YFT-01-02

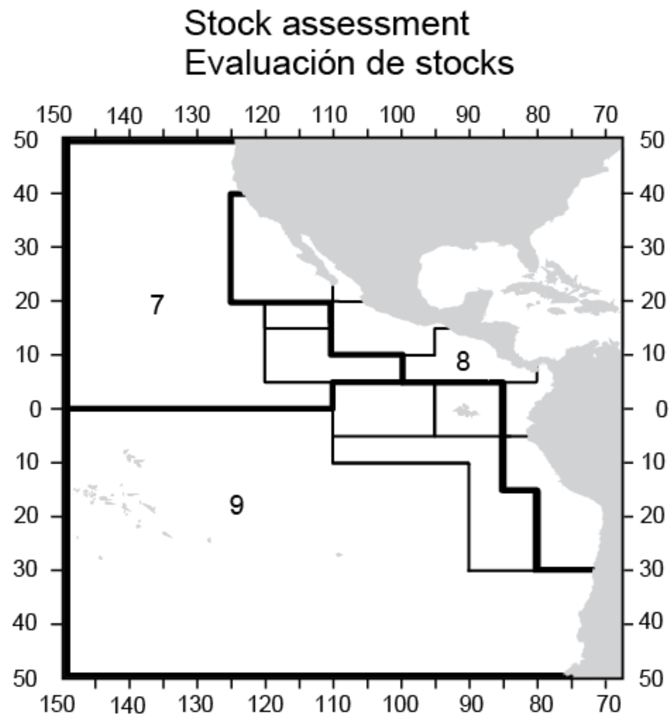
Exploring large-scale pattern in yellowfin tuna data from dolphin sets of the eastern Pacific Ocean purse-seine fishery

Outline of presentation

- Background: why large-scale spatial pattern is of interest
- General approach to studying large-scale spatial pattern
- Description of input data
  - Length-frequency distributions
  - Catch-per-unit-effort trends
- Description of the tree methods
- Results for yellowfin tuna in dolphin sets
- Future work

Why is large-scale spatial pattern of interest?

- Current IATTC stock assessment areas are formed from the areas used to guide collection of port-sampling data.
- Sampling spatial stratification was developed in 1960s (refined in 1990s).
- As fisheries evolve over time, it is useful to consider alternative strata.
- There are three types of purse-seine sets and species composition of the catch differs by set type; yellowfin tuna is the dominant catch of dolphin sets.
- This work focuses on data for yellowfin tuna from purse-seine sets on dolphins for 2000-2011 (large-vessels).



General approach to studying large-scale pattern

Identify similar large-scale structure:

- length-frequency distributions, and
- catch-per-unit-effort trends.

Need length-frequency distributions and catch-per-unit-effort trends on a fine-scale spatial-temporal grid for the EPO fishery area.

Spatial-temporal grid: 5° latitude by 5° longitude by quarter-of-the-year.

Why?

- 5° spatial information available for virtually all length-frequency data.
- Assessment model has quarterly time step and would like to know if large-scale spatial pattern varies seasonally.

CIAT IATTC

Length-frequency distributions

IATTC port-sampling data, 2000-2011 (2611 well samples).

