INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

SIXTH MEETING

La Jolla, California (USA)
11-15 May 2015

DOCUMENT SAC-06 INF-A

COMPUTATIONAL DETAILS OF THE SPATIAL POSTSTRATIFICATION AND THE PENNINGTON SAMPLE SIZE FOR THE 2014 EXPLORATORY ASSESSMENTS

Cleridy E. Lennert-Cody

1. SUMMARY

The 2014 exploratory assessments are based on different stock and fishery spatial definitions than previous assessments. This document describes the method used to define new spatial strata for catch estimation with the new stock and fishery definitions for 1975-2014. The document also describes the method used to compute the Pennington sample size for the 2000-2014. The Pennington sample size estimates are used as weights in the exploratory assessments. Both methods will be revised for the 2015 assessment.

2. SPATIAL POSTSTRATIFICATION

2.1. Background

The current catch estimation methodology is based on computing estimates of species and size composition within strata. Historically these strata have been defined by the IATTC sampling areas, month and gear (purse-seine set type -vessel size class) (Tomlinson 2004; Suter 2010). Catch estimates for each stratum are then combined across areas, month and vessel size class categories to obtain estimates for each fishery area and quarter. The IATTC sampling areas were developed based on the yellowfin fishery before it expanded to cover the whole EPO and do not adequately represent the western parts of the EPO, thus limiting their usefulness for generating data for spatially defined fisheries. To adapt the current catch estimation methodology to the new stock and fishery spatial definitions (SAC-06-04c), new catch areas were created within fishery areas.

Creating these new catch areas was a two-step process. The first step was to merge stock and fishery areas for bigeye and yellowfin tuna, by purse-seine set type, thereby forming a common spatial stratification for both species within a set type (skipjack was not considered at this step because there is no skipjack assessment). Those combined fishery areas are shown in Figures 1-3. There was no change to the yellowfin areas for dolphin and unassociated sets from this first step because the bigeye stock and fishery definitions are the same for these two set types and align with the yellowfin boundaries. The second step of the process was to further divide some of these areas into two sub-areas, based on analysis of species composition and average weights. The analysis used for this further substratification is described below.

2.2. Data and method of analysis

Port-sampling data for bigeye, yellowfin and skipjack tuna from 1980-2013 were used in this analysis. Data prior to 1980 were excluded to avoid any influence of the CYRA/XCYRA regulations (Suter 2010 and references therein) on the spatial structure identified by this analysis. Unlike the boundaries of the IATTC sampling areas used for previous catch estimation (Tomlinson 2004; Suter 2010), the CYRA/XCYRA boundaries do not coincide with the new stock and fishery area boundaries. The data for 1980-1999 were analyzed separately from that for 2000-2013 because different sampling protocols were used in the two time periods (Suter 2010).

The port-sampling data were summarized by the following well-level quantities:

$$\text{species composition: } \frac{\hat{N}_{ij}}{\hat{N}_{.j}}$$

$$\text{average weight: } \frac{\hat{W}_{ij}}{\hat{N}_{ij}}$$

where $\hat{W}_{ij}$ is the estimated weight of species i in well j, $\hat{N}_{ij}$ is the estimated number of fish of species i in well j and $\hat{N}_{.j}$ is the estimated number of all fish (yellowfin, bigeye and skipjack) in well j. Species composition estimates are only available for 2000-2013. These species composition and average weight quantities are equivalent to, or similar to, the well-level summaries used in the catch estimation and were computed following methods outlined in Tomlinson *et al.* 1992 and Tomlinson 2004. Because some species are not commonly caught in all set types, the species included in the analyses varied by set type: the analysis of dolphin-set data used only yellowfin tuna summaries; the analysis of unassociated-set data used yellowfin and skipjack tuna summaries; and, with one exception, the analysis for floating-object set data used summaries for all three tuna species. For comparison, separate analyses of species proportions and average weights were also conducted for the 2000-2013 data.

