YFT-01-03

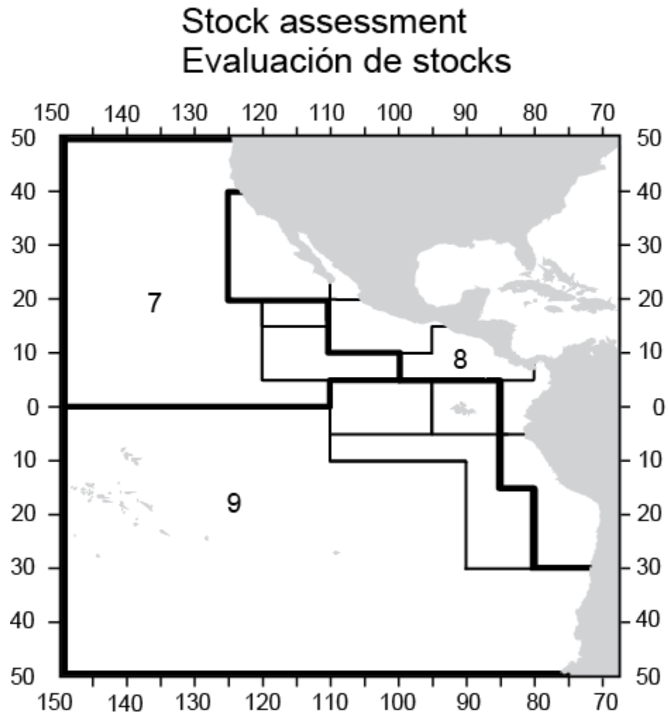
Poststratification of purse-seine port-sampling data from dolphin sets

Outline of presentation

- Background
  - Port-sampling data
  - Ratio-based estimator of species catch
  - Brief mention of some previous results
- Present analyses
  - Brief overview of recent analyses
  - Regression tree analysis of average weight (yellowfin) used to define sub-strata
  - Linear model analyses of average weight data (yellowfin, skipjack) used to evaluate candidate postratifications
- Future work

Why consider poststratification?

- Presently considering revising the stock assessment areas.
- Revised stock assessment areas may not be aggregates of the current sampling areas.
- As a result, the spatial ‘strata’ used to compute total catch by species may need to be modified, based on poststratification of the port-sampling data.
- As a added benefit, poststratification may simplify treatment of the ‘missing data’ problem (strata with catch but no/insufficient sample data).
- The present work focuses on the purse-seine fishery on tunas associated with dolphins, from 2000-2011.



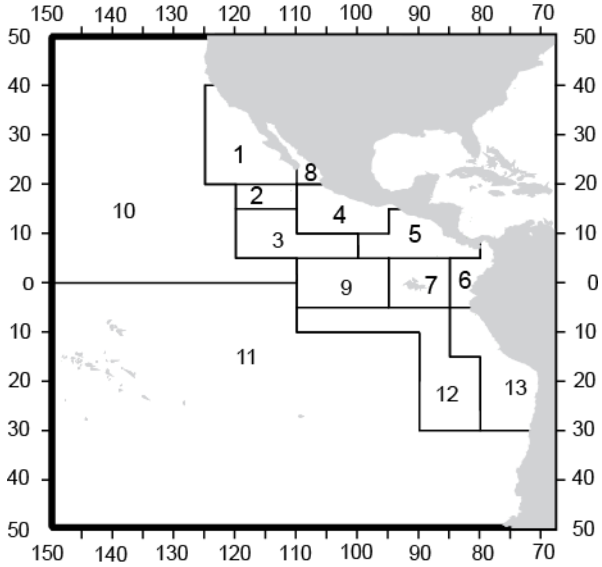
Background: port-sampling data

To obtain a representative data set, the purse-seine fishery is divided into categories (“strata”) that are defined by:

Area (13)

Month (12)

Mode of fishing (6)



Type of vessel	Type of set
small purse-seiner	floating-object
"	unassociated
"	dolphin
large purse-seiner	floating-object
"	unassociated
"	dolphin

