD by grid cell. The vector of first-differences was used because it is the trends in CPD that are of interest, not the absolute magnitude of CPD.

3. METHODS OF ANALYSIS

Large-scale spatial-temporal pattern in the length-frequency distributions and in the CPD trends was explored using tree-based methods. All tree analyses used the 5° latitude, the 5° longitude and the quarter-of-the-year as predictor variables (all treated as numeric). Cyclic combinations of quarters (*e.g.*, October-December and January-March *versus* April-June and July-September)

were also considered for the quarter variable. Five tree analyses were conducted: 1) an analysis of only length-frequency distributions ("length-frequency" analysis); 2) an analysis of only CPD trends ("unweighted trends" analysis); 3) an analysis of only CPD trends, taking into consideration variability about the estimated trends ("variance-weighted" trends analysis); 4) an analysis of length-frequency distributions and CPD trends, simultaneously ("unweighted simultaneous" analysis); and 5) an analysis of length-frequency distributions and CPD trends, simultaneously, taking into consideration variability about the estimated trends ("variance-weighted simultaneous" analysis). The tree-based methods attempt to subdivide the data into smaller and smaller subgroups that are more homogeneous, based on the predictor variables values. Details of the methodology can be found in Lennert-Cody *et al*. (2010) and Lennert-Cody *et al*. (*In press*).

4. RESULTS AND DISCUSSION

Overall, the results of the tree analyses (Figures 10-14) show some agreement with the current stock assessment areas (Figure 1). Just as in the stock assessment stratification, a common feature of the tree results is that there is a tendency to divide the offshore region of the EPO into northern and southern areas near the equator, and also to separate the inshore region of the EPO from offshore. However, the simultaneous tree results suggest some modifications to the current spatial stratification could be considered (*e.g*., compare Figure 1 with Figure 14). For example, the tree analyses indicate differences in the inshore area, south and north of 5ºN, and that the size of the inshore area in the south might be increased slightly westward. In addition, the boundary between the inshore and offshore areas around the main east-west axis of the fishery (along ~10º-15ºN) could be revised to expand the inshore area. On the other hand, although the tree analyses indicate that data from north of 20ºN appear different, it may not be practical to create such a small stratum in the stock assessment model.

The tree partitions from the two simultaneous analyses (Figures 13-14) are defined only by spatial variables suggesting that large-scale spatial structure could be more important in these data than temporal (quarterly) structure. However, for the CPD data the analyses show a competition among spatial partitions (latitude and longitude) and those based on quarter (Tables 1-2; Figures 11-12), which does not happen for the length-frequency data (Tables 1-2; Figure 10). This may reflect unavoidable confounding of spatial and seasonal effects in the CPD data or indicate that a revised treatment of the CPD data is necessary. For example, there is considerable variability in CPD within grid cells (Figures 5-8). The CPD trees from unweighted and variance-weighted analyses (Figures 11-12) show different structure due to the variance-weighting, which will down weight grid cells in the analysis where there is greater variability about the estimated trends. In contrast to the differences between the unweighted and variance weighted CPD trees, the unweighted and the variance-weighted simultaneous trees are nearly identical (Figures 13-14), which reflects the influence of the length-frequency data on the simultaneous tree analyses. It would be worthwhile to explore other trend models for CPD (e.g., other transformations, distributional models) or perhaps aggregate CPD to the quarter before fitting the smooth trend models, and to consider other relative indices such as catch-per-set.

Several sensitivity analyses need to be conducted. First, growing/shrinking the length-frequency bins introduces variability which was not accounted for in the tree analyses. Previous analysis of data from 2003-2007 (Lennert-Cody, unpublished) did not find major differences in the main

splits when unadjusted monthly length-frequency data were used, but this analysis needs to be repeated with the 2000-2011 data. Second, previous analysis of the data for 2003-2007 showed some year to year variability (Lennert-Cody *et al*. 2010). However, it may be difficult to separate inter-annual spatial variability from sampling variability when analyzing the length-frequency data of individual years. Therefore, repeating the tree analyses with length-frequency data from groups of years, where the groups of years are defined based on differences in EPO-wide environmental forcing (*e.g*., El Niño, La Niña) could be useful. Finally, sensitivity of the current partition definitions to data of grid cells at the margins of the fishery region needs to be explored by running the tree analysis with different definitions of sufficient data.

ACKNOWLEDGEMENTS

Special thanks to Christine Patnode for help with graphics and Nickolas Vogel for data base assistance.

REFERENCES

Aires-da-Silva, A. and Maunder, M.N. 2010. Status of bigeye tuna in the eastern Pacific Ocean in 2008 and outlook for the future. Inter-American Tropical Tuna Commission Stock Assessment Report 10, pages 116-228.

Aires-da-Silva, A. and Maunder, M.N. 2012a. Status of yellowfin tuna in the eastern Pacific Ocean in 2011 and outlook for the future. Inter-American Tropical Tuna Commission Stock Assessment Report 12, pages 3-110.

Aires-da-Silva, A. and Maunder, M.N. 2012b. Status of bigeye tuna in the eastern Pacific Ocean in 2010 and outlook for the future. Inter-American Tropical Tuna Commission Stock Assessment Report 12, pages 111-122.

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

Lennert-Cody, C.E. Maunder, M.N., Tomlinson, P.K., Aires-da-Silva, A., Pérez, A. 2012. Progress report on the development of poststratified estimators of total catch for the purse-seine fishery port-sampling data. IATTC Document SAC-03-10, 3rd Meeting of the Scientific Advisory Committee, La Jolla, California, USA, 15-18 May, 2012.

Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., Minami, M. *In press*. Defining population spatial units: simultaneous analysis of frequency distributions and times series. Fisheries Research.

Maunder, M.N., Lennert-Cody, C.E., Aires-da-Silva, A., Bayliff, W.H., Tomlinson, P.K., and Schaefer, K.M. 2010. Summary of data available for bigeye tuna in the eastern Pacific Ocean

and its use in stock assessment. Document BET-01-07. External Review of IATTC Bigeye Tuna Assessment, 3-7 May, 2010, La Jolla, California, U.S.A.

Suter, J.M. 2010. An evaluation of the area stratification used for sampling tunas in the eastern Pacific Ocean and implications for estimating total annual catches. Inter-American Tropical Tuna Commission Special Report 18.

Tomlinson, P.K. 2004. Sampling the tuna catch of the eastern Pacific Ocean for species composition and length-frequency distributions. Inter-American Tropical Tuna Commission Stock Assessment Report 4, pages 311-333.

Watters, G.M. 1999. Geographical distributions of effort and catches of tunas by purse-seine vessels in the eastern Pacific Ocean during 1965-1998. IATTC Data Report 10.

Wild. A. 1986. Growth of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean based on otolith increments. Inter-American Tropical Tuna Commission Bulletin 18, 421-482.

Wood, S.N. 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC. 391 pp.

Table 1. Tree details for the unweighted simultaneous analysis (tree shown in Figure 13). At each of steps (a)-(d), the scaled improvement in reducing data heterogeneity achieved by each candidate partition of each data type (1.0 = greatest improvement) for the four 'best' candidates (ranks in parentheses, from 'best'= 1), and the ranks of the first few candidate splits for the final tree (from 'best' = 1), are provided. Step (a) shows the results from the initial partition of the full data set. For steps (b)-(d), the region that is to be partitioned at that step is indicated. For example, the two sub-regions created by the partition at 5ºN selected in step (a) are further partitioned in step (b) (north of 5ºN) and step (c) (south of 5ºN). In addition, to provide context for the split ranks, the number of possible split-variable values at each step is provided in parentheses after the region to be partitioned (*e.g.*, at step (b), there are 22 possible split-variable values).