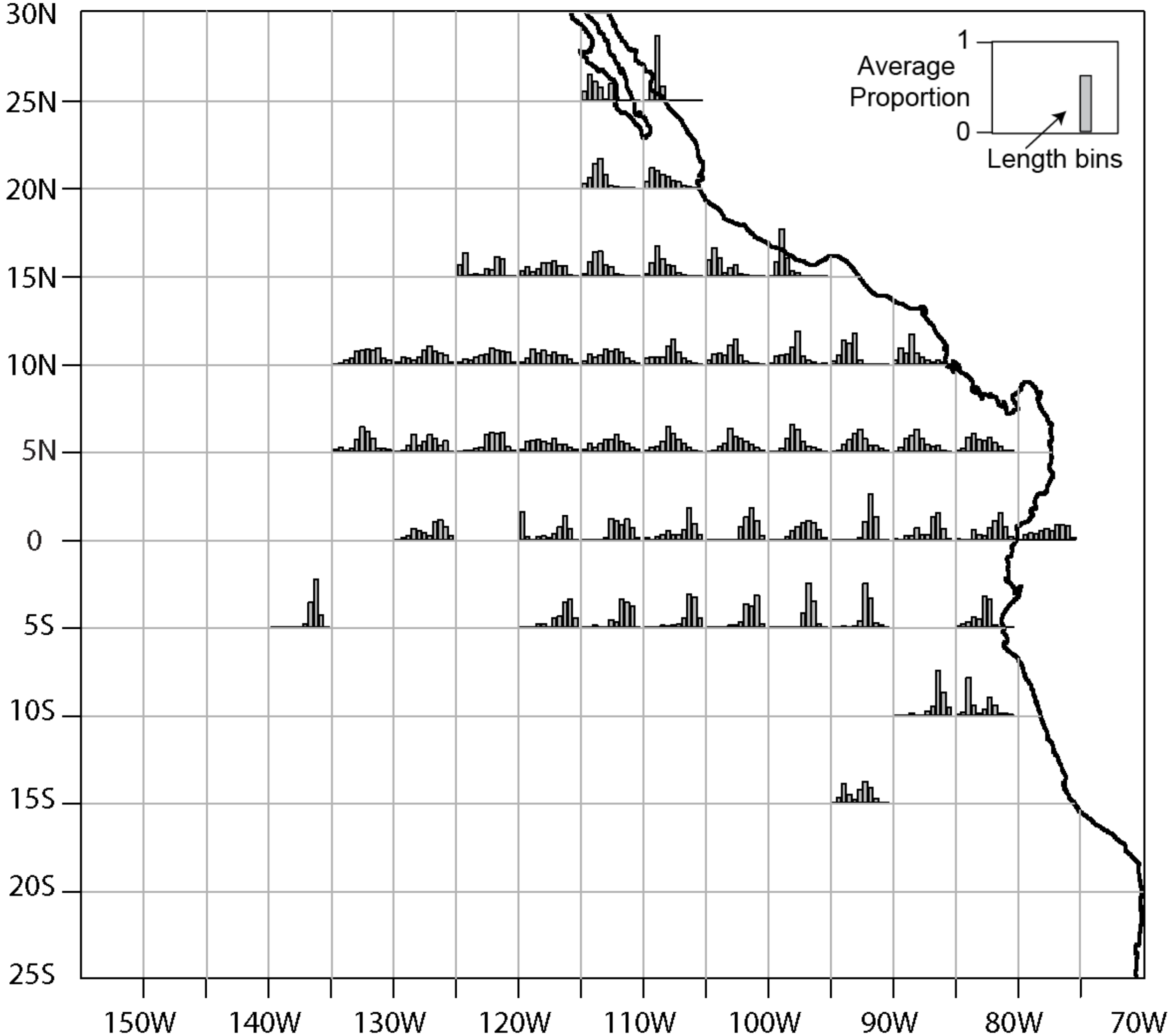
For each sample, have the month, 5° latitude and 5° longitude of fishing.

Regard samples from months of the same quarter as 'replicates' for the quarter.

Data processing:

- Raise sample data to the well catch (to accommodate samples from 'sorted' unloadings).
- Grown/shrink lengths to mid-month of each quarter (Gompertz growth model).
- Summarize each sample (l) by proportion of fish in each of 11 length intervals, $\{p_l(j), j = 1, \ldots, 11\}$: ≤ 58cm, 59-69cm, …,136-146cm, 147-159cm, ≥ 160cm.

Summary of yellowfin tuna length-frequency distributions
October-December



Catch-per-unit-effort trends

Catch and effort data from observer and logbook data bases, 1975-2011.

Catch per day fishing (CPD) for each month and 5° area was computed by the same method as that of the assessment.

Regard samples from months of the same quarter as 'replicates' for the quarter.

Within each 5° area by quarter (i) with sufficient data:

- Fit a smooth model to the temporal trend in nominal CPD using penalized cubic regression splines:

$$\textit{square root}\left(cpd_{i,y,n_y}\right) = f\left(year_{i,y,n_y}\right) + error_{i,y,n_y}$$

where

cpd is catch divided by days fishing;

