APAIS data calibration methodology report

John Foster
National Marine Fisheries Service

F. Jay Breidt
Colorado State University

Jean D. Opsomer
Colorado State University

March 11, 2018

1 Background

In 2013, new design and estimation procedures were implemented for the Access Point Angler Intercept Survey (APAIS) conducted by the National Marine Fisheries Service (NMFS). The new procedures were introduced as part of on-going efforts to improve the statistical validity and reliability of the recreational marine fisheries estimates produced by NMFS, which followed the recommendations of a National Academies of Sciences panel review (Sullivan et al. 2006). The most important design changes include improved protocols for interview assignments in terms of interview sites and times of day, and changes to the randomization of assignments so that they better covered the target population, again in terms of sites and times of day. Associated with those design changes were changes to the estimation methods, which are now fully weighted to reflect the unequal probability sampling design. APAIS data collected since March (wave 2) 2013 follow the new design and estimation procedures.

APAIS data have been collected since 1981, and NMFS staff clearly recognize the importance of preserving the integrity of the time series of catch estimates despite these design and estimation changes. Because of this, an adjustment procedure was developed to create "pseudo-weights" for APAIS data collected between January 2004 and March 2013. These weights were constructed based on a combination of site pressures and empirical site visit frequencies, and on estimated expected fractions of trips that took place during the time the interviewer was on site relative to the daily total number of trips. Weighting the observed trips on a given site-day assignment by the inverse of this estimated fraction was meant to correct for differential representation of sampled trips within site-days. The fractions of trips were predicted by a small area estimation model fitted to data from the

Coastal Household Telephone Survey (CHTS), see Hernandez-Stumpfhauser et al. (2016). The combination of modeled site-day selection probabilities and within-day probabilities resulted in weights that better reflect the population of trips during the period 2004-2013, with respect to its overall size and distribution across states, waves and modes.

However, an implicit assumption underlying the validity of this approach is that the trips occurring during the time period the interviewer is on site are representative of those that take place during the full day. This is satisfied if either the time on site is randomly selected within the day, or the trip characteristics are not related to the time of day. The first condition was definitely not satisfied, because the large majority of site visits were made at what was considered the busiest time of the day and were also subject to a degree of interviewer discretion. The second condition appears not to be satisfied either, according to analysis reported in a technical report (see MRIP Staff 2014). Hence, there is a need to supplement the weighting procedure that is based solely on fraction of daily trips within selected site-days by a procedure that accounts for differences in trip characteristics between those that were observed during the site visit intervals and those outside of it.

There is also a desire to adjust the time series for the period prior to 2004. For that earlier period, not only are the selection probabilities within site-days unknown as above, but information allowing the construction of site-day visit probabilities is incomplete or missing, with the required design information becoming progressively more limited going back in time. Further complicating matters, the sampling procedures, including site selection and sampling intensity, underwent changes during that period, and documentation for these changes is no longer available. Hence, separate procedures are needed to calibrate estimates prior to 2004.

Correcting time series of survey data following changes in design, data collection and/or estimation methods is a challenging statistical issue. The "gold standard" approach involves conducting side-by-side measurements under the old and new methods, fitting a suitable calibration model relating estimates under both methods, and developing and applying adjustment factors based on the model results. This approach is currently being implemented by NMFS to calibrate the trip estimates obtained under CHTS and its replacement survey, the Fishing Effort Survey (FES). See NMFS Staff (2015) for more details on the CHTS and FES surveys and the transition between them.

While explicit statistical calibration would in principle be attractive for the APAIS time series as well, there are a number of reasons why that is not possible. First and most critically, there is no overlap period between the old and new designs, so that the data needed for fitting a calibration model are not available. Second, unlike CHTS and FES, which primarily involve estimating the total number of trips for a given region and time period, APAIS is used to produce numerous different estimates, covering a wide range of trip characteristics and detailed catch by species, location and type. Hence, even if an overlap period were available, it is not clear that statistical calibration would be feasible, since multiple models would likely be required for the different types of variables.