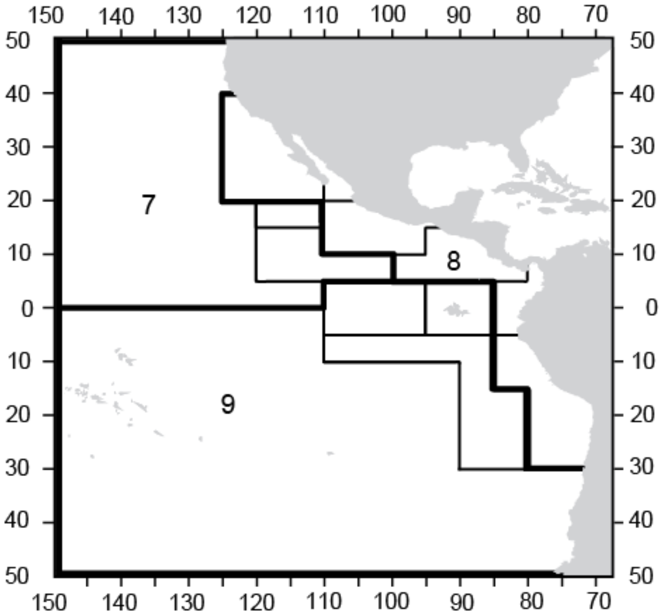
Background: port-sampling data

- A two-stage sampling procedure is used to sample tuna catches in port:
  - 1st stage: vessel wells
    - sampled largely opportunistically;
    - only sampled if catch is from the same "stratum."
  - 2nd stage: fish within a well
    - individual fish selected from an opportunistically-established starting point during unloading;
    - approximately 50 fish of each species are measured;
    - independent of the measured fish, several hundred fish are counted for species composition;
    - sampling differs slightly for catches unloaded by species and size.
- The sample data contain information on both the sampling area and the 5º area.

Background: estimation of catch by species

Current estimator of species catch (in weight):

$$\widehat{W}_{hi} = W_h\,\widehat{p}_{hi} \quad = W_h \left[\frac{\sum_{j=1}^{q} W_{hj} \left(\dfrac{\frac{w_{hij}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}}{\sum_{i=1}^{3} \frac{w_{hij}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}} \right)}{\sum_{j=1}^{q} W_{hj}} \right]$$



W_h : total weight of all species combined in sampling stratum *h* (*h*: area x month x fishing mode);

$\hat{p}$: estimate of the fraction of species *i* (in weight);

W_{hj} : total weight of all species combined for the j^{th} well sample in *h*;

w : sum of the weights of fish measured (from lengths);

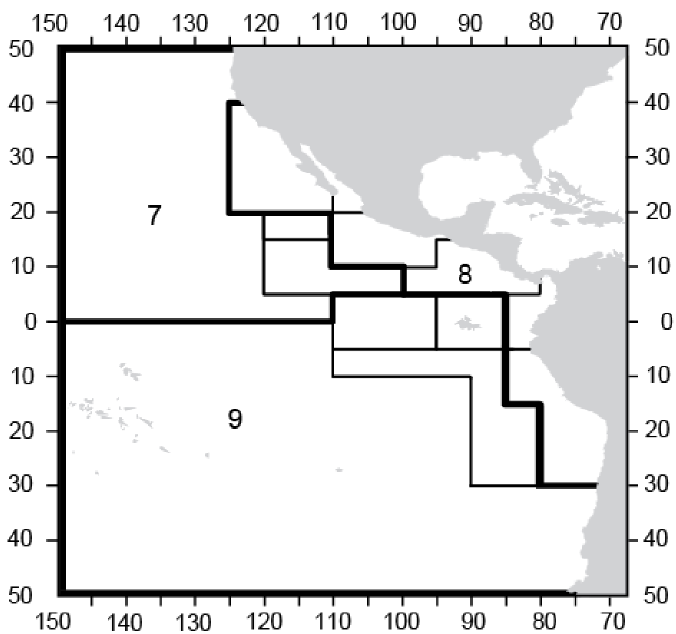
m : number of fish measured;

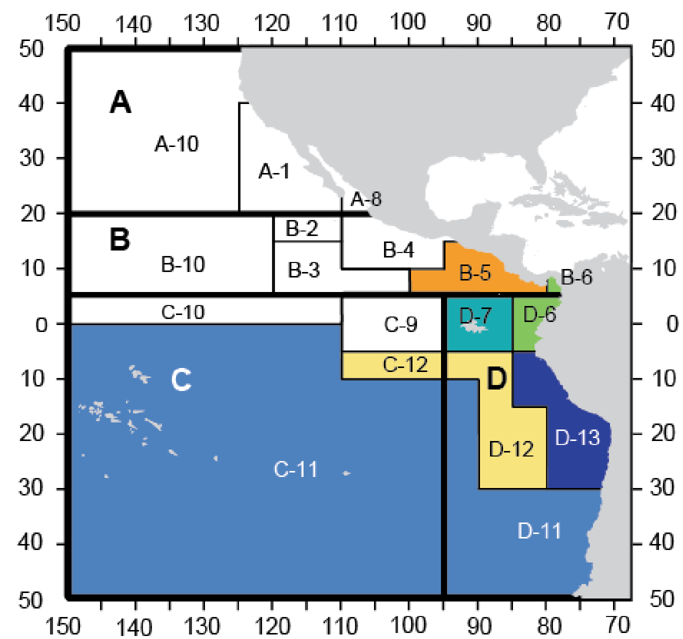
n : number of fish counted;