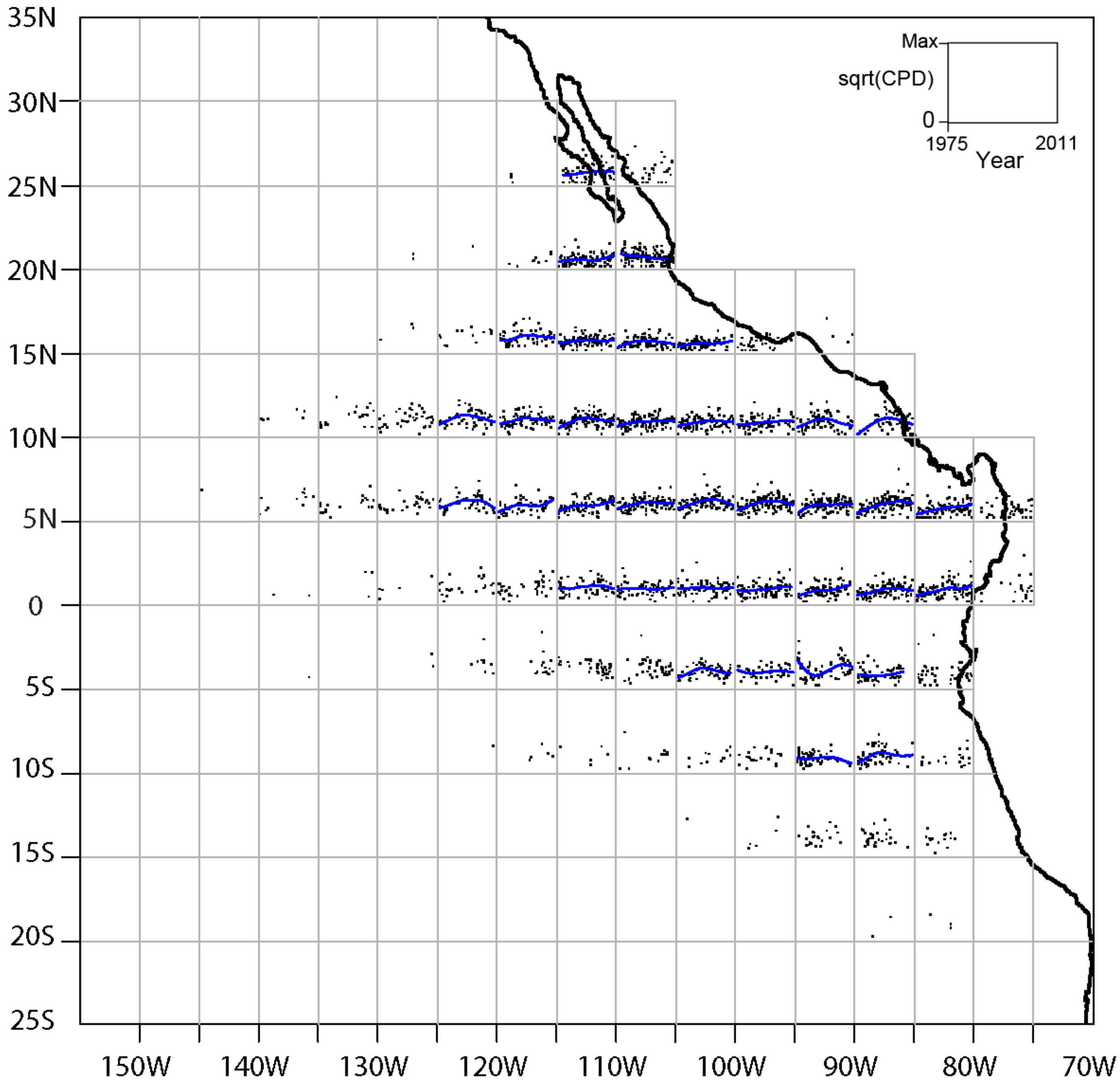
f is a smooth function;

y indexes year and n_y indexes data points;

basis dimension, knot locations and smoothing parameter same for all i.

- Predict annual time series of CPD (on scale of square root), $\hat{C}_i$.

Yellowfin tuna CPD trends
October-December



Methods of analysis

Use tree-based methods to study large-scale pattern in both length-frequency distributions and CPD trends.

Small trees are built by binary recursive partitioning (no pruning).

Predictors: 5° latitude (numeric), 5° longitude (numeric), quarter (numeric, but including cyclic quarter values).

Build three types of trees:

- using only the length-frequency distributions;
- using only the CPD trends;
- using length-frequency distributions and CPD trends, simultaneously ('simultaneous' tree).

To build the trees, need:

- response variable(s), and
- measures of impurity (heterogeneity) to define the split criterion.

Response variables and impurity measures

Length-frequency distributions

- Response
  - Proportion of fish in each size bin; $\{p_l(j)\}$.
- Impurity
  - Kullback-Leibler divergence ('KLD'): $I_{KLD} = \sum_l \sum_j p_m(j) log\left(\frac{p_l(j)}{\bar{p}_{.}(j)}\right)$

CPD trends

- Response
  - Vector of first-differenced annual CPD time series ($\hat{C}_i$), $\Delta\hat{C}_i$.
- Impurity

$$I_{SS\_weighted} = \sum_i \left(\hat{C}_i - \tilde{C}\right)^T \Delta^T \left(\hat{\Sigma}_{\Delta\hat{C}_i}\right)^{-1} \Delta\left(\hat{C}_i - \tilde{C}\right)$$

where $\tilde{C}$ is estimated from the pooled data, and $\hat{\Sigma}_{\Delta\hat{C}_i}$ is a diagonal matrix of variance estimates.