	Scaled improvement length-frequency (split rank)	Scaled improvement CPD trends (split rank)	Simultaneous tree split rank
(a) Full data set (26)			
Latitude 20°N	0.537 (3)		4
Latitude 15°N	0.531 (4)		
Latitude 10°N	0.638 (2)	(9)	2
Latitude 5°N	1.000 (1)	0.820 (4)	1
Latitude 0°	(5)	(7)	3
Longitude 120°W		(5)	
Quarters1; 2-4		1.000 (1)	
Quarters 1-2; 3-4		0.892 (2)	
Quarters 1-3; 4		0.847 (3)	
(b) North of 5°N (22)			
Latitude 20°N	0.960 (2)	1.000 (1)	1
Latitude 15°N	(5)	(5)	2
Longitude 125°W	0.857 (4)		
Longitude 120°W	0.924 (3)		4
Longitude 115°W	1.000 (1)		3
Longitude 100°W		0.872 (3)	
Longitude 95°W		0.929 (2)	
Quarter 1; 2-4		0.711 (4)	
(c) South of 5°N (17)			
Longitude 120°W		1.000 (1)	2
Longitude 115°W		0.805 (3)	
Longitude 110°W		0.550 (5)	

Longitude 105°W	(5)		
Longitude 100°W	0.560 (4)		
Longitude 95°W	0.946 (2)	(10)	1
Longitude 90°W	0.563 (3)		
Longitude 85°W	1.000 (1)		
Quarter 1-2; 3-4		0.809 (2)	4
Quarter 1-3; 4		0.781 (4)	3
(d) Between 5°N and 20°N (20)			
Latitude 10°N		1.000 (1)	2
Longitude 130°W	(5)		
Longitude 125°W	0.896 (4)		
Longitude 120°W	0.942 (2)	0.769 (8)	1
Longitude 115°W	1.000 (1)		3
Longitude 110°W	0.929 (3)		
Longitude 100°W		(6)	
Longitude 95°W		0.926 (3)	
Longitude 90°W		(7)	
Quarter 1; 2-4		0.924 (4)	
Quarter 1-2; 3-4		0.951 (2)	4
Quarter 1-3; 4		(5)	

Table 2. Tree details for the variance-weighted simultaneous analysis (tree shown in Figure 14). At each of steps (a)-(d), the scaled improvement in reducing data heterogeneity achieved by each candidate partition of each data type (1.0 = greatest improvement) for the four 'best' candidates (ranks in parentheses, from 'best'= 1), and the ranks of the first few candidate splits for the final tree (from 'best' = 1), are provided. Step (a) shows the results from the initial partition of the full data set. For steps (b)-(d), the region that is to be partitioned at that step is indicated. For example, the two sub-regions created by the partition at 5ºN selected in step (a) are further partitioned in step (b) (north of 5ºN) and step (c) (south of 5ºN). In addition, to provide context for the split ranks, the number of possible split-variable values at each step is provided in parentheses after the region to be partitioned (*e.g.*, at step (b), there are 22 possible split-variable values).

	Scaled improvement length-frequency (split rank)	Scaled improvement CPD trends (split rank)	Simultaneous tree split rank
(a) Full data set (26)			
Latitude 20°N	0.537 (3)	1.000 (1)	2
Latitude 15°N	0.531 (4)		
Latitude 10°N	0.638 (2)	0.723 (4)	3
Latitude 5°N	1.000 (1)	(9)	1
Latitude 0°	(5)		
Longitude 115°W		0.851 (2)	4
Quarters1; 2-4		(5)	
Quarters 1-2; 3-4			
Quarters 1-3; 4		0.782 (3)	
(b) North of 5°N (22)			
Latitude 20°N	0.960 (2)	1.000 (1)	1
Latitude 15°N	(5)	0.600 (4)	3
Latitude 10°N		0.598 (3)	
Longitude 125°W	0.857 (4)		
Longitude 120°W	0.924 (3)		4
Longitude 115°W	1.000 (1)	(5)	2
Longitude 100°W			
Longitude 95°W			
Quarter 1-3; 4		0.629 (2)	
(c) South of 5°N (17)			
Longitude 120°W		0.891 (2)	
Longitude 115°W		(5)	

Longitude 110°W			
Longitude 105°W	(5)		
Longitude 100°W	0.560 (4)	0.723 (4)	2
Longitude 95°W	0.946 (2)	(6)	1
Longitude 90°W	0.563 (3)		
Longitude 85°W	1.000 (1)		
Quarter 1-2; 3-4		0.751 (3)	4
Quarter 1-3; 4		1.000 (1)	3
(d) Between 5°N and 20°N (20)			
Latitude 10°N		1.000 (1)	2
Longitude 125°W	0.896 (4)		
Longitude 120°W	0.942 (2)	(6)	3
Longitude 115°W	1.000 (1)	(5)	1
Longitude 110°W	0.929 (3)		
Longitude 100°W			
Longitude 95°W			
Longitude 90°W			
Quarter 1; 2-4		0.609 (4)	
Quarter 1-2; 3-4		0.705 (2)	4
Quarter 1-3; 4		0.690 (3)	

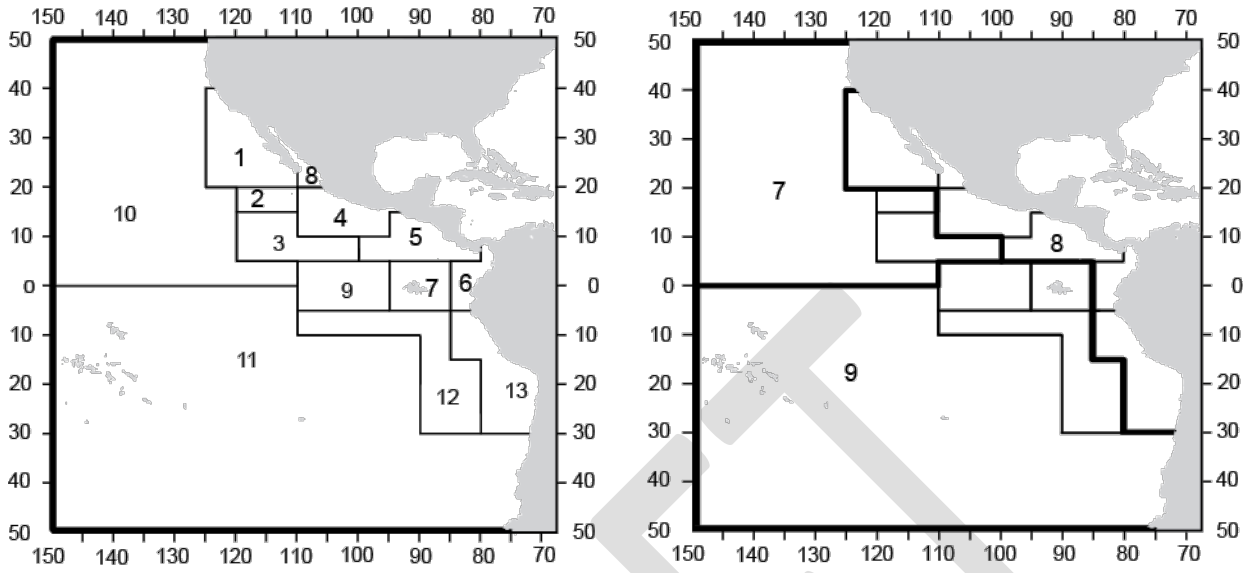


Figure 1. Sampling areas (left) and yellowfin tuna stock assessment areas for dolphin sets (right).

