AN ADAPTIVE SAMPLING DESIGN FOR THE ESTIMATION OF THRESHER SHARK CATCH AND ANGLER EFFORT IN A RECREATIONAL FISHERY IN CALIFORNIA

Vincent Gallucci and Aneesh Hariharan

University of Washington, Seattle, WA

Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

The history of fisheries statistics is rich in sampling designs for the estimation of the catch of a given species or a group of species. The sampling design used is suggested by the distribution of the animals being sampled and the distribution of sample units. In commercial fisheries, the distribution of sample units is usually a relatively small number of portside facilities where catch is unloaded, in which case a sampler would monitor catch as it is taken off a vessel. The distribution of time spent portside is determined by the sample coordinator in such a way that the catch from different locations or sample units can be synthesized. The sample design would either consider each of the vessels to be a sample unit or each landing site to be a sample unit. This type of scenario applies to multistage sampling designs where the sizes of vessels might be weighted to take into account increased or decreased potential catch. In general, catch would be measured in terms of thousands of metric tons or thousands of fish.

The sampling design scenario described in this report for estimating catch and effort of a recreational fishery for common thresher sharks, is different in a number of ways. For one thing, it involves the capture of an uncommon species as a relatively rare event. In addition, in the case of commercial fishery port sampling, there is no assumption about the distribution of fish at sea since the vessels are assumed to have roved over wide ranging territory and not to have focused their fishing activity in the vicinity of the port. In the current situation, private recreational vessels are the predominant fishing platform, and the vessels are likely to have much less mobility away from their home marinas and launch ramps due to being relatively small and unsuited to fish far offshore. Thus, the capture of even one animal is considered to be a significant event. These factors combined say something about sampling intensity to record the capture of even one shark. Adding the presumption that the spatial and temporal distribution of the sharks adds predictability to which marinas and launch ramps are more likely see the return of vessels with sharks suggests possible improvements in the sampling design to estimate the number of sharks captured, in the sense that if they tend to appear in a clustered pattern in the ocean, then one might presume that this would be reflected in the vessels returning with sharks.

The annual longline survey conducted by the NOAA Southwest Fisheries Science Center suggests clustered distributions of thresher shark presence based on longline captures of neonates. Extrapolation of these data seem to suggest that marinas and launch ramps in the vicinity of the clusters will provide recreational fishers from these sites with higher probabilities of catching a thresher. Ideally, the sampling design chosen will reflect this biological and logistical knowledge.

An adaptive cluster sampling design is chosen as the optimal methodology for accurately representing the amount of effort spent and number of thresher sharks captured in the recreational fishery. The overall objective is to maximize the number observations that will contain a thresher shark targeting trip and/or positive thresher catch relative to the number of observations without thresher shark effort or catch. In the particular design chosen, "prior" information is used, based on sampling effort from a 5 year California Recreational Fishery Survey (CRFS) of possible landing sites for recreational boats. These landing sites are "marinas" which may have many boat slips, where the boats remain in the water, and “ramps” where many boats are pulled onshore onto trailers and taken from the scene, along with “man-made piers”, as well as other possible landing sites. It should be noted that most private access marinas in southern California are not sampled by CRFS technicians due to sampler access limitations (i.e., private boat slips can be behind locked gates and there is no state requirement authorizing mandatory sampler access). It is assumed that access to these sites will continue to be restricted and maybe catered to at some time in the future.

The initial CRFS sampling effort was spread over several years and the design was essentially simple random sampling. Those sites from that historical database that yielded significant numbers of threshers are denoted as ‘priors’. There are not many of these, only 28 spread from the Mexican border to north of San Francisco. Linking this with the previous discussions about thresher sharks being clustered in their coastal realizations suggests that these clusters occurred within private boat range off of these 28 sites. Of course within these 28 sites (marinas and landing ramps) there exist many more sites (~250) that did not correspond in the historical database to locations where significant thresher catch was landed. Past identity as a prior is presumed to be based upon biological or physical factors that thresher sharks might choose as clustering locations due to food availability, temperature, or some other criteria.

