INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

2ND MEETING

La Jolla, California (USA)
9-12 May 2011

DOCUMENT SAC-02-10

POSTSTRATIFIED ESTIMATORS OF TOTAL CATCH FOR THE PURSE-SEINE FISHERY PORT-SAMPLING DATA

Cleridy E. Lennert-Cody, Mark N. Maunder, Patrick K. Tomlinson, Alexandre Aires-da-Silva, Alejandro Pérez, Marti McCracken

1. SUMMARY

Although the goals of stratification for stock assessment and data collection are often in agreement, it is useful to be able to consider different stratifications for the two, particularly when the characteristics of the fisheries have changed over time. In order to do this, it is necessary to develop an estimator for fishery totals based on the post-data collection strata. Two candidate poststratified estimators for catch species and size composition for the recent purse-seine fishery data are presented. An approach for selecting between the two is also discussed. Application of these methods to purse-seine data from 2000 to present will be undertaken in the future.

2. INTRODUCTION

Stratification is used in stock assessment to address differences in stock and fishery dynamics. In general, the fisheries data (catch, 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. Stratification also can be used during data collection to guard against skewed sample allocations (which might lead to bias) and to minimize variance of the estimators of population totals (*e.g.*, Holt and Smith 1979; Thompson 1992). Thus, the goals of stratification for stock assessment and data collection are often in agreement. However, they may differ if the characteristics of the fisheries have changed over time.

Presently, tuna stock assessment for all tuna species in the eastern Pacific Ocean (EPO) (*e.g*., Maunder and Aires-da-Silva 2010; Aires-da-Silva and Maunder, 2010) uses large areas formed by aggregating the spatial sampling strata (Figure 1a-b). However, because of recent changes in the purse-seine fishery, it is desirable to be able to consider alternative spatial partitions of the EPO fishery area constructed from units other than the sampling strata. For example, the purse-seine fishery on floating objects in the EPO has expanded considerably offshore since the early-1990s (Watters 1999). By contrast, the sampling strata used for surface fisheries in the EPO were primarily developed in the late 1960s (Suter 2008, and references therein) when the fishery was more coastal (Watters 1999) and dominated by yellowfin and skipjack tuna catches using purse seines that set on tuna associated with dolphins and tuna in unassociated schools, and by pole-and-line vessels. Although these sampling strata were refined in the late 1990s (Suter 2008), they may not be optimal now for use in stock assessment in fisheries with a strong offshore component, such as the purse-seine fishery on floating objects, which produces the greatest catches of bigeye tuna (Anonymous 2010).

Poststratification (*e.g*., Holt and Smith 1979; Valliant 1993) is a technique used in data analysis to group samples, after the data have been collected, when estimates of population totals are desired for groups whose definitions were not expressly part of the data collection protocol. Two candidate poststratified estimators of the total catch by species and size, for the port-sampling data collected since 2000, are presented. A data analysis procedure for selecting between the two is proposed. It is assumed herein that candidate poststrata already exist. Such poststrata could be defined, for example, based on analysis of spatiotemporal patterns in length-frequency and/or CPUE trends (*e.g*., Lennert-Cody *et al*. 2010). However, poststrata might also be defined based on other data and criteria.

3. BACKGROUND

3.1. Data collection

Data on the species and size composition of the tuna catch of purse-seine vessels are collected when vessels arrive in port to unload (Tomlinson 2004; Suter 2008). To try to insure that the samples collected are representative of the entire fishery, the purse-seine fishery is divided into categories, or ‘sampling strata’. These sampling strata are defined by the location of fishing (13 areas, Figure 1a), the month of fishing and the mode of fishing (6 modes based on purse-seine set type and size of vessel), for a total of 936 possible strata. Not all strata have fishing activity in any given year. Samples are collected by stratum 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. Because the number of wells in a stratum is not known in advance and because some vessels may unload in ports where logistics make sampling prohibitively difficult, wells to be sampled are selected opportunistically. However, a well is sampled only if all the catch it contains is from the same sampling stratum (*i.e*., same area, month and fishing mode). Over the course of a year, unequal numbers of wells will be sampled per stratum. Once a well of a vessel has been selected, 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. From the same well, and independently of the measured fish, several hundred fish are counted for species composition. The fish sampled from the well are selected one at a time, from an opportunistically established starting point, as circumstances permit. Details of the port-sampling data collection procedures can be found in the Appendix of Suter (2008).