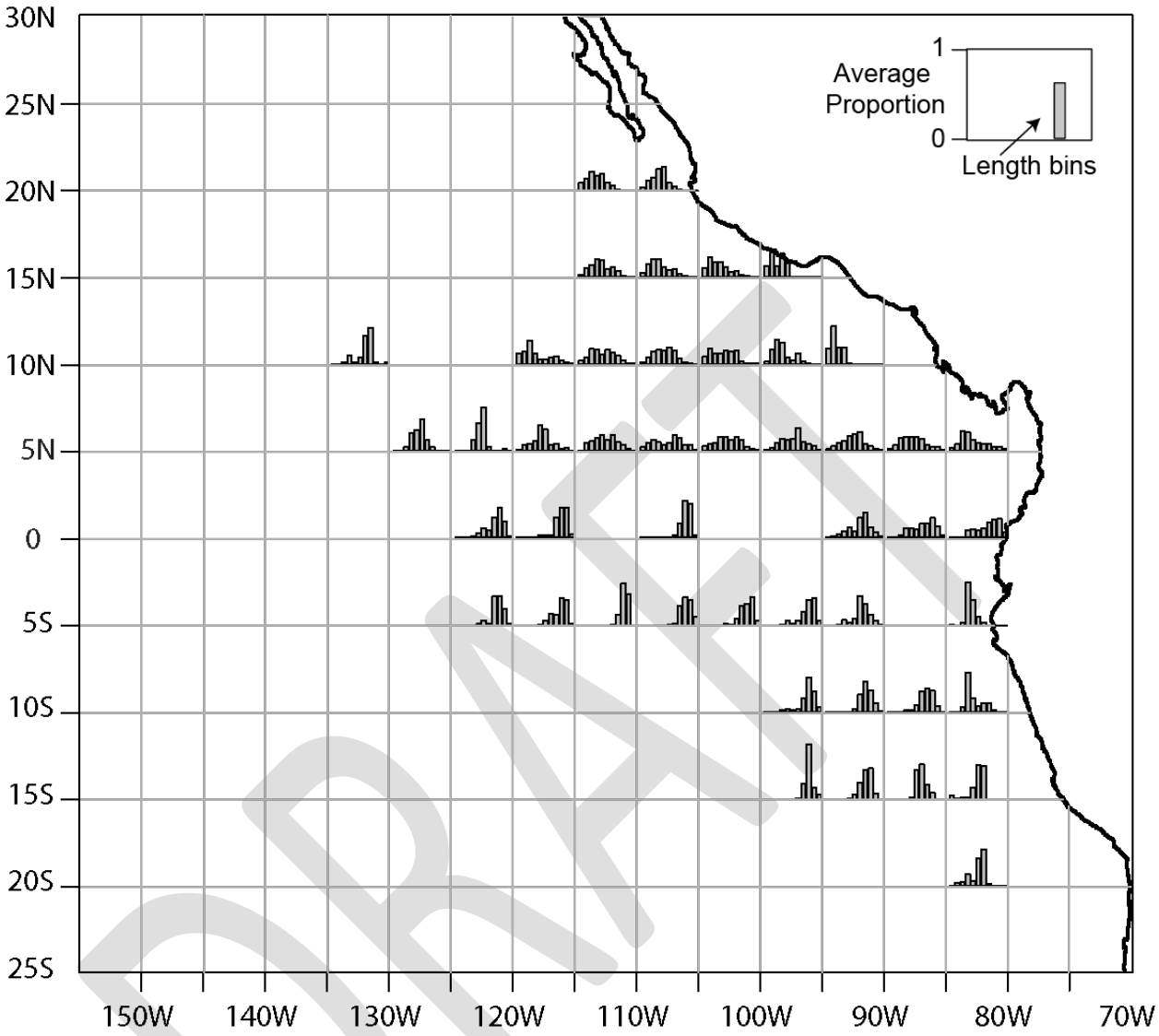


Figure 2. Summary of the yellowfin tuna binned length-frequency sample data for the 1st quarter-of-the-year. The values shown are the average proportion of fish in each binned length interval, where the average was computed over samples from months and years of the same grid cell. The ranges of the x- and y-axes are the same; the y-axis for the length-frequency data ranges from 0-1.

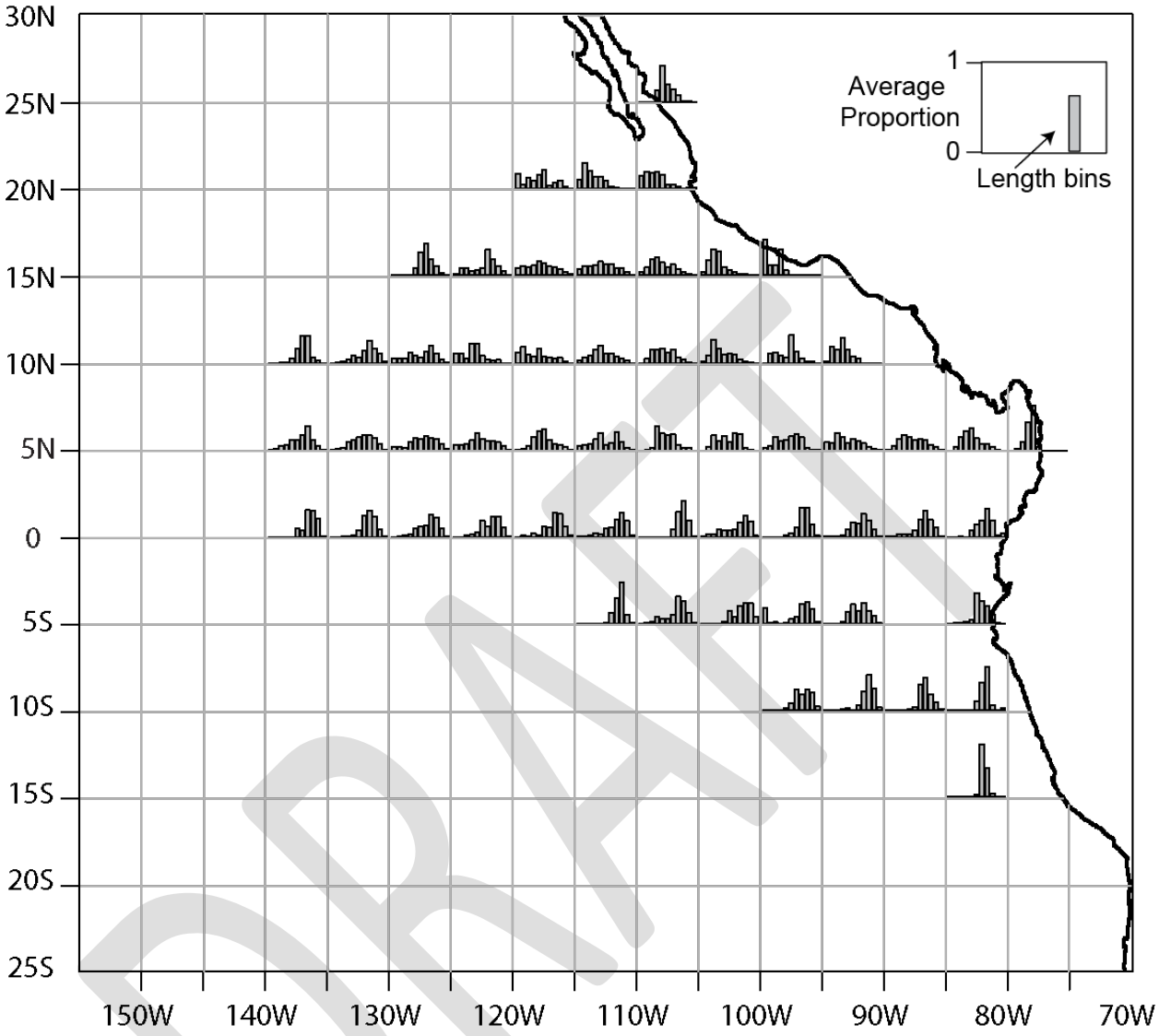


Figure 3. Summary of the yellowfin tuna binned length-frequency sample data for the 2nd quarter-of-the-year. The values shown are the average proportion of fish in each binned length interval, where the average was computed over samples from months and years of the same grid cell. The ranges of the x- and y-axes are the same; the y-axis for the length-frequency data ranges from 0-1.