To determine sub-strata within fishery areas for each purse-seine set type, regression tree analyses were conducted within each fishery area, by set type, where there were sufficient sample data. Multivariate regression trees (De'ath 2002) for 2000-2013, and classical regression trees (Breiman *et al*. 1984) for 1980-1990, were used to partition the data into more homogeneous sub-groups. To avoid over-stratification of the data, only the first split of each tree was considered. The sub-areas were primarily selected based on results of the analysis of the 2000-2013 data because both species proportions and average weights were available. However, sub-area definitions also took into consideration results from analysis of 1980-1990 average weights, separate analyses of average weights and species proportions for the 2000-2013 data, and best competitor split information for both time periods. Predictor variables used in the regression tree analyses were: 5° latitude, 5° longitude, month (categorical), vessel size-class (categorical: class 1-5; class 6). Data were analyzed in *R* (R Development Core Team 2012) with *vegan* (Okansen *et al*. 2013) and *mvpart* (Therneau and Atkinson 2012) packages.

2.3. Results and future work

The resulting sub-stratification of the fishery areas is shown in Figures 1-3. In most cases there were similarities between the first partitions selected for the 2000-2013 data and those selected for the 1980-1990 data. For the sub-division of the southern inshore dolphin-set area (Figure 1), the first tree partition for 2000-2013 data was at 85°W (first competitor split was on month) and the first competitor split for the 1980-1990 data was at 90°W (the first split was on month). Similar results were obtained for the subdivision of the southern inshore area for unassociated sets (Figure 2), which emphasizes the strong spatial-seasonal aspect of this region. The partition of the northern offshore area for floating-objects sets (Figure 3) represents a compromise between the first competitor split for 2000-2013 for all three species, the first split for bigeye tuna proportion and average weight, the first split for proportions for all three species, and the first split for bigeye and skipjack average weights for 1980-1990. The partition of the southern offshore area for floating-object sets (Figure 3) was also a compromise between the first partition for all three species (115°W), the first partition for bigeye tuna 2000-2013 (110°W), and the first partition for 1980-1990 for bigeye tuna (105°W), yellowfin tuna (115°W) and skipjack tuna (110°W). The partition of the northern inshore area for floating-object set data (Figure 3) was based on data for yellowfin and skipjack tuna. There was less agreement in this area between analyses and time periods. The partition selected presents the results for yellowfin average weight and proportion for 2000-2013.

For the 2015 assessment, catch spatial strata will be determined based on a combination of methods. More detailed regression tree analyses will be used to define candidate sub-areas. In addition, over-stratification will be evaluated using simulations to investigate stability and error of the estimates for different levels of stratification, including changes to spatial strata (stock areas, fishery areas, sub-division of fishery areas), time periods (months *versus* quarter) and vessel size class. Evaluating over-stratification is important because the finer the level of stratification of the data, the more strata there will be for which no port-sampling data are available in any given year. In such situations, port-sampling data from “neighboring” strata have to be used to estimate size composition, and for 2000 onwards, species catch. The trade-offs between high-resolution strata/high-levels of substitution and lower-resolution strata have not been evaluated previously due to the computational challenges. However, because the catch estimation programs have recently been implemented in R, more rigorous evaluation will be possible.

3. PENNINGTON SAMPLE SIZE

3.1. Background

Size composition estimates used in the stock assessments are weighted to give more influence to those size compositions that are considered to be more reliable. There are several options for computing weights. One option for weights is the Pennington sample size (Pennington *et al.* 2002). The Pennington sample size was originally defined to provide a measure of the precision of size composition estimates from fishery-independent survey data. To adapt the Pennington sample size to the present yellowfin and bigeye tuna stock assessments, we follow the work of Chris Francis (unpublished).

3.2. Data and method of analysis

Port-sampling data and catch data for 2000-2014 were used to estimate the Pennington sample size for the recent assessment period. For this analysis, estimates were not made for the 1975-1999 period because of differences in sampling protocol between the two periods and uncertainty as to how to produce a consistent time series of estimates for 1975-2014. The Pennington sample size was estimated for each fishery stratum (*i.e.*, each year x set type x fishery area x quarter) for which there were at least two well samples available; in what follows below the stratum subscripting is suppressed.