3.2. Current estimators

The current estimators of total catch by species and size (Tomlinson 2004) have the general form of a ratio-type estimator of the stratum total (*e.g*., Thompson 1992) based on the amount of the catch in sampled wells. The total estimated annual catch (in weight) of species i in sampling stratum h is given by:

$$\widehat{W}_{hi} = W_h \hat{p}_{hi}$$

$$= W_h \left[\frac{\sum_{j=1}^{q} W_{hj} \left(\frac{\frac{w_{hij}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}}{\sum_{i=1}^{\# sp} \frac{w_{hij}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}} \right)}{\sum_{j=1}^{q} W_{hj}} \right]$$

$$= W_h \frac{\left[\sum_{j=1}^{q} W_{hj} \cdot g_1(w_{hij}, m_{hij}, n_{hij}; i = 1, \ldots, \#sp)\right]}{\left[\sum_{j=1}^{q} W_{hj}\right]} \qquad (1)$$

where W_h is the total weight of all species combined in sampling stratum *h* (assumed known), $\hat{p}$ is the estimate of the species fraction (derived from weight) in the stratum, W_{hj} is the total weight of all species combined in the j^{th} well sampled from sampling stratum *h* (also assumed known), *j*=1, …, *q* wells sampled, *w* is the sum of the weights of fish measured (converted from lengths), *m* is the number of fish measured, *n* is the number of fish counted for species composition, and g_1 represents the function of the sample means and proportions shown in curved brackets (*i.e.*, a function of only the *w*'s, *m*'s and *n*'s).

Similarly, the estimated total number of fish of species *i* in length interval *k* of sampling stratum *h* is given by:

$$\hat{N}_{hik} = \hat{N}_{hi}\hat{f}_{hik}$$

$$= W_h \left[\frac{\sum_{j=1}^{q} W_{hj} \left(\frac{\frac{m_{hijk}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}}{\sum_{i=1}^{\# sp} \frac{w_{hij}}{m_{hij}} \frac{n_{hij}}{n_{h.j}}} \right)}{\sum_{j=1}^{q} W_{hj}} \right]$$

$$= W_h \frac{\left[\sum_{j=1}^{q} W_{hj} \cdot g_2(w_{hij}, m_{hijk}, m_{hij}, n_{hij}; i = 1, \ldots, \#sp)\right]}{\left[\sum_{j=1}^{q} W_{hj}\right]} \qquad (2)$$

where $\hat{N}$ is the estimate of the total number of fish in the stratum, $\hat{f}$ is the estimate of the species fraction, derived from numbers of individuals, and g_2 represents a function of the sample means and proportions only. Thus, the weighting for both estimators is the same; they differ only in the form of *g*, the function of the sample means and proportions.

4. PROPOSED APPROACH

4.1. Candidate poststratified estimators

Two candidate poststratified estimators of catch by species in poststratum *c* are presented below which preserve the specific function g_1 of equation (1). The following deals specifically with poststratified estimators of total species catch; poststratified estimators of the total number of fish in a length interval (equation (2)) would follow by replacing g_1 with g_2.

The first poststratified estimator of total species catch was developed under the assumption that both the sampling strata and the poststrata may contain important information with respect to the estimation of catch. In other words, within a poststratum *c*, the distinction between sampling strata, or fractions thereof, is preserved such that effectively poststratum c is further subdivided by the sampling strata $\{h\}$. This first poststratified estimator was obtained by following the general approach of Valliant (1993) for data collected using stratified two-stage sampling, but modified to address an important difference between the examples of Valliant (1993) and the port-sampling data. Specifically, in the case of the port-sampling data, fish in an individual well sample cannot come from more than one sample stratum and are assumed not to come from more than one poststratum; in contrast, clusters of Valliant (1993), equivalent to wells, could span poststrata.

The first poststratified estimator of the total catch of species i in poststratum c, $\widehat{W}_{ps-I;\,ci}$, I s the sum of catch estimates of species i from entire sampling strata, or parts thereof, that belong to poststratum c. In other words,

$$\widehat{W}_{ps-I;\,ci} = \sum_{h:\,h\cap c} W_{h\cap c} \frac{\left[\sum_{j\in h\cap c} W_{hj} \cdot g_1(\ldots)\right]}{\left[\sum_{j\in h\cap c} W_{hj}\right]} \qquad (3)$$

where the outer summand is over sampling strata h that intersect poststratum c, $h \cap c$ refers to the region of sample stratum h that is also in poststratum c, and $W_{h\cap c}$ is the total fishery catch in that region. Note that this poststratified estimator would preserve the sampling strata $\{h\}$. For example, in Figure 1c is an example of four spatial poststrata (A-D) and it can be seen that spatial sampling strata 5, 7, 11, and 12 would be bisected to create the hypothetical inshore poststratum D, whereas sampling strata 6 and 13 are totally contained within poststratum D. $\widehat{W}_{ps-I;\,ci}$ might prove problematic if there are many h for which each $h \cap c$ is small and contains few samples.