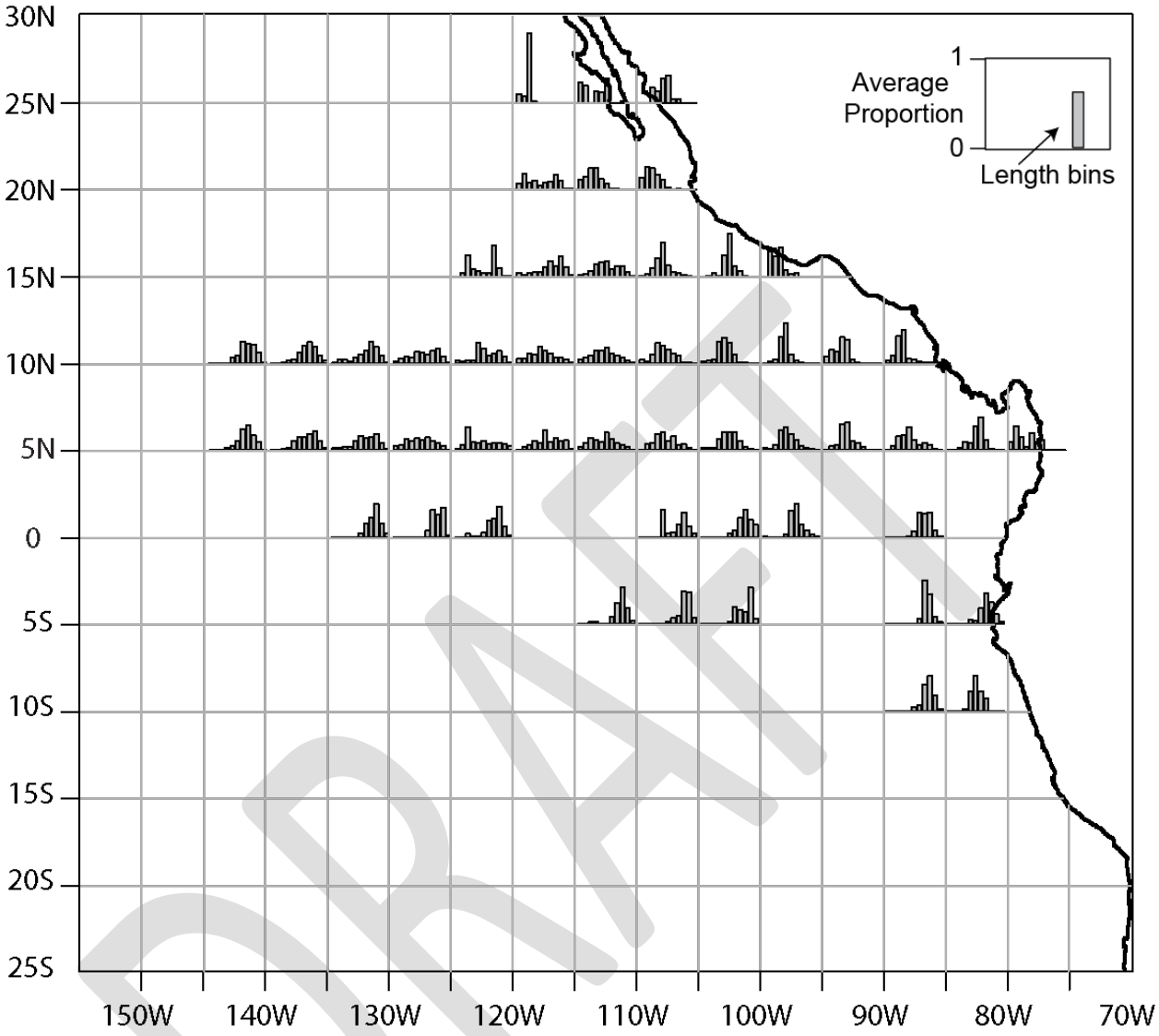


Figure 4. Summary of the yellowfin tuna binned length-frequency sample data for the 3rd quarter-of-the-year. The values shown are the average proportion of fish in each binned length interval, where the average was computed over samples from months and years of the same grid cell. The ranges of the x- and y-axes are the same; the y-axis for the length-frequency data ranges from 0-1.

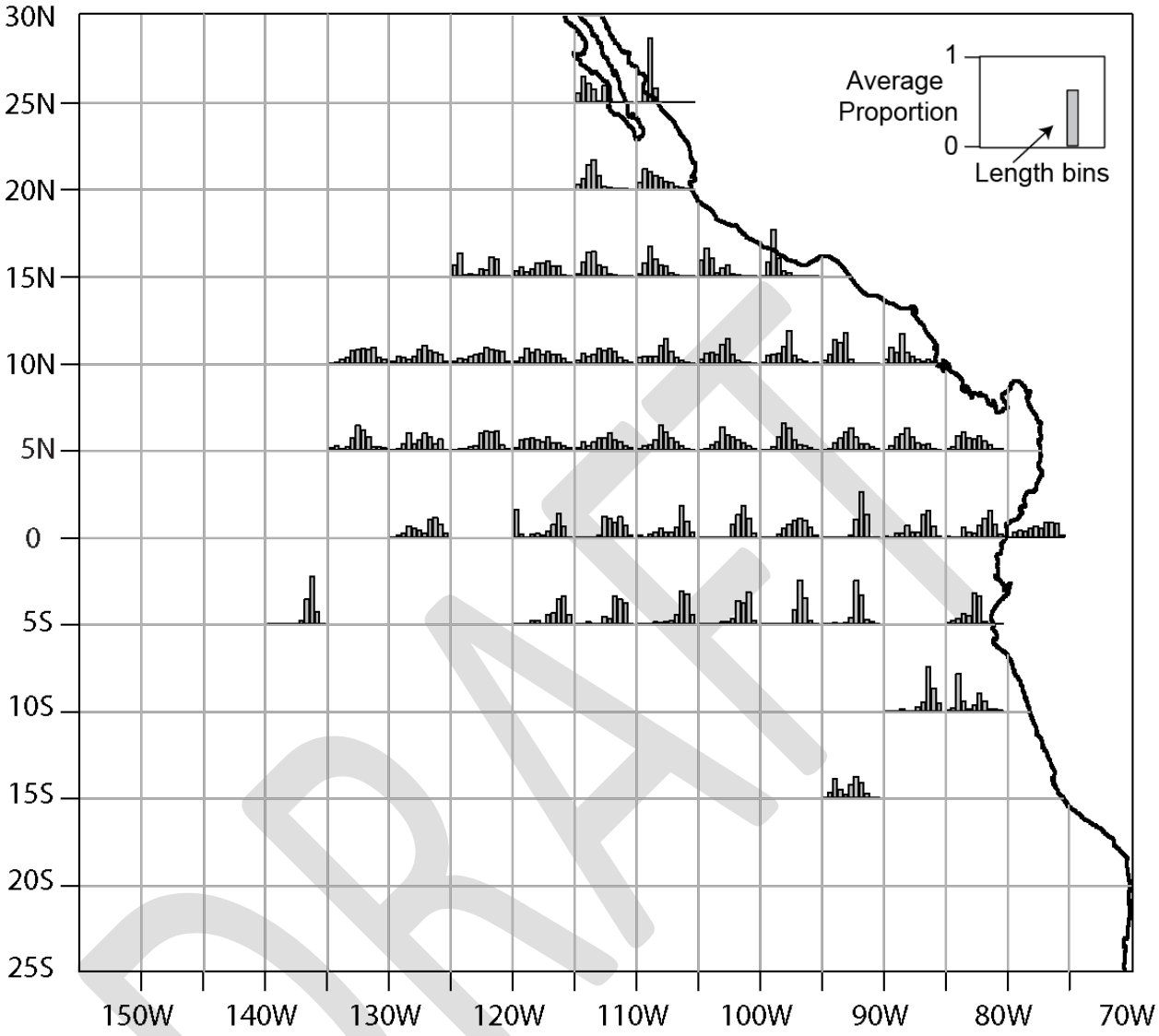


Figure 5. Summary of the yellowfin tuna binned length-frequency sample data for the 4th quarter-of-the-year. The values shown are the average proportion of fish in each binned length interval, where the average was computed over samples from months and years of the same grid cell. The ranges of the x- and y-axes are the same; the y-axis for the length-frequency data ranges from 0-1.