The Pennington sample size (Pennington *et al.*, 2002, equation 5), $\widehat{m}_{eff}$, is given by the ratio of the variance of the mean length estimated for data from a clustered sampling design (*e.g.*, fish clustered within wells), $var(\hat{R})$, to the variance of the population length distribution (X indicating the random variable length), $\hat{\sigma}_X^2$:

$$\widehat{m}_{eff} = \frac{\hat{\sigma}_X^2}{var(\hat{R})}$$

The Pennington sample size was defined based on mean length, but adapted here to mean weight. For binned data, the population mean weight can be expressed as:

$$\mu = \sum_k \pi_k w_k$$

where π_k is the probability of obtaining a fish of a length corresponding to the k^{th} 1-cm length bin (*i.e.*, the proportion of fish in length bin k) and w_k is the weight associated with the mid-point of the k^{th} length bin. The variance of the population length distribution:

$$\sigma_X^2 = E(X^2) - \left(E(X)\right)^2$$

was estimated as:

$$\hat{\sigma}_X^2 = \sum_k \hat{\pi}_k w_k^2 - \left(\sum_k \hat{\pi}_k w_k\right)^2$$

where $\hat{\pi}_k$ is the stratum estimate of the proportion of fish in the k^{th} 1-cm length bin.

The estimate of the variance of the ratio estimator of mean weight for data collected under a clustered sampling design was obtained by a bootstrap procedure where well-level data were resampled within each catch stratum (area x month x set type x vessel size class) for each year. The bootstrap estimate of variance from B bootstrap data sets is given by:

$$var(\hat{R}) = \frac{1}{(B-1)} \sum_i^B \left[\left(\sum_k \pi_{ik}^* w_k\right) - \left(\frac{1}{B}\sum_i^B \sum_k \pi_{ik}^* w_k\right)\right]^2$$

where π_{ik}^* is the estimate from the i^{th} bootstrap replicate of the proportion of fish in length bin *k*.

To deal with instability in the estimates of the Pennington sample size for strata with few well samples, the Pennington estimates, as a function of the number of samples with a species, were smoothed using a generalized additive model and smooth predictions generated.

3.3. Results and future work

Estimates of Pennington sample size and the relationship between the Pennington sample size and the number of wells with a species, by set type, are shown in Figure 4. Predicted Pennington sample size values used in the exploratory stock assessments were based on predictions from these smooth curves. Depending on the results of simulations to determine stratum “size” and minimum number of wells samples per stratum, the need to smooth the raw Pennington sample size values, as well as the smooth curves, may change for the 2015 assessment.

ACKNOWLEDGEMENTS

Special thanks to Chris Francis for sharing his method for the fishery-dependent version of the Pennington sample size and for taking the time to discuss implementation.

REFERENCES

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1984. Classification and Regression Trees, Chapman & Hall/CRC, New York.

De’ath, G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83, 1105-1117.

Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H. 2013. vegan: Community Ecology Package. R package version 2.0-7. http://CRAN.R-project.org/package=vegan

Pennington, M., Burmeister, L.-M., Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin 100: 74-80.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

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. Special Report 18. Inter-American Tropical Tuna Commission. http://www.iattc.org/PDFFiles2/SpecialReports/SpecialReport18.pdf

Therneau, T.M. and Atkinson, B. 2012. rpart by Terry M Therneau, Beth Atkinson. R port of rpart by Brian Ripley <ripley@stats.ox.ac.uk>. Some routines from vegan – Jari Oksanen <jari.oksanen@oulu.fi> Extensions and adaptations of rpart to mvpart by Glenn De'ath. (2012).

mvpart: Multivariate partitioning. Rpackage version 1.6-0. http://CRAN.R-project.org/package=mvpart