The second postratified estimator follows from disregarding the sampling strata and uses instead the approach taken for two-stage simple random sampling (*e.g.*, Thompson 2002). This might be considered reasonable given the opportunistic nature of the actual sample collection. This results in a poststratified estimator identical in form to equation (1), but with W_h and W_{hj} replaced by W_c and W_{cj} , respectively. In other words, this second poststratified estimator of the total catch of species i in poststratum c, $\widehat{W}_{ps-II;\,ci}$, is given by:

$$\widehat{W}_{ps-II;\,ci} = W_c \frac{\left[\sum_{j=1}^{q^*} W_{cj} \cdot g_1(\ldots)\right]}{\left[\sum_{j=1}^{q^*} W_{cj}\right]} \qquad (4)$$

where q^* is the number of samples in poststratum c.

4.2. Selecting the specific estimator

To select between the two poststratified estimators (*i.e.*, between equation (3) and equation (4)) it is proposed that generalized linear models be fitted to the port-sampling data (average weights, species fractions) to evaluate whether it is necessary to retain sampling strata within poststrata. For average weight, a global evaluation of the utility of the sampling strata within poststrata can be obtained by fitting the following two models (by species) and comparing the relative performance of the two models for each species with the Akaike Information Criterion (AIC; Burnham and Anderson, 2002):

$$\text{mean}(w_{chij}/m_{chij}) = \text{overall mean} + \text{poststratum effect}$$

and

$$\text{mean}(w_{chij}/m_{chij}) = \text{overall mean} + \text{poststratum-sample stratum effect}$$

where 'poststratum-sample stratum' indicates sample strata within poststrata. The same global evaluation can be done for species composition by fitting a logistic regression model (polytomous response) to the species fractions (with the same independent variables as above). The sample average weights and species fractions are components used to evaluate g. If the models with the poststratum-sample stratum effect lead to only a small reduction in AIC over the respective models with the simpler poststratum effect, this would suggest that within poststrata the sampling strata could be disregarded. This is plausible for poststrata that are defined based on analysis of spatiotemporal structure in length-frequency distributions and CPUE trends. The above generalized linear models would be fitted with weights equal to the individual well catch amounts (to be consistent with the ratio estimator weighting). In addition, and particularly if the fitted models with poststratum-sample stratum effects result in large reductions in AIC, the utility of sampling strata within poststrata will be further studied by fitting generalized linear models to the average weights and the species fractions within each poststratum (the independent variable would be a categorical sample-stratum effect). If necessary, tests of the significance of the sample-stratum effect by poststratum can be combined across poststrata to get an overall p-value with the truncated product method (Zaykin *et al.* 2002) for multiple comparisons.

4.3. Feasibility

The feasibility of implementing a poststratified estimator depends in part on the spatial resolution of both the total landed catch (for all species combined) and the port-sampling data. The total landed catch is allocated to the sampling strata using information from observer data and vessel logbooks. Observer and logbook spatial information is recorded in terms of latitude and longitude, with a coarsest resolution of 5° area (unless completely unavailable). Therefore, the total landed catch generally will be known equally well with respect to the spatial sampling strata as for any spatial poststratification that is similarly derived from combinations of 5° areas. The one exception to this can occur for samples associated with fishing in the Gulf of California (*i.e.*, the inner-Gulf region of sampling stratum 8; Figure 1a).

If the spatial poststrata are large, it is anticipated that most, if not all, port-sample spatial information will be known to the spatial poststratum level. At the time of collection of the port-sampling data, not only the sampling area but also the 5° area have been recorded for most samples since 2000. A preliminary comparison of the 5° areas of port-samples to actual set positions of sets that went into the sampled wells indicates that about 82% of all samples from 2000-2009 were in agreement with actual set positions at the level of the 5° area, and about 98% of all samples were within one 5° square area of the corresponding set 5° square area.