The cluster design used makes specific use of the 28 priors as well as the ~250 non-priors. Thus, all possible landing sites are utilized but in different ways. The first step divides the prior sites from S-N into 8 strata of different sizes as measured by the distribution of priors. Eight is chosen based on observation of geographical distances between prior sites that were spread from the Mexican border to Point Conception. It is not necessary for the number of strata to be 8, it is more of a guideline derived solely from the geographical distance between priors across the entire coast of California. Therefore some strata will have more priors than others and will be surrounded by many of the other (~250) sites. The stratification is not unique but rather attempts to place priors into groups within a stratum boundary that usually occur where there is a sufficiently large gap between priors. In effect, it is doubtful that a slightly different specification of the stratum boundaries would make any difference whatsoever in the sampling results. With additional information on financial constraints, initial survey data, along with required levels of precision in estimates, one could derive an optimal sample size per stratum.

The sampling here may be considered to be multistage in that there is successive sampling from the distribution of the prior sites and from the non-prior sites. In the end, the number of “hits” or thresher catch found in each stage must be added together. It certainly follows that in sampling from the first stage (priors) all sample units must be equally likely, as is the case in sampling the non-priors. Thus far, the adaptive aspect has not arisen. Further, as in the historical database, it is necessary for the sample plan to draw samples from locations over the length of coastal California which does make for challenging logistics.

Thus far the focus is upon the spatial distribution of sample units of available marinas, man made piers, and landing ramps of two different types and has not included a temporal component. This component enters with the introduction of the adaptive elements. A part of the logistical problem is that the adaptive elements occur on successive sample days. It is assumed that a maximum number of sample days exist due to financial constraints and thus the adaptive sampling cannot be carried out indefinitely.

In the document a stopping rule of a fixed number of days to carry out adaptive sampling is introduced. Once the stopping rule is reached, the entire adaptive sampling process is is re-iterated. In this design, the same number of sample days is used as was used in the historical sampling database, viz, 6-8 sample days per month. This number was also the agreed upon number in the Carlsbad Oct 21st meeting. This is one of those decision nodes where financial constraints enter.

The procedure is as follows: On day 1, a sample is taken from the distribution of priors over the whole coastal range. Samplers are literally stationed at n* sites for the whole day. It is believed that having covered the whole coastal range with the initial sample it is likely that if threshers are available somewhere, hits will occur in a way that also reflects the history of hits. A random component is in the sampling from the selection of sample units identified as priors. A further element of randomness is introduced by sampling from the non-priors. Clearly criteria must be developed for the number of prior samples that are selected as well as for the non-prior sites that are selected. Here is another point where financial constraints enter. We developed a metric which takes into account the 'denseness' of a stratum in terms of the number of priors present to decide upon the number of random non-priors to choose for that particular stratum. It is again pointed out that this metric isn't unique, it is a result of estimating geographical distances between priors, pre-selecting strata and increasing sampling effort in strata (such as Southern California) where the priors are densely distributed. As a consequence, a greater number of random non-priors are chosen to be sampled in this stratum. The number of samplers available was a major topic of discussion in the Carlsbad meeting. A specific number was not identified and thus none is used in this report. In general, we assume that less than 15 would be available. These persons are located over the range of sites. Thus, e.g., a sample size of 3 prior sites (stage 1) and 7 non-prior sites (stage 2) over all strata are sampled. Hopefully, by sampling over the full range of possible occurrences, some "hits" occur. These allow the introduction of the adaptive component which then diverges significantly from the standard methods of Simple Random Sampling.

On day 2, the vicinities of sample sites with hits are resampled with the adaptive methodology of sampling neighboring sites. Guidelines for specification of rules for neighborhood expansions to form networks have been suggested, independently for estimation of catch and effort. This, of course, puts the second day sampling into only those strata where threshers were captured. The neighborhood sites may or may not yield additional hits. According to the biology, if it was true that clusters remained in coastal areas for more than a few days, one may assume that neighboring sites will yield additional hits. If this is the case, on day 3, more sampling will occur in the expanded neighborhoods. In addition to sampling in the neighborhoods on day 2, additional sites are chosen to sample from among the priors and separately, from among the non-priors.

On day 3, the neighborhoods which consisted of one neighboring sample unit on each side of the day 1 hits will be expanded with an additional two sites, one on each side of the old neighborhood. Sampling will occur in this expanded neighborhood. Ultimately, these expanded neighborhoods will construct a network of sites. It again should be noted that picking two neighborhood sites is merely a practical consideration related to financial constraints and the linearity assumption of sites spread across the California coast. One might well have picked two adjacent sample units on each side on either or both days. Thus, if the stopping rule is 3 days, sampling in the network stops after the third day. If one believes that the cluster of threshers has remained in the same vicinity over this period of time, if recreational vessels are being attracted to this apparent hotspot, and if finances permit, extending sampling to fourth and fifth days could occur. However, it is impossible to tell at this time how this will work at the practical level, but it appears to make logical sense as viewed from the perspectives of thresher and human behavior.