Tomlinson, P.K., Tsuji, S. and Calkins, T.P. 1992. Length-frequency estimation for yellowfin tuna (Thunnus albacares) caught by commercial fishing gear in the eastern Pacific Ocean. Inter-American Tropical Tuna Commission Bulletin 20(6): 357-398. http://www.iattc.org/PDFFiles2/Bulletins/Bulletin-Vol.20-No.6.pdf

Tomlinson, P.K. 2004. Sampling the tuna catch of the eastern Pacific Ocean for species composition and length-frequency distributions. Stock Assessment Report 4, pp. 311-333. Inter-American Tropical Tuna Commission, La Jolla, California, USA. http://www.iattc.org/StockAssessmentReports/StockAssessmentReport4ENG.htm

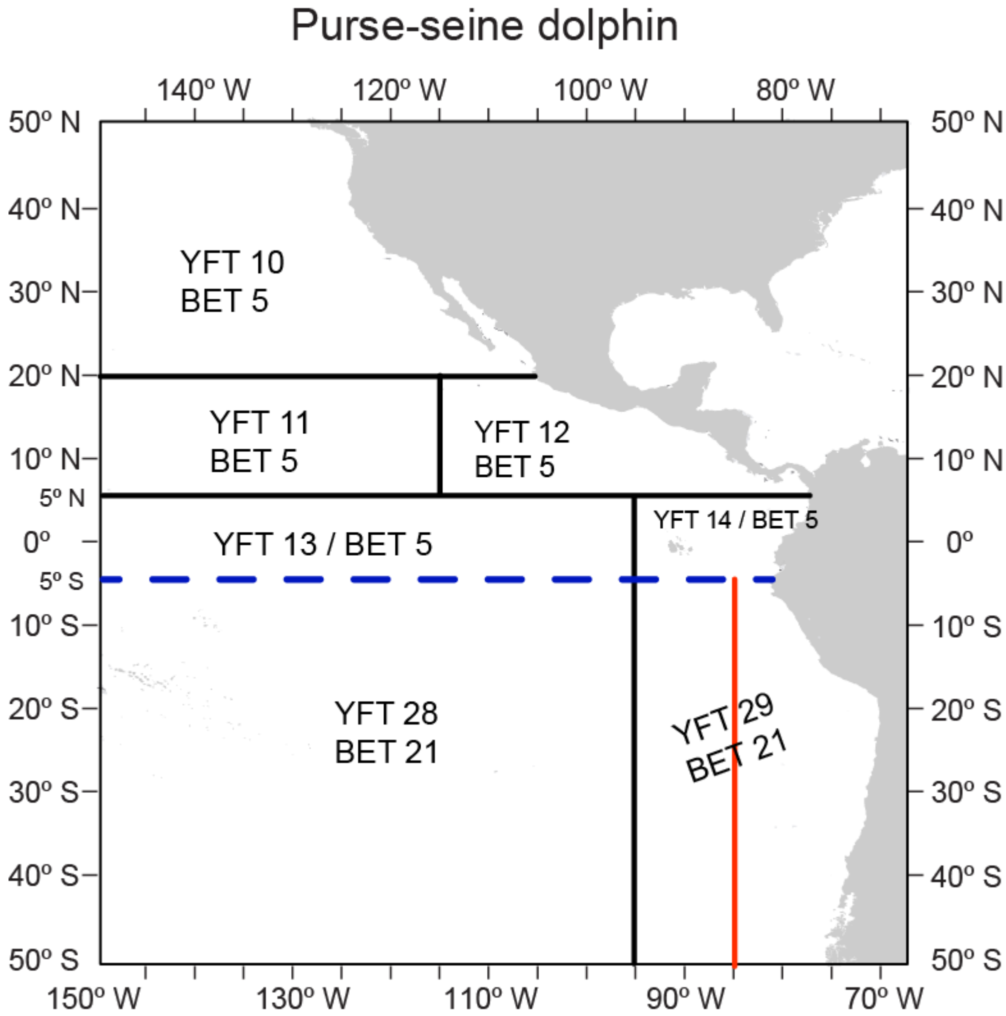

FIGURE 1. Catch spatial stratification for dolphin-set data. Dark blue, dashed lines: stock boundaries; black solid lines: fishery boundaries; orange solid lines: additional boundaries for catch strata.