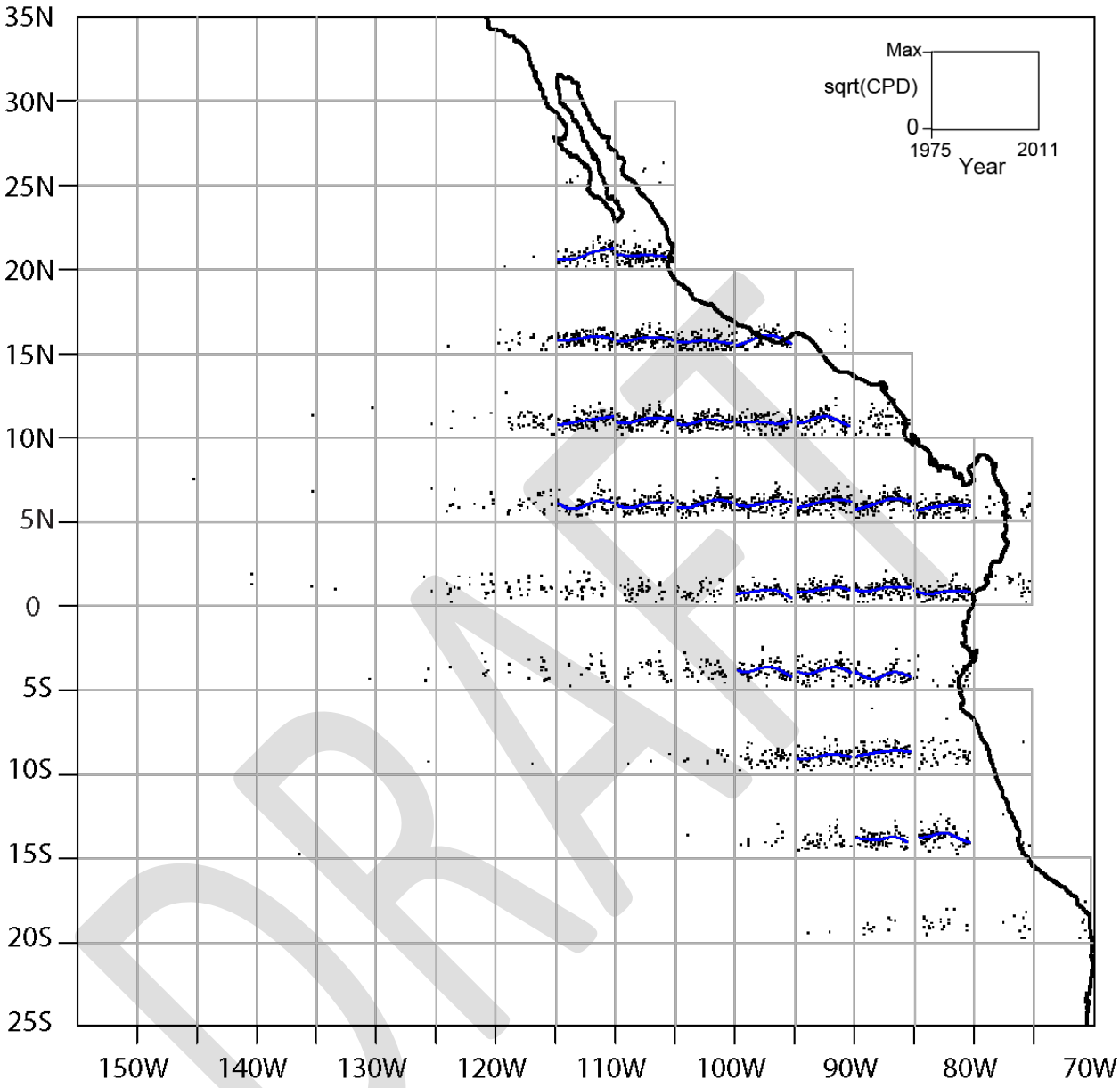


Figure 6. CPD trend data for the 1st quarter-of-the-year. The black dots are the data points (monthly CPD in the 1st quarter), and for grid cells with sufficient data, the blue lines show the smooth trends. The ranges of the x- and y-axes are the same; the x-axis for the CPD data ranges from 1975-2011 and the y-axis ranges from 0 to the maximum(sqrt(CPD)) value, with the maximum taken over all grid cells (*i.e.*, over all areas and quarters). 'sqrt' = square root.

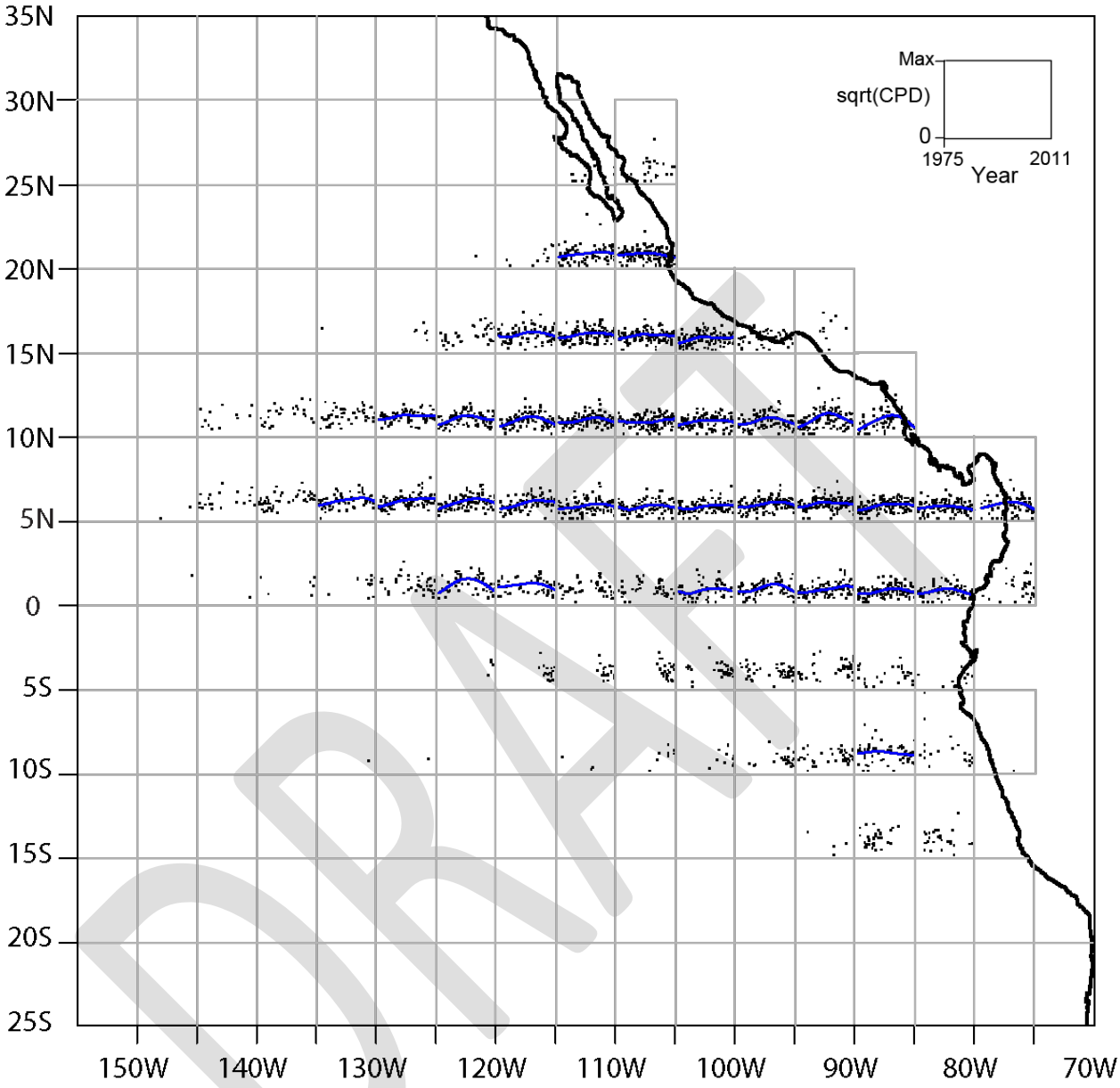


Figure 7. CPD trend data for the 2nd quarter-of-the-year. The black dots are the data points (monthly CPD in the 2nd quarter), and for grid cells with sufficient data, the blue lines show the smooth trends. The ranges of the x- and y-axes are the same; the x-axis for the CPD data ranges from 1975-2011 and the y-axis ranges from 0 to the maximum(sqrt(CPD)) value, with the maximum taken over all grid cells (*i.e.*, over all areas and quarters). 'sqrt' = square root.