A few observations are possible now. First, all sampling is done without replacement so that it is not possible to be sampling the same location more than once in a time period ("Tau" in the text). All prior and non-prior sites are equally likely to be sampled, respectively. The appropriate equations for sampling without replacement are used in the estimators. It appears that expansion of the hits in the neighborhoods (networks) of the priors will amplify to the entire coastline in a straightforward way leading to unbiased estimators of catch. Similarly, simple random sampling of the non-priors and their networks will amplify to the entire coast in a straightforward way and these two estimates may be summed with their estimated variances. The sampling estimators presented in the following material do not depend upon the introduction of specific numbers as are used above, but are general and are easily adapted by a lead sample person, or "samplemeister". The use of specific or plausible numbers in the above suggests the possibility of a simulation, but also judging from the above, a simulation of the entire process would defeat its own purpose because of its complexity, and therefore is not done. Instead, the fundamental criterion, expressed as an inequality, that specifies the conditions under which adaptive sampling is more efficient than Simple Random Sampling is given in an appendix. We provide the implications for the use of adaptive sampling to illustrate when it is more efficient than simple random sampling. There is a tendency to calculate a "Catch per Unit Effort" (CPUE) in sampling designs, analogous to the same term applied to fishing efficiency. The entire section above is written solely in terms of "catch." However, the following statistical design has a significant component devoted to the estimation of effort, for no other reason than the tendency to look for the CPUE estimator. It is not clear how this estimator will be used in the sampling design whereas in the stock assessment sense it has clear implications.

The description of the above sample design is clearly different from that used in the CRFS Sampling Plan (see CRFS Manual 2009). In examining that sample methodology, certain characteristics of it appear to be improvable. The details of what was considered unsatisfactory about that methodology are not the focus of this report. Rather, the focus here is on development of estimators in the adaptive design presented. Further, this design will not include a step by step procedure for a sampling plan including distances between sites and other details unique to the California coast. This too cannot be done efficiently without the use of extensive details of each site. A possible criticism of this design is that it does not account for the differences in sample unit size between marinas, between landing ramps and between man-made piers. One might account for these differences by the use of sampling with probability proportional to size. However, done properly, one would also need to account for the time distribution of returning boats to marinas and landing ramps. Data on the relative sizes of the ~250 sites along with logistic constraints, such as time to commute from each site to neighboring sites depending upon the size, are not available.

In conclusion, this report is a hybridization of two situations. One parent is the introduction of one interpretation of the CRFS database. The other parent is an experimental design known as adaptive sampling which is a general sampling methodology, particularly useful for sampling from a rare population. The underpinning of the use of adaptive sampling is the team's interpretation of the biological field data for the capture of neonates, the observation that neonates seem to congregate off of certain landing sites for unknown numbers of days, and that fisherpersons take advantage of this in the distribution and intensity of fishing effort. Fundamentally, the foundation is a plausibility argument based upon interpretations of thresher shark and human behavior. An appendix is presented from which one may conclude the circumstances under which adaptive sampling is more efficient (lower variance) than simple random sampling; which it seems was the design used in the CRFS survey methods. Subject to the above, the necessary statistical sampling estimators for a stratified clustered adaptive design are presented in the text.

Introduction

This is a project report for a sampling design to estimate catch and effort in a recreational fishery on the common thresher shark, *Alopias vulpinus*. The geographic scope of the sampling area covers approximately 350 miles (ignoring the coastline north of Pt. Conception) during the months (April to October) when threshers aggregations are present and accessible to recreational vessels. This project had the advantage of drawing upon an existing dataset collected by the California Recreational Fisheries Survey (CRFS). The status quo CRFS sampling design is based on a Stratified Simple Random Sampling (SRS) methodology which was carried out annually in the years 2005-2009. After reviewing these data and methodology and observing the clustering of boats returning with thresher hits, along with understanding some of the spatial distribution of thresher shark schools off of (near shore) coastal California, it was decided that some form of cluster sampling was desirable vs SRS. Nevertheless, the difficulty with the use of cluster sampling in its usual context is that a very large proportion of the clusters may not contain any thresher hits. The choice of adaptive cluster sampling methodology seems to alleviate this problem. This is discussed by Pollard et al. (2002) in carrying out an adaptive line transect sampling design.