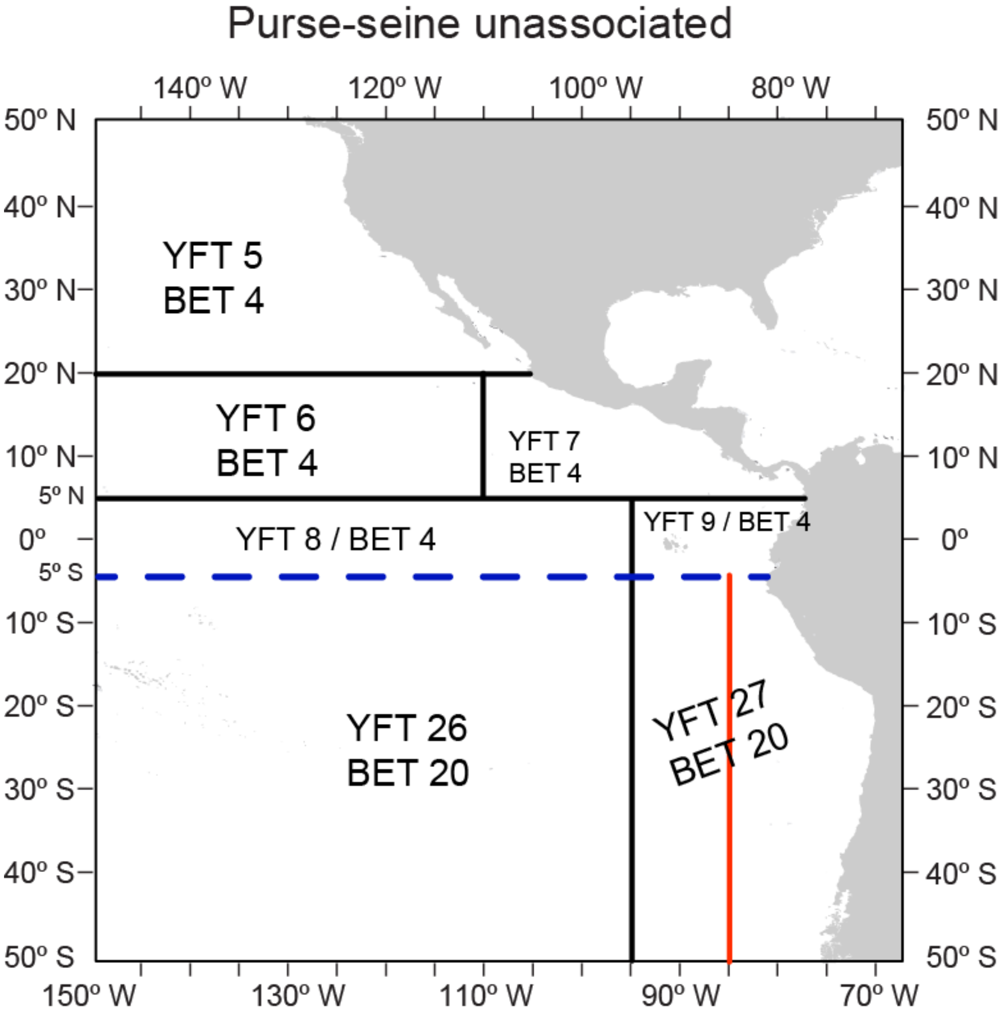

FIGURE 2. Catch spatial stratification for unassociated-set data. Dark blue, dashed lines: stock boundaries; black solid lines: fishery boundaries; orange solid lines: additional boundaries for catch strata.