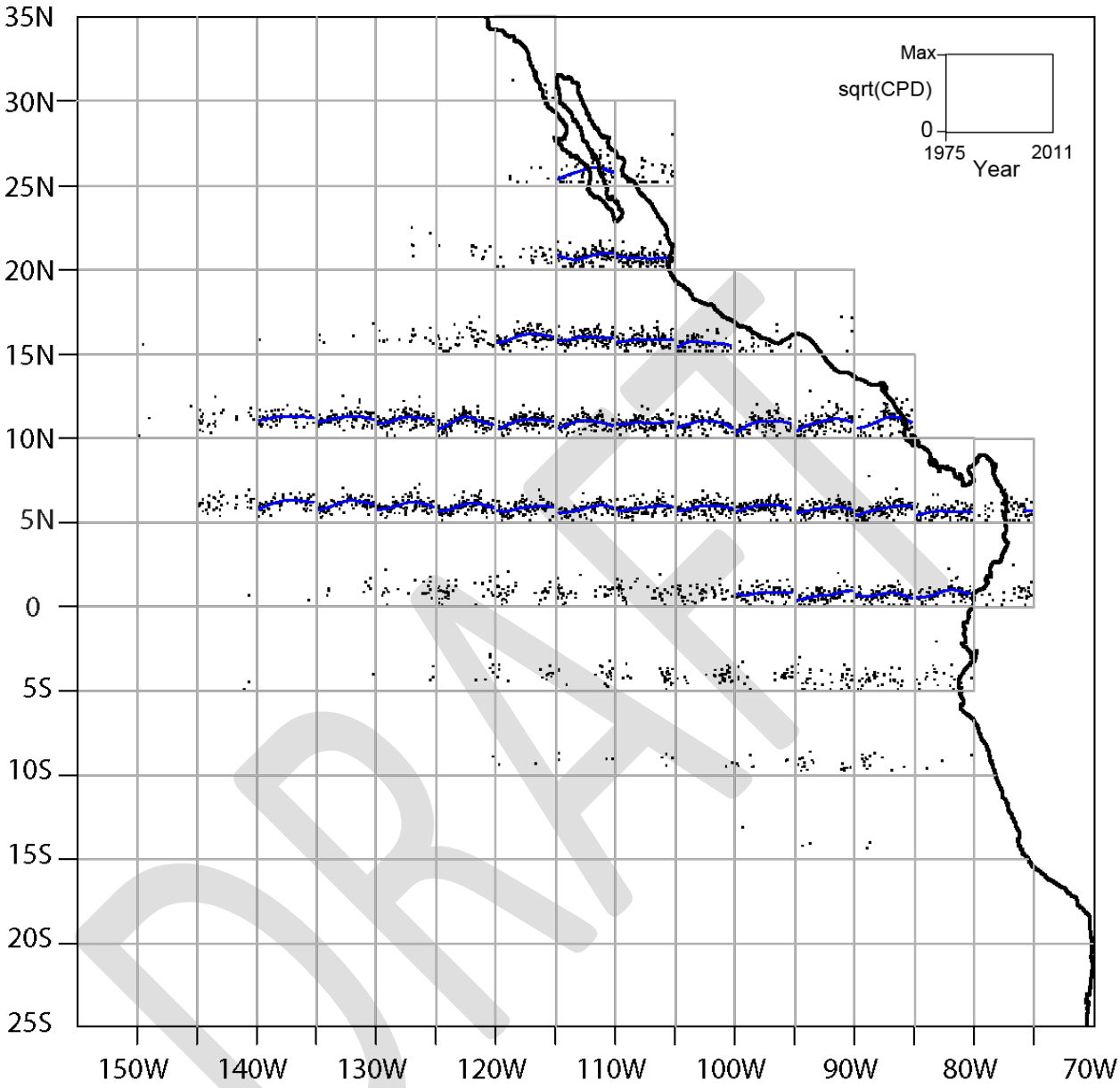


Figure 8. CPD trend data for the 3rd quarter-of-the-year. The black dots are the data points (monthly CPD in the 3rd quarter), and for grid cells with sufficient data, the blue lines show the smooth trends. The ranges of the x- and y-axes are the same; the x-axis for the CPD data ranges from 1975-2011 and the y-axis ranges from 0 to the maximum(sqrt(CPD)) value, with the maximum taken over all grid cells (*i.e.*, over all areas and quarters). 'sqrt' = square root.

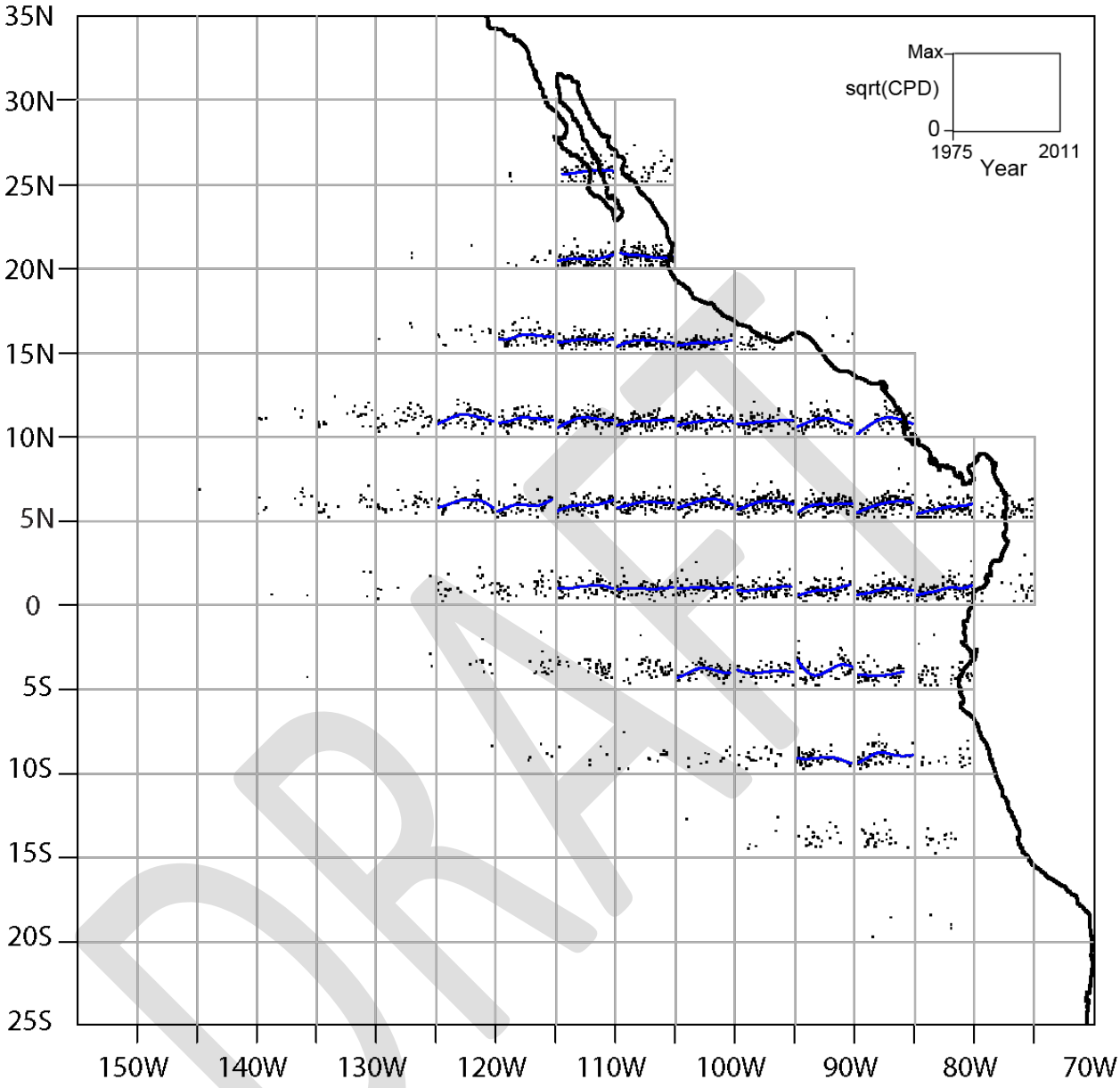
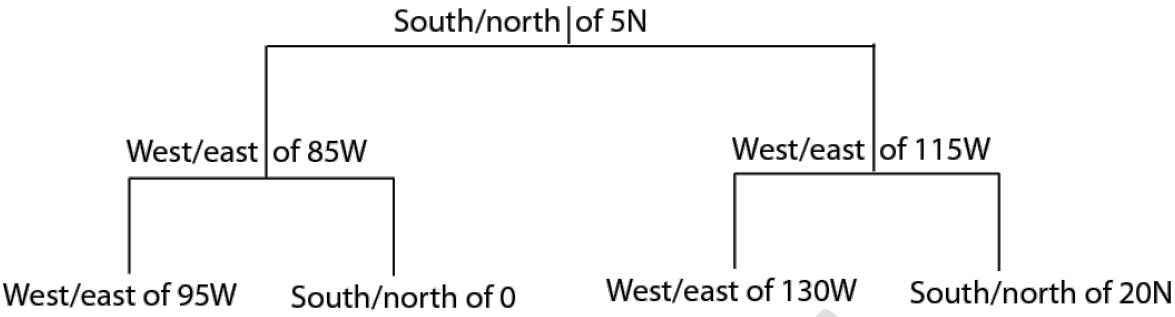


Figure 9. CPD trend data for the 4th quarter-of-the-year. The black dots are the data points (monthly CPD in the 4th quarter), and for grid cells with sufficient data, the blue lines show the smooth trends. The ranges of the x- and y-axes are the same; the x-axis for the CPD data ranges from 1975-2011 and the y-axis ranges from 0 to the maximum(sqrt(CPD)) value, with the maximum taken over all grid cells (*i.e.*, over all areas and quarters). 'sqrt' = square root.