For these reasons, the proposed APAIS calibration will rely on a weight adjustment approach, which is conceptually similar to the pseudo-weight approach described above for the 2004-2013 data. By adjusting weights rather than modeling the estimates themselves, the data collected prior to 2013 are preserved but their weights are suitably modified so that the distribution of trips better reflects the actual population distribution. By incorporating the calibration adjustments into the survey weights, the historical data can continue to be made available as survey public-use (micro) datasets, greatly facilitating their acceptance among the current data users.

2 Adjustment approach for 2004-2013 data

We first consider the adjustment of the weights for the period 2004-2013 (wave 1). Because there is no overlap period between the old and new designs and the CHTS contains only limited information on trip characteristics, no direct comparison distribution is available on which to calibrate. Instead, calibration will be performed using the trip distribution for the period 2013 (wave 2)-2016 as the target distribution. This is reasonable if the mix of trip characteristics has remained constant over time, at least over the periods being considered. The validity of this assumption cannot directly be assessed, because differences in observed trip characteristics before and after 2013 can be explained by both the design and estimation changes as well as by actual changes in the fishery. However, we will modify the proposed method in situations in which we observed a significant "drift" in important trip characteristics over time, see Section 3 below. For now, assume that it is reasonable to work under the assumption that differences in trip characteristic distributions between the periods 2004-2013 and 2013-2016 are likely primarily due to the design and estimation method changes. Hence, the weight adjustment method will calibrate the weights for trips in 2004-2013 (wave 1) to the weight distribution for 2013 (wave 2)-2016.

The key decision in the proposed method is which trip characteristics to adjust for. Following the analysis results shown in MRIP Staff (2014), the following trip-level variables were identified as both important trip characteristics and ones for which the distribution in the data collected prior to 2013 deviated from those under the new methods:

- state and sub-state region (if applicable)
- year and wave

- mode
- area fished
- coastal/non-coastal household
- for-hire boat frame membership.

The values for each of these variables defines categories of trips. For instance, there are 4 modes (shore, private boat, headboat, charter) and each trip belongs to one of those modes. Taken in combination, the values for these variables define large numbers of trip domains. Under the new design and estimation approach, by summing up the weights of the trips corresponding to a given set of values for these variables, we obtain an estimate of the number of trips of that type.

As an example to explain the adjustment procedure, let $U_{D,2014}$ present the domain of all trips that occurred, say, in a particular substate region in Florida during wave 2 on a private boat by a coastal household in state waters, in 2014. The true total number of such trips that took place is equal to $N_{D,2014}$. It is unknown but it can be estimated based on APAIS intercepted trips under the new design and estimation methods, by $\widehat{N}_{D,2014} = \sum_s w_i I_{\{i \in U_{D,2014}\}} = \sum_{s_{D,2014}} w_i$. This can be repeated for any combination of values of the classification variables. However, while statistically valid, these estimates are likely to be quite variable for some of these domains because they contain only small numbers of observed trips.

Likewise, we can compute estimates for the same domains for years prior to the design change, e.g. 2012: $\widehat{N}_{D,2012} = \sum_{s_{D,2012}} w_i$. This estimate might not be valid, however, because of the recognized shortcomings of the design and estimation methods in effect at that time. If $N_{D,2012}$ were known, we might therefore decide to adjust the weights so that they sum up to $N_{D,2012}$. This is readily accomplished by replacing all w_i for $i \in s_{D,2012}$ by

$$w_i^* = \frac{N_{D,2012}}{\sum_{j \in s_{D,2012}} w_j} w_i. \tag{1}$$

The weights w_i^* in sample domain $s_{D,2012}$ now sum to the new control total $N_{D,2012}$, and can be applied to any variable y_i collected in the survey. This type of calibration to known control totals is commonly applied in surveys, to improve the precision of estimators.