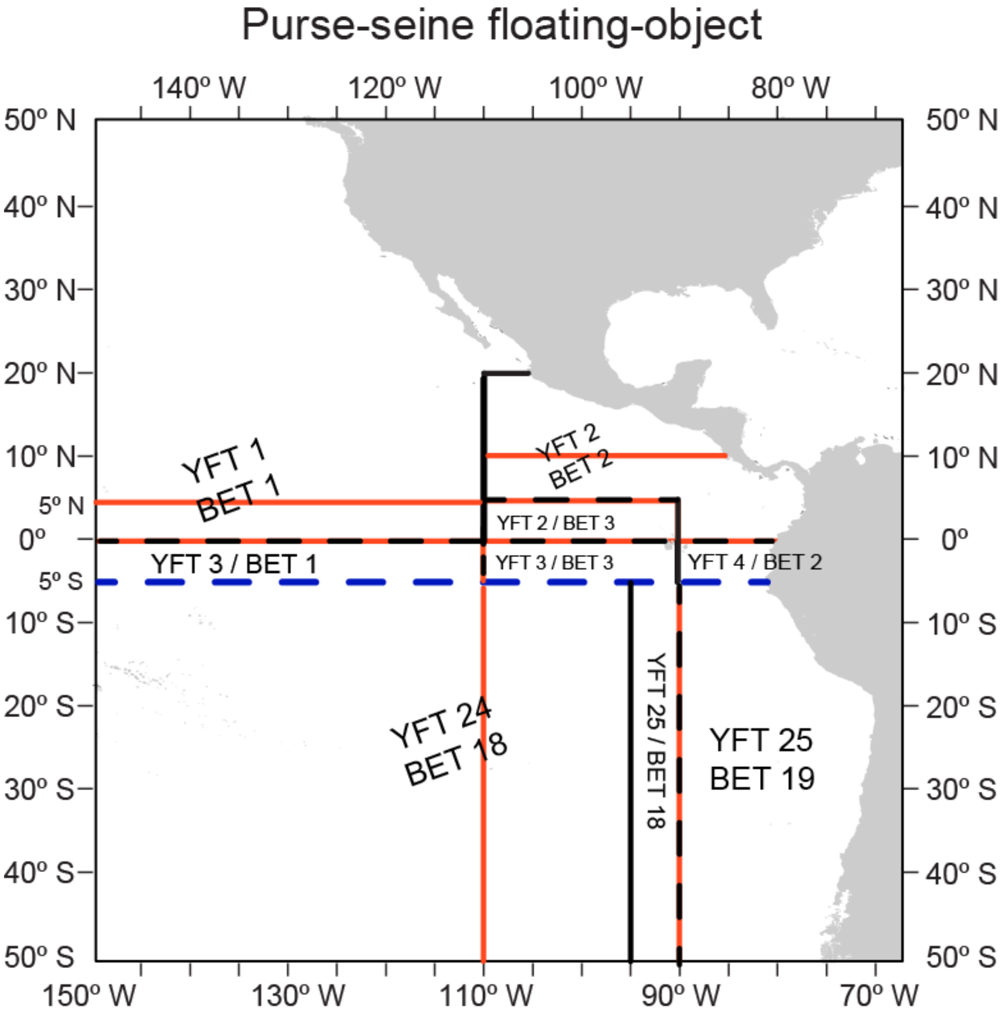

FIGURE 3. Catch spatial stratification for floating-object-set data. Dark blue, dashed lines: stock boundaries; black solid lines: fishery boundaries; orange solid lines: additional boundaries for catch strata; orange-black dashed lines: shared fishery boundaries and catch stratum boundaries.