Building the trees

Repeatedly partition the data set(s) at predictor values that maximize the following split criteria:

- Length-frequency distributions:

$$Imp\_KLD = n_{left} \sum_j \bar{p}_{left}(j) \, log\left(\frac{\bar{p}_{left}(j)}{\bar{p}_{.}(j)}\right) + n_{right} \sum_j \bar{p}_{right}(j) \, log\left(\frac{\bar{p}_{right}(j)}{\bar{p}_{.}(j)}\right)$$

- CPD-based trends:

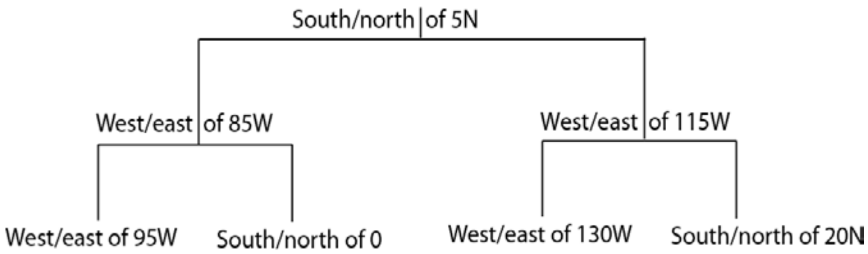
$$Imp\_SS = I_{SS;\,all} - \left(I_{SS;\,left} + I_{SS;\,right}\right)$$

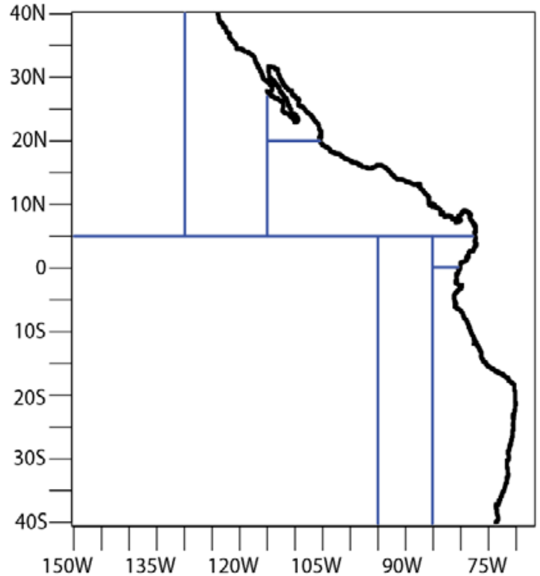
- Both data types, simultaneously (0<γ<1; used γ=0.5):

$$\gamma\left[\frac{Imp\_KLD}{max_{candiate\ splits}(Imp\_KLD)}\right] + (1-\gamma)\left[\frac{Imp\_SS}{max_{candidate\ splits}(Imp\_SS)}\right]$$