Since we do not know $N_{D,2012}$, implementing this ratio-type adjustment requires that it be replaced by a sample-based quantity. As noted above, we propose to use estimates based on the data collected under the new design since 2013. In order to reduce the variability of the control total estimates and also because individual years are not meaningful targets (i.e. we are not interested in adjusting 2012 weights to match the 2014 totals, but rather, adjust pre-2013 years to post-2013 years), both the control targets and the adjustment ratios are averaged across years. Hence, the unfeasible adjustment in (1) is replaced by

$$w_i^* = \frac{\widehat{N}_{D,\text{new}}}{\widehat{N}_{D,\text{old}}} w_i, \tag{2}$$

where $\widehat{N}_{D,\text{new}}$ is the average of $\widehat{N}_{D,2013}$, $\widehat{N}_{D,2014}$, $\widehat{N}_{D,2015}$ and $\widehat{N}_{D,2016}$ (with the first of these omitted if the domain is in wave 1) and $\widehat{N}_{D,\text{old}}$ is the average of $\widehat{N}_{D,2004}, \ldots, \widehat{N}_{D,2012}$ (and $\widehat{N}_{D,2013}$, only if the domain is in wave 1). Unlike the (unfeasible) adjusted weights in (1), the weights w_i^* do not sum to a control total for a particular year. Instead, they correct for the overall under- or over-representation of trips in domain U_D under the old design and estimation methods relative to the new methods implemented since 2013, which is expected to lead to improved estimates for variables of interest that are related to the domain that is being adjusted.

While averaging the adjustment ratios across years as in (2) reduces their variability, the fine definition of the domains (as intersections of numerous control variables) is still expected to lead to unreliable adjustments in many domains. Therefore, the full ratio adjustment in (2) is replaced by a *raking ratio* adjustment, originally proposed in Deming and Stephan (1940) and widely used in survey calibration. The motivation for this procedure is that instead of adjusting at the finest domain level, adjustments are made iteratively on a set of a coarser domains. These coarser domains are determined by a subset of the variables mentioned above. For each of them, it is possible to compute the averages of the annual estimates as described for $\widehat{N}_{D,\text{new}}$ above. We denote the ones we use in our adjustment procedure as follows:

- AF (state, wave, mode and area fished): $\widehat{N}_{D,\text{new,AF}}$
- HS (state, wave, mode and coastal/non-coastal household status): $\widehat{N}_{D,\text{new,HS}}$
- FH (state, wave, mode and for-hire boat frame status): $\widehat{N}_{D,\text{new,FH}}$
- RE (state, wave, mode and substate region): $\widehat{N}_{D,\text{new,RE}}$

While not explicit in this notation, for each of these domains, the averages are for each of the categories of these variables. So for instance, $\widehat{N}_{D,\text{new,AF}}$ are averages of estimates for each state-wave-mode-area fished combination, and so on for the other domain definitions above.

The raking ratio algorithm, also sometimes called iteratively proportional fitting, then proceeds as follows:

1. Initialize: set $t = 0$, set the adjusted weights $w_i^{(t)}$ equal to the initial weights w_i for the period 2004–2013 (wave 1), and compute the $\widehat{N}_{D,\text{new,AF}}$, $\widehat{N}_{D,\text{new,HS}}$, $\widehat{N}_{D,\text{new,FH}}$ and $\widehat{N}_{D,\text{new,RE}}$.

2. Let $\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$ be the averages of the estimated AF domain totals for the period 2004–2012 (include 2013 for wave 1) using weights $w^{(t)}_i$, compute the ratios $R^{(t)}_{\text{AF}} = \widehat{N}_{D,\text{new},\text{AF}}/\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$, and set $w^{(t)}_{i,\text{AF}} = R^{(t)}_{\text{AF}} w^{(t)}_i$.

3. Starting from the weights $w^{(t)}_{i,\text{AF}}$, do the same as in 2 for the HS domains, resulting in ratios $R^{(t)}_{\text{HS}}$ and weights $w^{(t)}_{i,\text{HS}}$.

4. Starting from the weights $w^{(t)}_{i,\text{HS}}$, do the same as in 2 for the FH domains, resulting in ratios $R^{(t)}_{\text{FH}}$ and weights $w^{(t)}_{i,\text{FH}}$.

5. Starting from the weights $w^{(t)}_{i,\text{FH}}$, do the same as in 2 for the RE domains, resulting in ratios $R^{(t)}_{\text{RE}}$ and weights $w^{(t)}_{i,\text{RE}}$.