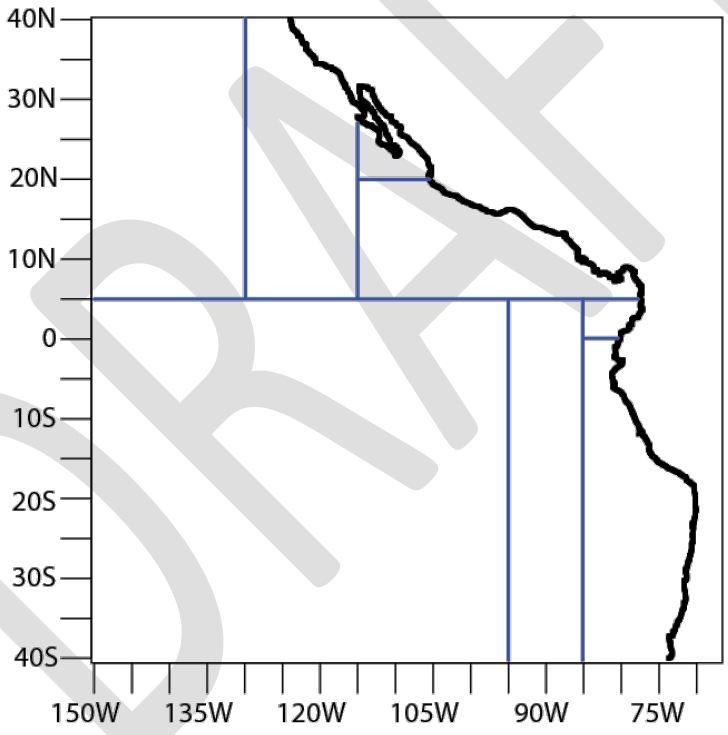


Figure 10. Tree produced from the length-frequency analysis. All left branches include the end point; branch length is displayed as uniform. *: no further splits explored.

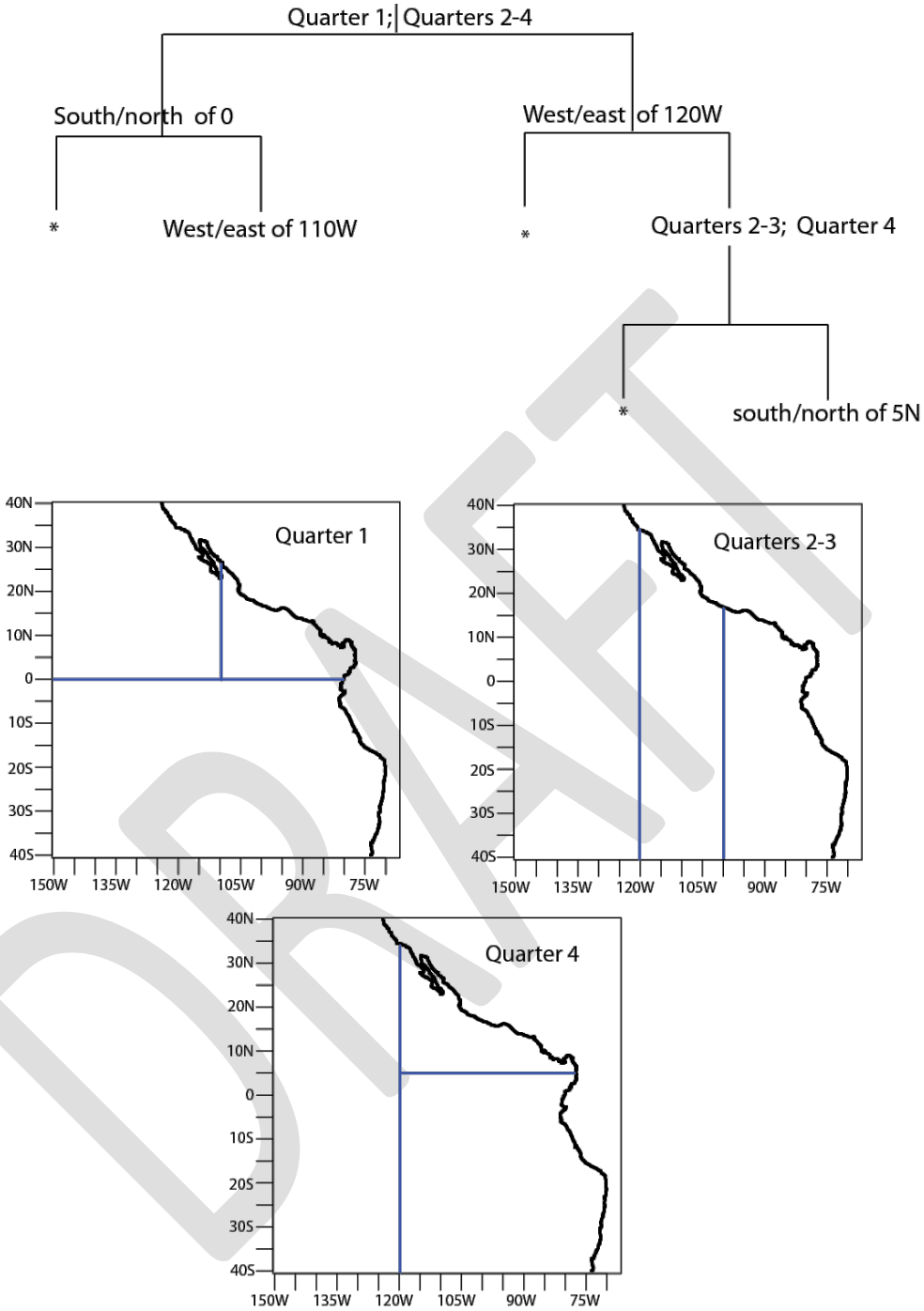


Figure 11. Tree produced from the unweighted trends analysis. All left branches include the end point; branch length is displayed as uniform. *: no further splits explored.

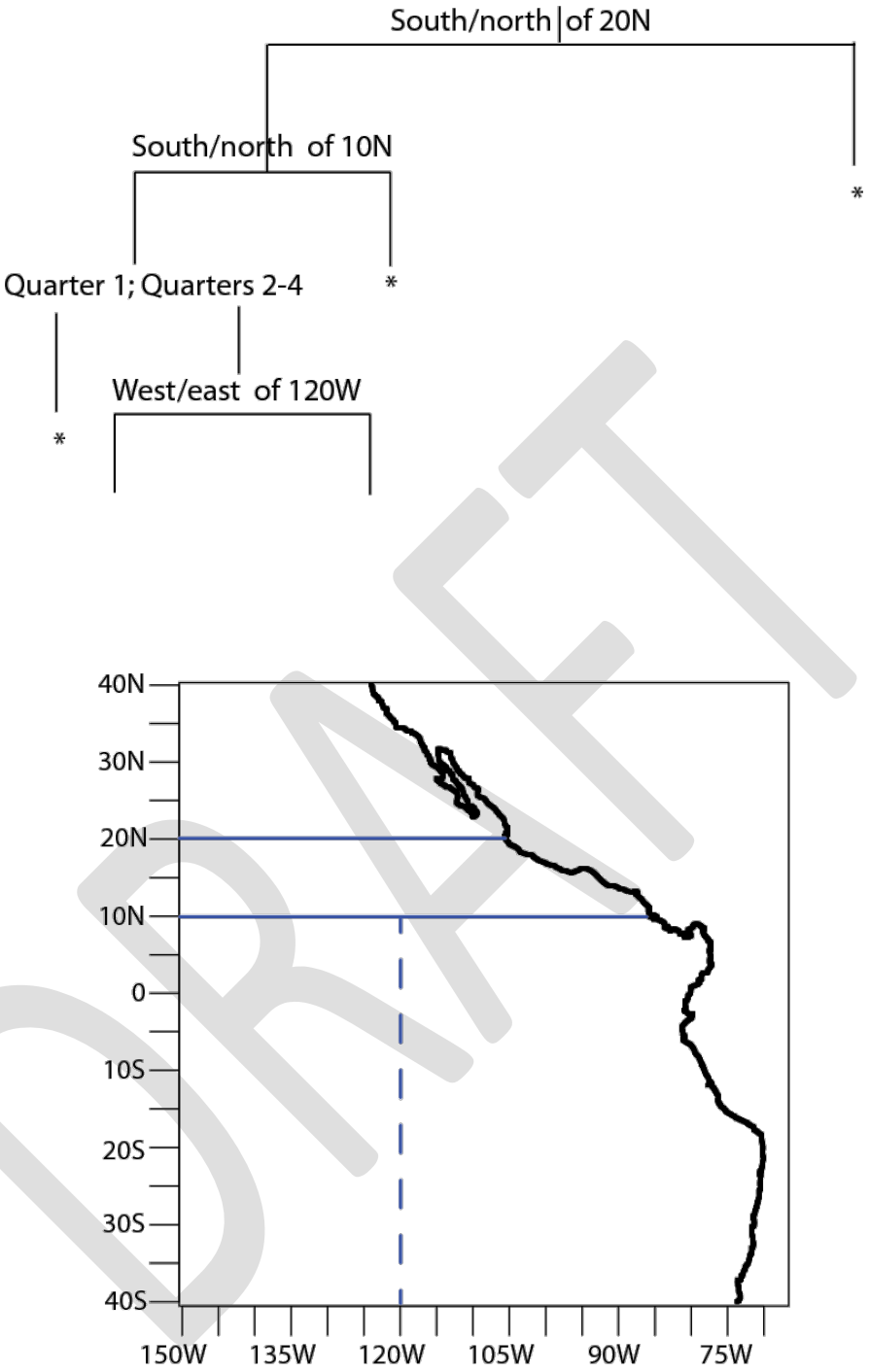


Figure 12. Tree produced from the variance-weighted trends analysis. All left branches include the end point; branch length is displayed as uniform. *: no further splits explored.