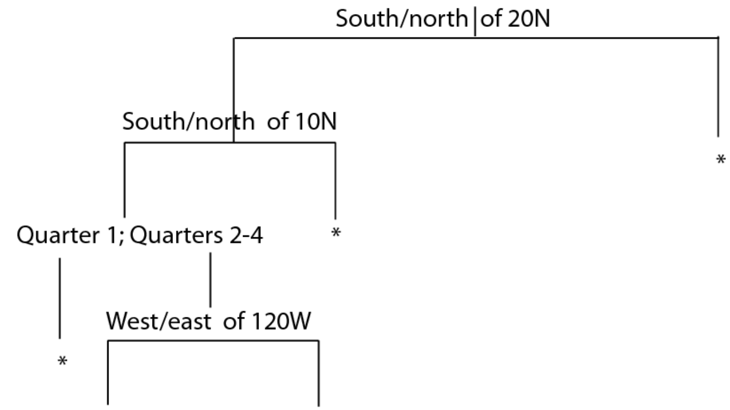
CIAT IATTC

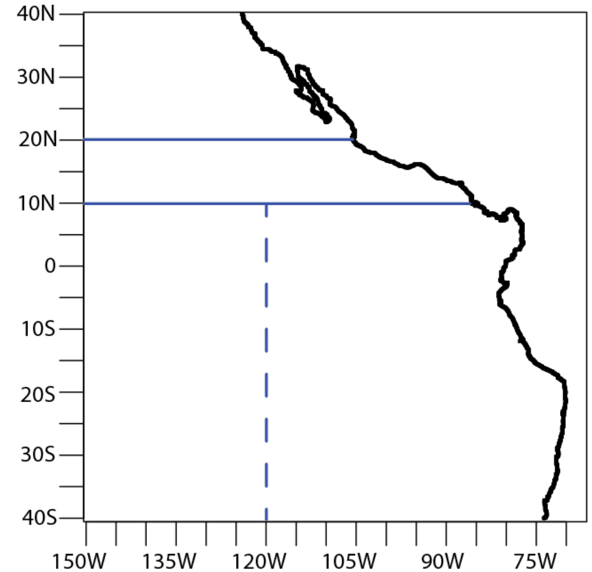
Length-frequency





CPD trends, variance-weighted





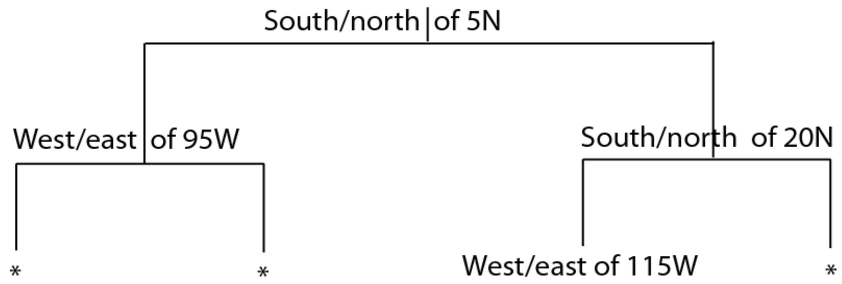
Simultaneous tree, variance-weighted
(YFT-01-02 Table 2)

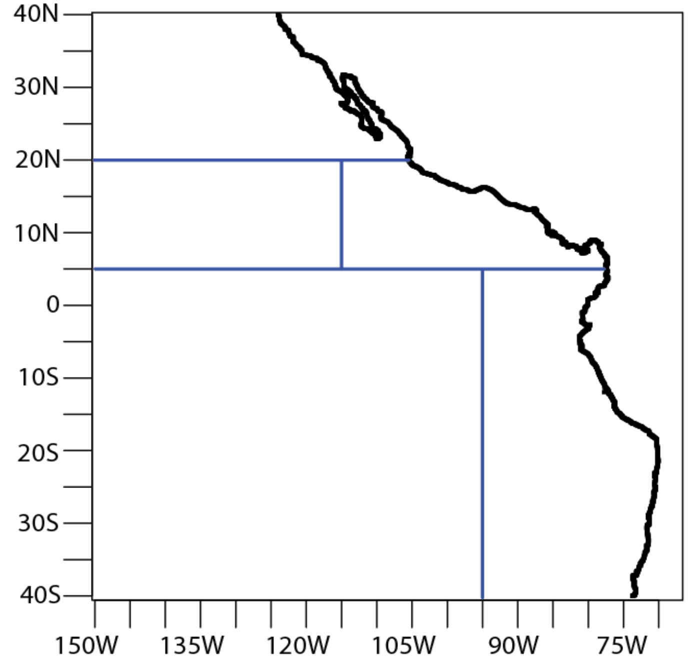
	Scaled improvement length-frequency (split rank)	Scaled improvement CPD trends (split rank)	Simultaneous tree split rank
(a) Full data set (26)			
Latitude 20°N	0.537 (3)	1.000 (1)	2
Latitude 15°N	0.531 (4)		
Latitude 10°N	0.638 (2)	0.723 (4)	3
Latitude 5°N	1.000 (1)	(9)	1
Latitude 0°	(5)		
Longitude 115°W		0.851 (2)	4
Quarters1; 2-4		(5)	
Quarters 1-2; 3-4			
Quarters 1-3; 4		0.782 (3)	

Simultaneous tree, variance-weighted (YFT-01-02 Table 2)

	Scaled improvement length-frequency (split rank)	Scaled improvement CPD trends (split rank)	Simultaneous tree split rank
(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)	

Simultaneous tree, variance-weighted





Future work

Sensitivity analyses:

- criteria used to select 5°-quarters with sufficient data;
- revisions to trends model (other transformations; revisit CPD calculation);
- effects of growing/shrinking length bins;
- inter-annual variability.

Apply the tree-based methods to fishery data from other purse-seine sets types:

- yellowfin tuna, skipjack tuna in unassociated sets;
- bigeye tuna, skipjack tuna in floating-object sets.