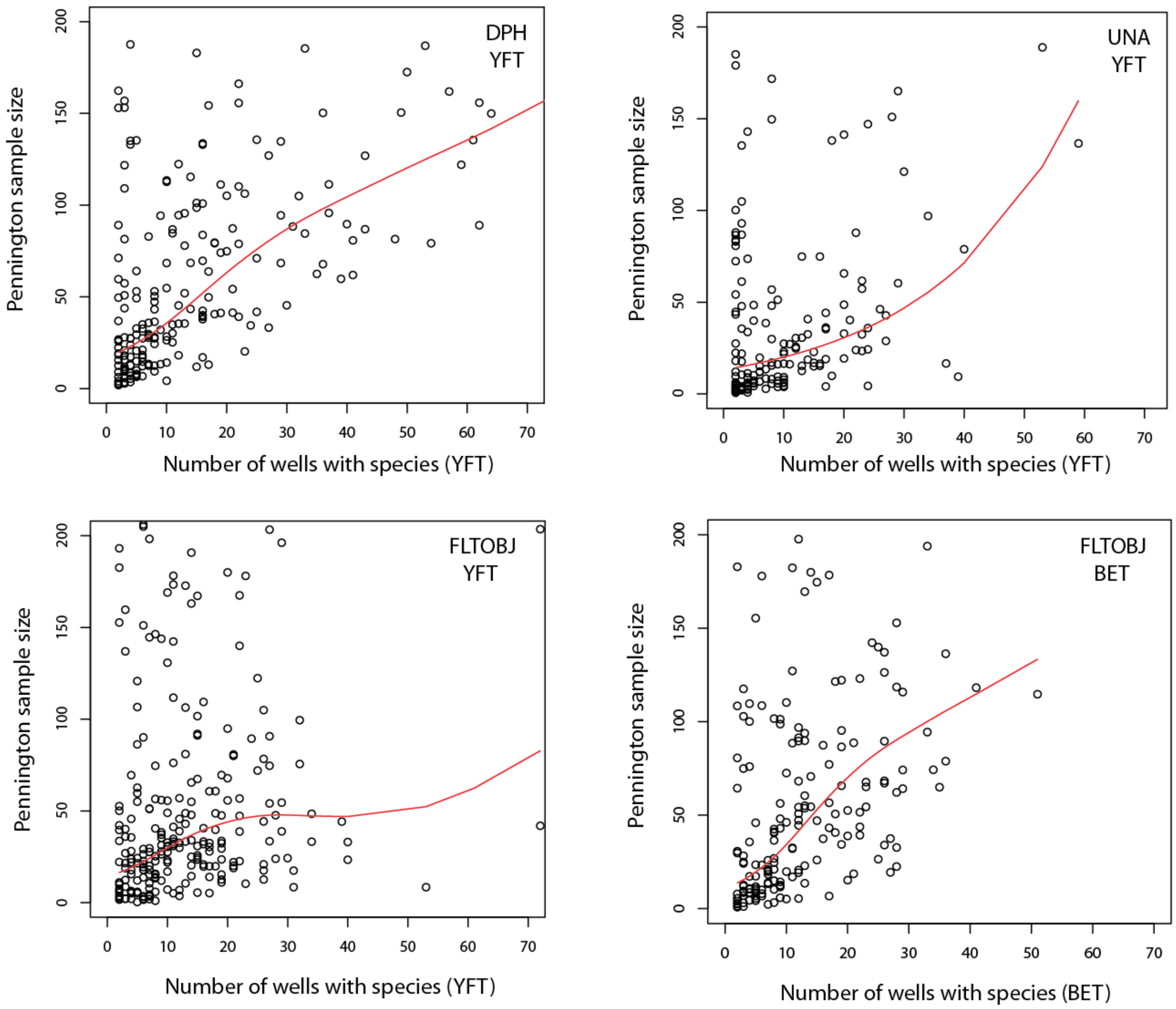

FIGURE 4. Raw Pennington sample size estimates *versus* number of wells with the species, by purse-seine set type, and smooth curves (red lines). Each point in each panel represents the Pennington sample size for a given year x area x quarter. For the purposes of presentation, values of the Pennington sample size larger than 200 are not shown in the figures. “DPH”: dolphin sets; “UNA”: unassociated sets; “FLT”: floating-object sets; “YFT”: yellowfin tuna; “BET”: bigeye tuna.