It is noted that for a spatial poststratification that does not match the sampling stratification areas, there may be vessel wells with catch that came from multiple spatial sampling-strata (and so are not sampled) but might have been from just one spatial poststratum. These vessel wells would not be sampled. However, this is consistent with the assumption presently made that the sampled vessel wells adequately represent the fishery.

5. FUTURE WORK

5.1. Treatment of sampling strata with catch but insufficient samples

On average, annually since 2000, approximately 20% of the total landed catch is associated with sampling strata for which it was not possible to obtain port-sampling data (due to logistic constraints) or for which fewer than two samples were available. Once an appropriate poststratification of the data is determined, the poststratified estimator may simplify the treatment of landed catch in sampling-strata with no sampling data. For example, assuming large spatial poststrata, if the poststratified estimator used has the form of equation (4), then it is likely that catch will always be associated with sampling data in a poststratum area. Exact treatment under poststratification of catch from sampling-strata with no sampling data will be investigated in the future.

5.2. Variance estimation

Approximate estimates of variance of the poststratified estimator will initially be computed using a bootstrap 'half-sample' procedure (Efron 1982) where one less than the number of samples in a stratum is drawn (with replacement) in order to remove bias (underestimation of the variance). If the overlap of sampling-strata and poststrata are limited for some poststrata, $W_{ps\text{-}I}$ may have larger variance due to smaller sample sizes than $W_{ps\text{-}II}$, but perhaps be less likely to be biased if the sampling-strata remain relevant within poststrata. Resampling will be done from the empirical distributions of lengths and species composition, instead of implementing parametric bootstraps. Three different resampling scenarios will be considered, the first using the sampling strata, the second using the poststrata, and the third using the sampling strata within the poststrata. This procedure does not include a finite population correction factor. However, given that the estimated annual level of sampling coverage of the catch is roughly 8% (computed as the sum of catch in sampled wells divided by the total fishery catch), this is probably not critical. In the future, however, other variance estimation procedures will be explored (*e.g*., resampling procedures for survey data that include a finite population correction; *e.g*., Sitter 1992), and an estimate of sampling coverage will be derived based on number of vessel wells.

5.3. Assessing feasibility for pole-and-line and pre-2000 purse-seine port-sampling data

As with the recent port-sampling data, the feasibility of poststratifying the pre-2000 port-sampling data collected from purse-seine and pole-and-line vessels will depend in part on the resolution of the spatial information associated with these samples. This will be investigated in the future.

ACKNOWLEDGEMENTS

We thank Nancy C. H. Lo for helpful discussions.

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.

Anonymous, 2010. Tunas and billfishes in the eastern Pacific Ocean in 2009. Fishery Status Report 8.

Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer.

Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics.

Holt, D. and Smith, T.M.F. 1979. Post stratification. Journal of the Royal Statistical Society, Series A 142, Part 1: 33-46.

Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A. 2010. Preliminary analysis of spatial-temporal pattern in bigeye tuna length-frequency distributions and catch-per-unit-effort trends. Document BET-01-02a, External Review of IATTC Bigeye Tuna Assessment, 3-7 May, 2010.

Maunder, M.N. and Aires-da-Silva, A. 2010. Status of yellowfin tuna in the eastern Pacific Ocean in 2008 and outlook for the future. Inter-American Tropical Tuna Commission Stock Assessment Report 10, pages 3-109.

Sitter, R.R. 1992. A resampling procedure for complex survey data. Journal of the American Statistical Association 87: 755-765.

Suter, J.M. 2008. An evaluation of the area stratification used for sampling tunas in the eastern Pacific Ocean and implications for estimating total annual catches. Masters thesis, San Diego State University, San Diego, California, USA.

Thompson, S.K. 1992. Sampling. Wiley.

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.

Valliant, R. 1993. Poststratification and conditional variance estimation. Journal of the American Statistical Association 88: 89-96.

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

Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H. and Weir, B.S. 2002. Truncated product method for combining p-values. Genetic Epidemiology 22: 170-185.