6. Set $w^{(t+1)}_i = w^{(t)}_{i,\text{RE}}$.

7. Repeat steps 2–6 until convergence, which is evaluated by measuring the change in the ratios $R^{(t)}_{\text{AF}}, R^{(t)}_{\text{HS}}, R^{(t)}_{\text{FH}}, R^{(t)}_{\text{RE}}$ for different t. Set the final adjusted weights w^*_i equal to the iterated weights $w^{(t)}_i$.

This raking ratio procedure ensures that the weights w^*_i are adjusted to match each of the "marginal" raking variables (AF, HS, FH, RE), but not the fine domains defined by the combinations of these raking variables. This prevents adjusting to overly small domains, with associated overfitting and weight instability issues.

3 Modification for temporal changes in fishery characteristics

We now return to the assumption of constant trip characteristics over time. As noted, the raking procedure described in Section 2 is based on the assumption that the estimated trip distribution since 2013 is a reasonable target for the trip distributions prior to 2013. However, if the trip characteristics in the fishery have changed over that time period, observed differences between the pre-2013 and post-2013 periods are likely to be due to a combination of the design-estimation changes and actual fishery changes. Raking as in Section 2 in this situation will result in a weight adjustment that is too large, because it will remove both the design-induced change and the actual fishery change. We therefore implemented a two-step procedure to decrease the risk of over-adjusting the weights, described in this section.

Consider a single set of control domains first, say AF above. Prior to raking, for a given state, mode and area fished, we create a dataset containing the estimated domain totals for each year and wave combination between 2004 and 2013 (wave 1), resulting in a time series of 145 data points. There are multiple such time series, for each combination of state, move and area fished. We perform a simple linear regression of the totals against a time index for each time series, and test whether the slope is significantly different from zero at the 97.5% confidence level. If the null hypothesis cannot be rejected for a given time series, we maintain the raking adjustment described in Section 2 for the AF domains. If the null hypothesis is rejected for a time series, step 2 in the raking algorithm is modified for that particular control domain, so that only the years 2010–2013 (wave 1) are used in the computation of $\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$. Hence, the AF ratio adjustment $R^{(t)}_{\text{AF}} = \widehat{N}_{D,\text{new},\text{AF}}/\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$ is only based on the most recent years instead of the full time period, in those domains for which a significant time trend is detected.

The same testing and modifications are applied to the remaining three control domains (HS, FH, RE). The full adjustment procedure that accounts for temporal trends therefore consists of the linear regression tests followed by shortening of the time period used for computing the ratio adjustments in any of the control domains for which a non-zero slope is detected, following by the raking algorithm.

4 Adjustments for prior periods: 1993-2003

Weight adjustments for data collected prior to 2004 were performed following the computation of the adjusted weights for 2004-2013 (wave 1). The major difficulty for the earlier periods was that unlike for 2004-2012, it is not possible to construct meaningful initial sample weights for the APAIS data. As such, the weight adjustment method described in sections 2 and 3 could not be applied directly and needed to be extended to address effects of the 2013 APAIS design change as well as any effects associated with initial weighting of the 2004-2012 APAIS data. Using 1 as the initial base weight for intercepted angler-trips was not adequate as the sample sizes, in terms of sampled site-days, were known to vary considerably over time. Unfortunately, the exact sample sizes were unavailable for these earlier years.

It was decided to divide this period in two pieces overall. This provides a hedge against incorrect time trend adjustments masking actual changes in the fishery, as well as unaccounted-for changes in design. This is similar to the argument in Section 3, but was applied globally prior to any further adjustments. Hence, we performed the adjustments for 1993-2003 and 1981-1992 separately.

Considering first 1993-2003, we investigated two approaches for creating initial weights. In a first approach, these weights were calculated by using the MRFSS effort estimates as counts of angler-trips and dividing this by the number of intercepted angler-trips. This calculation was performed in cells defined by state, year, wave, mode, area fished and sub-state region. However, while these initial weights account for the overall magnitude of the fishing effort in a cell, they completely miss relative changes in the number and distribution of site-day assignments that occurred during this period. This lead to stability issues in the development of final weights.