g : function of the sample means and species fractions.

Background: estimation of catch by species

- Previous work found that:
  - poststratifying the data by sampling areas *within* hypothetical assessment areas was probably not necessary;
  - poststratifying the data *only* by the assessment areas was probably not sufficient.





Present analyses

Purpose of the present analyses: determine a reasonable level of poststratification of the sample data, considering *all* species caught in dolphin sets.

General approach: study spatial and temporal variability average weights by species, and species occurrence.

Focus on yellowfin and skipjack tuna because dolphin sets catch almost no bigeye tuna.

Present analyses

Two large-scale spatial stratifications were considered:

- current stock assessment areas
- tree areas from Document YFT-01-02.

Analyses were conducted to:

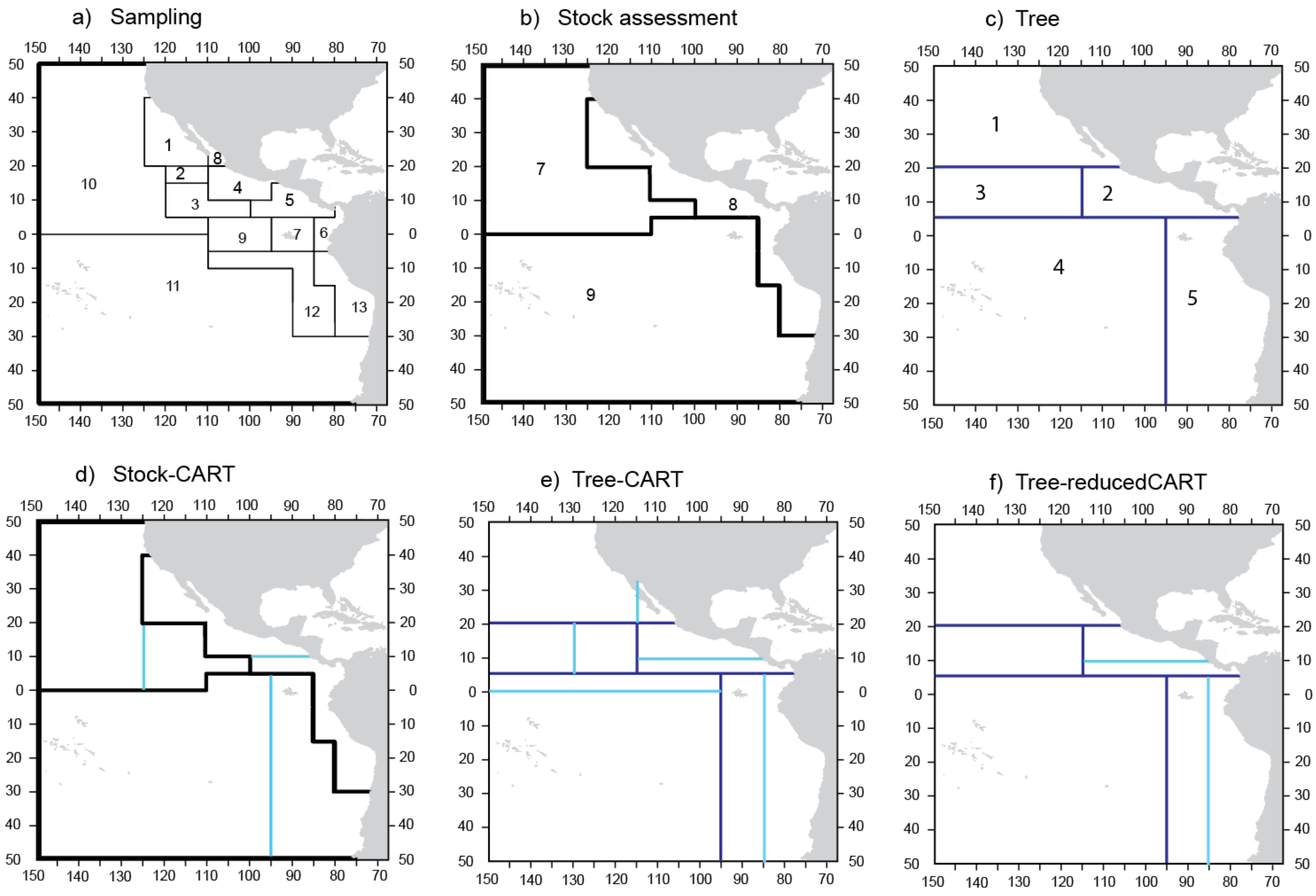
1) explore possibilities for sub-stratification of each large area:

   - regression tree analysis of average weight (yellowfin, skipjack);
   - classification tree analysis of presence/absence of skipjack in the samples.

2) compare the performance of candidate poststratifications from year to year, using linear model analysis of average weight (yellowfin, skipjack).

Linear model analysis of average weight

Spatial poststratifications used in the linear model analyses of average weight:



Linear model analysis of average weight

For average weight ($\bar{w}$), the following linear models were fitted:

$t(\bar{w}_j)$ = overall constant + stock assessment area effect + error
$t(\bar{w}_j)$ = overall constant + tree area effect + error
$t(\bar{w}_j)$ = overall constant + stock-CART area effect + error
$t(\bar{w}_j)$ = overall constant + tree-reducedCART area effect + error
$t(\bar{w}_j)$ = overall constant + tree-CART area effect + error
$t(\bar{w}_j)$ = overall constant + sample area effect + error
$t(\bar{w}_j)$ = overall constant + stock assessment area*quarter + error
$t(\bar{w}_j)$ = overall constant + tree area*quarter + error
$t(\bar{w}_j)$ = overall constant + stock-CART area*quarter + error
$t(\bar{w}_j)$ = overall constant + tree-reducedCART area*quarter + error
$t(\bar{w}_j)$ = overall constant + tree-CART area*quarter + error
$t(\bar{w}_j)$ = overall constant + sample area*quarter + error

t: data transformation (yellowfin: square root; skipjack: Box-Cox with λ=0.5)

Models were fitted separately by year, with weights equal to the individual-well total catch.

Within year and species, models compared with: $\Delta AIC = AIC - AIC_{min}$.

Linear model analysis of average weight

Example of results:

2005	YFT ΔAIC	SKJ ΔAIC
Stock (model (i))	96.7	25.4
Tree (model (ii))	29.4	8.4
StockCART (model (iii))	48.8	14.2
Tree-reducedCART (model (iv))	0.3	0.0
TreeCART (model (v))	1.2	1.8
Sample (model (vi))	15.0	12.0
Stock * quarter (model (vii))	75.0	27.8
Tree * quarter (model (viii))	2.1	16.9
StockCART * quarter (model (ix))	44.9	10.2
Tree-reducedCART * quarter (model (x))	0.0	5.9
TreeCART * quarter (model (xi))	3.3	9.1
Sample * quarter (model (xii))	9.3	0.3

Linear model analysis of average weight

Results of the linear model analyses of the average weight data can be summarized as follows:

- For the dominant species, estimation of total catch based on a spatial stratification with fewer areas than the present 13 areas is likely adequate (…déjà vu).
- For the dominant species, linking the assessment areas and catch estimation strata by sub-stratifying the assessment areas seems reasonable.
- For minor species, port-sampling data may be insufficient spatial-temporal analyses.

What's next?

Catch estimation strata have to be based on: area, time period, purse-seine set type (not tuna species).

Different set types catch the three tuna species in very different amounts:

2011 catch (mt)	**Yellowfin tuna**	**Skipjack Tuna**	**Bigeye tuna**
Dolphin	134220	5148	3
Floating object	39094	178262	55479
Unassociated	29022	95590	1044

What's next?

This suggests the following as a possible approach to defining poststrata for estimation of catch by species.

- Refine current assessment areas for a particular set type based on detailed spatial-temporal of the fisheries data for the dominant specie(s) (probably collapsing assessment areas for the minor specie(s)).

- Define candidate sub-strata within assessment areas based on spatial-temporal analysis of fishery data for the dominant catch specie(s) (average weight, sample species proportions).

- Evaluate the annual performance of candidate poststrata, for every species of a particular set type, based on the estimated variance and coefficient of variation of total species catch.