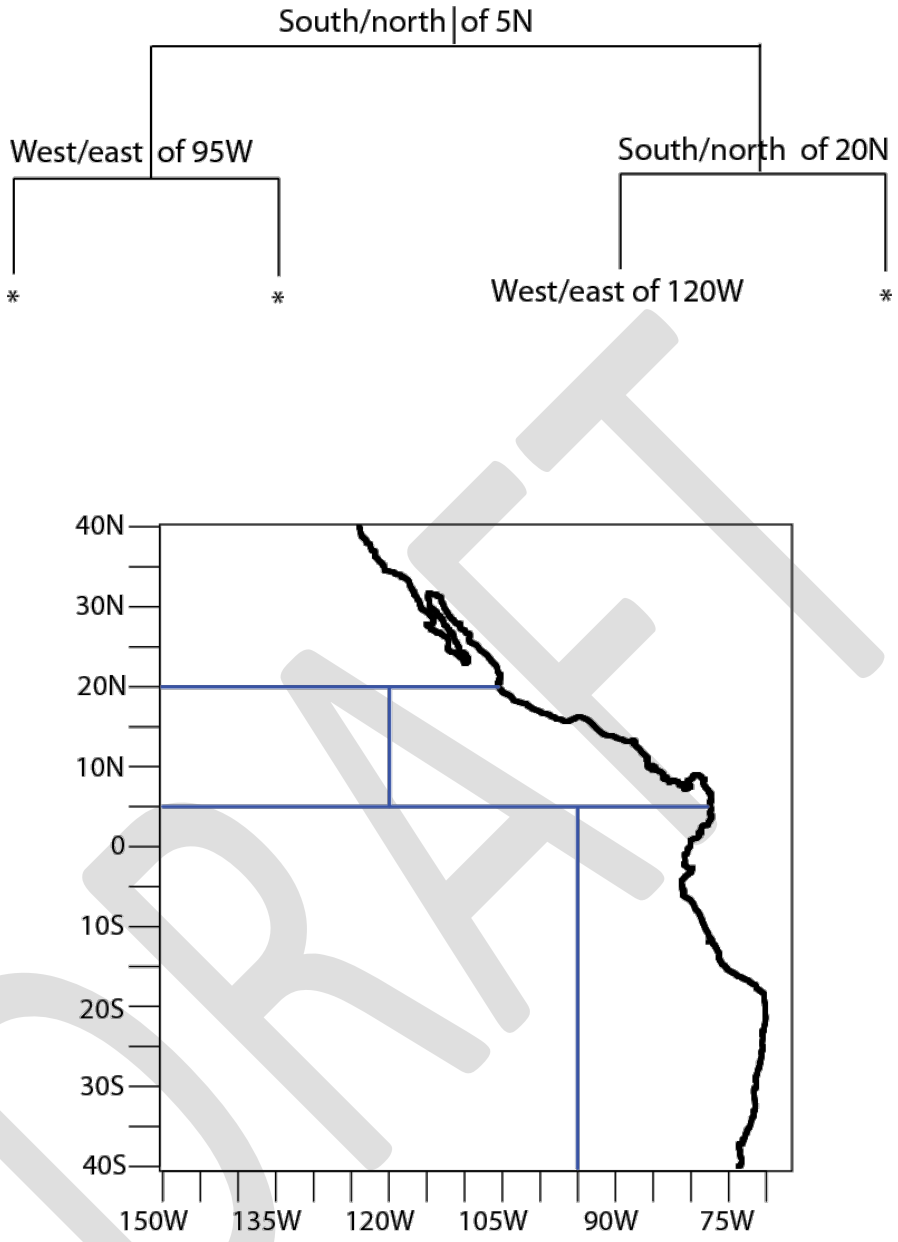


Figure 13. Tree produced by the unweighted simultaneous analysis. All left branches include the end point; branch length is displayed as uniform. *: no further splits explored.

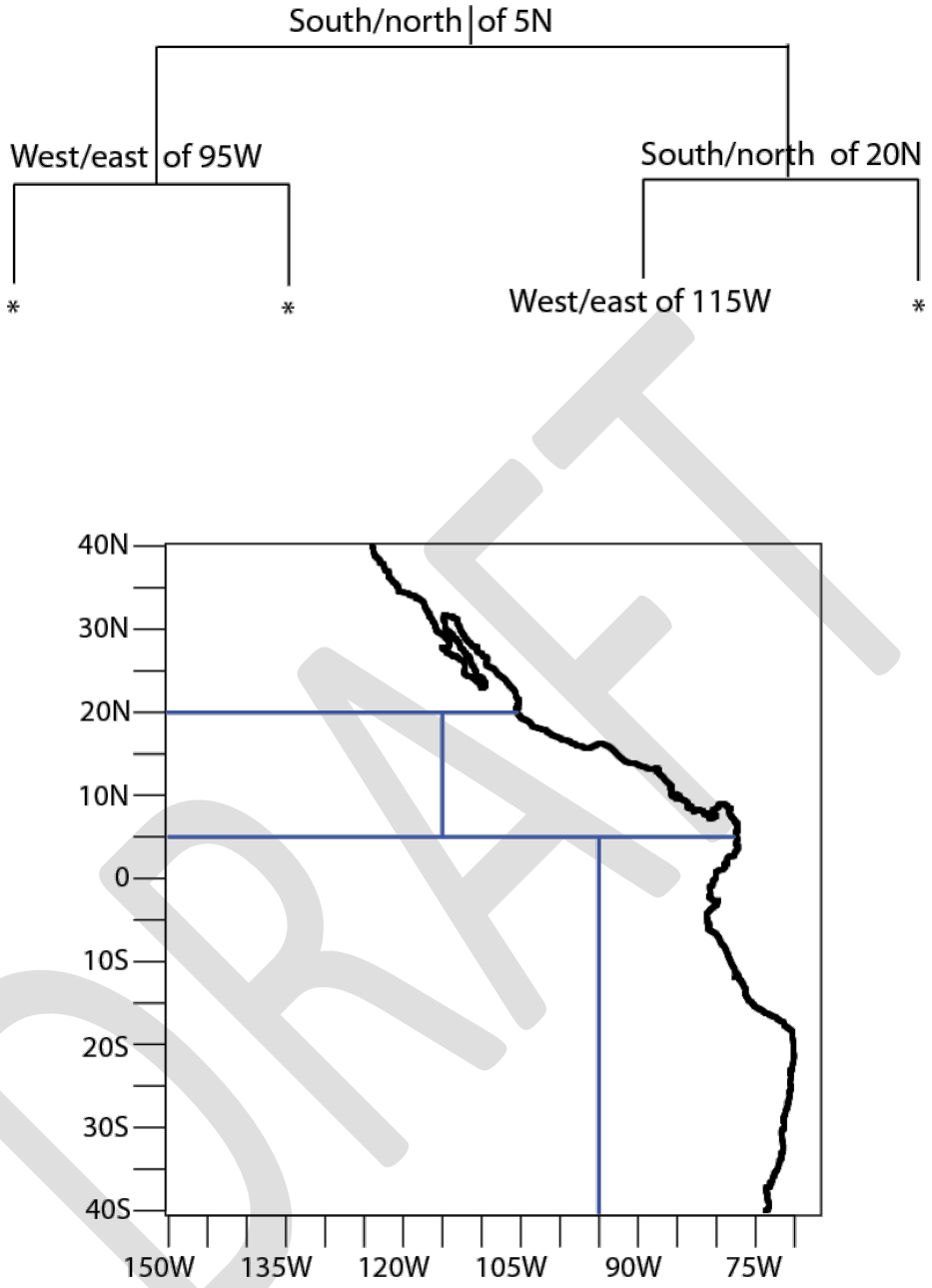


Figure 14. Tree produced from the variance-weighted simultaneous analysis. All left branches include the end point; branch length is displayed as uniform. *: no further splits explored.