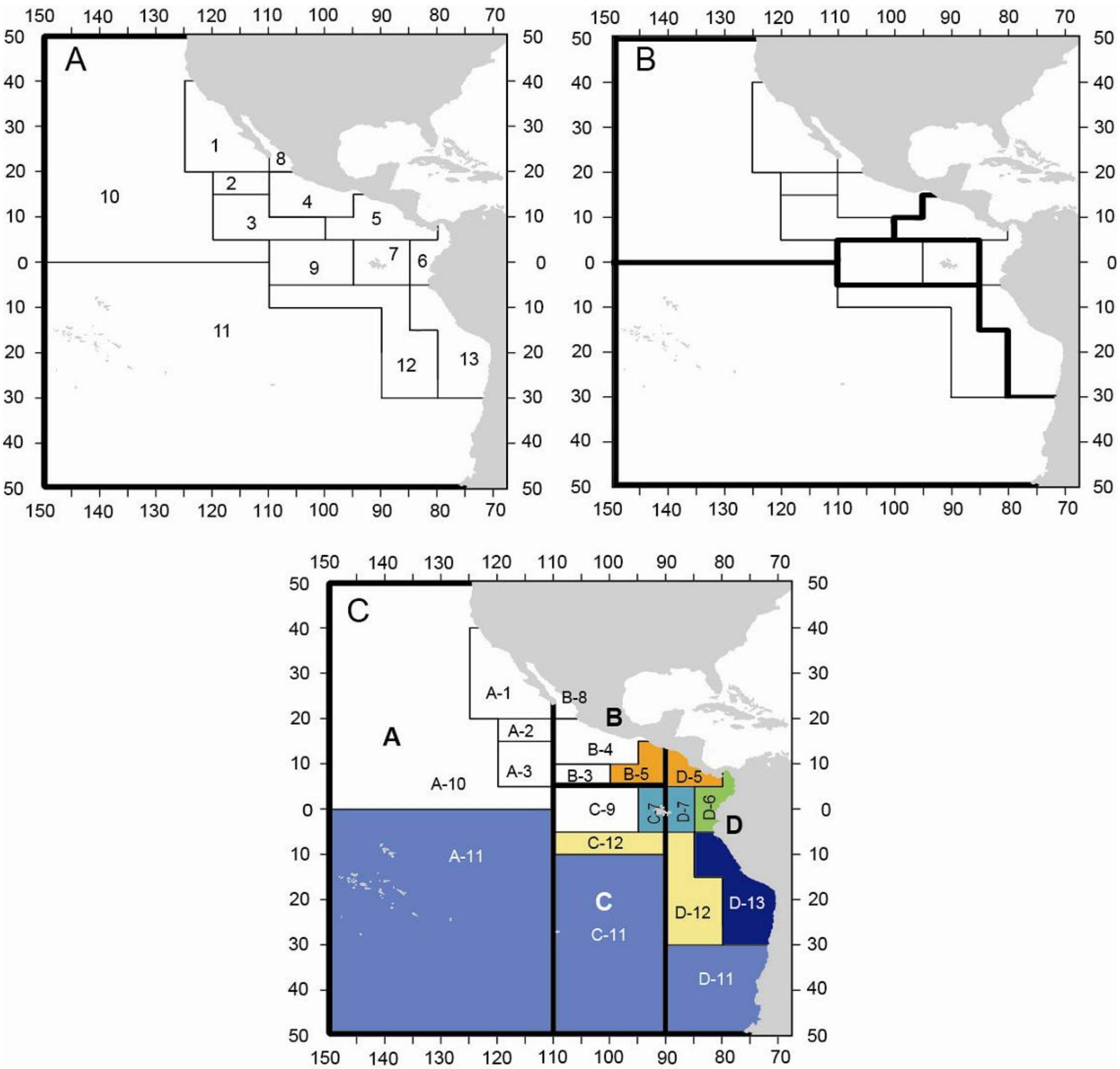

FIGURE 1. a) Sampling strata (Tomlinson 2004); b) stock assessment strata for bigeye tuna in floating-object sets (Aires-da-Silva and Maunder 2010); and c) an example of hypothetical spatial poststrata (A-D) based on results of a multivariate regression tree analysis of bigeye tuna length-frequency distributions (Lennert-Cody *et al*. 2010). The spatial sampling strata that are partially or totally contained in poststratum D are colored to illustrate spatial sampling strata that cross poststratum boundaries.

FIGURA 1. a) Estratos de muestreo (Tomlinson 2004); b) estratos de la evaluación de la población de atún patudo en lances sobre objetos flotantes (Aires-da-Silva and Maunder 2010); y c) un ejemplo de posestratos espaciales hipotéticos (A-D) basados en los resultados de un análisis de árbol de regresión multivariable de las distribuciones de frecuencia de talla de atún patudo (Lennert-Cody *et al*. 2010). Se coloran los estratos espaciales de muestreo parcial o totalmente contenidos en el posestrato D para ilustrar estratos espaciales de muestreo que cruzan límites de posestrato.