Hence, a second approach was developed using counts of site-days with intercepts to account for changes in site-day assignments. For this approach, counts of site-days with intercepts were tallied in cells defined by state, year, wave and mode. While the exact sampling design was unknown, these counts are a useful proxy for it, in the sense that changes in the number of site-days in these cell over time very likely correspond to changes in the underlying sampling design.

In order to incorporate the design changes, the maximum count was identified within each unique combinations of state, mode, and wave across years. Initial weights at the angler-trip level in a state-year-wave-mode cell were calculated as the count of site-days with intercepts in that cell, divided by the maximum count for that state-wave-mode combination. Hence, for cells corresponding to the year with the maximum count, the angler-trip weight is set equal 1, and for any other cell, the weight is greater than 1.

Under this approach, the initial weights will not be correct for the total number of trips, since they only account for relative changes in the design over time. This is justified by the fact that the overall "scale" of the weights, accounting for the volume of angler-trips, is not of interest in APAIS estimation, in which only rates are estimated.

Starting from these initial weights, a raking algorithm was again implemented to create updated weights. As a further adjustment for unobserved design effects, several raking control domains were added to those used for the 2004-2013 period:

- KOD (state, wave, mode and kind-of-day)
- MG (state, wave, mode and month groups)
- AC (state, wave, mode and site activity class).

The first of these corresponds to the usual weekday-weekend/holiday classification of angler-trips, but the other two require further explanation. For the MG domains, raking was attempted using individual month cells, but there were cases that would not converge. Months were therefore grouped into three classes: (1) January, March, October, December; (2) May, June, July, August; and (3) February, April, September, November. Class 1 represents the traditionally lower activity month during transition periods (month 1 in waves 1 and 2, month 2 in waves 5 and 6). Class 2 represents the peak activity period when sample sizes are generally similar or equally allocated among months within waves 3 and 4. Class 3 represents traditionally higher activity month during transition periods (month 2 in waves 1 and 2, month 1 in waves 5 and 6). For the AC domains, sites are divided into two groups, high activity and low activity, based on annual counts of intercepts by fishing mode. Sites with counts above the annual mean within cells defined by state, mode, year and sub-state region were classified as high; sites at or below the mean were classified as low.

The raking algorithm described in Section 2 was applied including these additional control variables, with the adjusted estimates for period 2004-2013 (wave 1) as the "new" estimates and those obtained with the initial weights described above for the period 1993-2003 as the "old" estimates. The linear regression testing for trend described in Section 3 was also performed, but with the modification that it was applied for both the new and the old periods. For any domains where a trend was detected in the old period, the adjustment ratio was computed on the years 2001-2003 instead of on the full period. Similarly, for domains where a trend was detected in the new period, the adjustment ratio was computed using 2004-2006 instead of the full period.

5 Adjustments for prior periods: 1981-1992

The adjustment procedure for 1981-1992 follows the same procedure as that for 1993-2003. The initial weights are again created based on relative counts of site-day assignments, and the raking procedure uses the additional control domains described in Section 4. "New" estimates are those obtained with the adjusted weights for 1993-2003 and "old" estimates are those for 1981-1992. Significant trends resulted in shortening of the period used for the raking ratios to 1990-1992 for the old period and to 1993-1995 for the new period.

References

Deming, W. E. and F. F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics 11*, 427–444.

Hernandez-Stumpfhauser, D., F. J. Breidt, and J. D. Opsomer (2016). Hierarchical Bayesian small area estimation for circular data. *Canadian Journal of Statistics 44*, 416–430.

MRIP Staff (2014, August 20). A descriptive analysis of the Access Point Angler Intercept Survey 2013 design change. Internal report, National Marine Fisheries Service.

NMFS Staff (2015, May 5). Transition plan for the Fishing Effort Survey. Internal report, National Marine Fisheries Service.

Sullivan, P. J., F. J. Breidt, R. B. Ditton, B. A. Knuth, B. M. Leaman, V. M. O'Connell, G. R. Parsons, K. H. Pollock, S. J. Smith, and S. L. Stokes (2006). *Review of Recreational Fisheries Survey Methods*. Washington, DC: National Academies Press.