The design presented is based upon utilizing the existing dataset to define a collection of sample units are considered "prior sample units" within which at least one hit (i.e. where a thresher catch was sampled) has occurred between 2005-2009. All sample units are considered to lie linearly along the California coast. The coastline is divided into 8 strata from S-N, each of which contains one or more prior sample units in the collection. Since adaptive cluster sampling is a relatively new methodology, it became necessary to derive many of the estimators and their variances for the first time. In particular, this report derives these estimators for stratified adaptive cluster sampling, since the CRFS survey

This report reviews the general methodology of stratified adaptive cluster sampling. This method of sampling was introduced relatively recently and has been developed in several books, for example Thompson (1992) and Thompson and Seber (1996). Recent applications of this methodology include works by Smith, Brown and Lo (2004). The methodology is especially useful when events of interest are relatively rare and have some spatially clustered appearances. Fundamentally, when a target element of interest is found in a sample unit, additional sample units are taken in the vicinity (neighborhood) of the first. Although it may seem that the sampling design defies the requirement for randomization, this is not considered a problem since some of the sample units selected are chosen randomly. The relationship between adaptive and simple random sampling for a patchily distributed population is seen in Appendix F, where adaptive sampling is shown to be more precise when the 'within network variance' is large where a network is defined as a sample unit along with its set of neighbors. It is recognized it is premature to introduce such words as "networks" but it seemed important to note the increased efficiency of adaptive sampling early on.

In any complex sample design, execution for data collection can be more difficult than in Simple Random Sampling when sample sites can be determined *apriori* . Under the circumstances of this design a "samplemeister" must decide when additional samples must be taken, where and how many. This is not an easy decision because of the practicalities of the distances that may exist in a neighborhood (one or more adjacent landing sites surrounding the one where a thresher was landed) and because the design itself adapts to the data as it comes in and so the sample size is not pre-determinable. This report does not cover this issue. To do so would require information about roads and distances between sample units (i.e. landing sites), which is a different type of problem than the derivation of estimators for statistical analysis. It is therefore the recommendation from this report that the data collection aspect of the project be carried out taking into account the logistical and financial tradeoffs, bearing in mind that the larger the sample size, the more the cost, and the lower the variance.

To maximize the number of boats sampled in any sample unit, the design includes a new sample collection form that is easy for the sampler to complete and which addresses all of the parameters that need to be estimated. (Appendix D)

Methods for Adaptive Sampling

Some Definitions

The database from CRFS was used to define the necessary sample design quantities, including what is a Sample Unit (SU) or a Sample Site, what is a stratum, what is a collection of priors, what is a neighborhood etc.

A sample unit is a boat launching/landing site such as a marina, a landing ramp, a man-made pier etc. that is accessible by CRFS samplers (i.e., not behind locked gates) A stratum is a collection of sample units demarcated by geographical boundaries. The size of each stratum is defined by a metric, to be discussed later in this report. A collection of prior sample units is some or all sample units that yielded at least one thresher "hit" or a thresher catch in the past 5 years. A neighborhood is a collection of sample units, formed by expanding initial sample units to networks where the probability of observing a thresher "hit" or thresher fishing effort is relatively high.

Distribution of Sample Universe

We pooled all the sites recorded from the U.S.-Mexico border to the California-Oregon border. In terms of the CRFS classification, this included all of PR1 (Primary Party and Private Rental sites), PR2 (Secondary Private and Party Rental sites), MM (Man Made Piers), BB (Beach and Banks with public launch sites) etc. This number was approximately 2000 sites and moreover, most of the sites were not sites where physical sampling could take place. Keeping this in mind, we compiled a list of PR1 and MM possible sites, the total of which was 28, all of which were known to have received hits in the past. Since, this was too small a number for any reasonable sampling design to take effect, we pooled the list of PR1, PR2 and MM sites to obtain a total of approximately 275 sites (Primary Party Rental (PR1), Secondary Party Rental (PR2) and Man Made(MM) sites from the CRFS database where sampling could potentially take place . The 28 PR1 and MM sites (Appendix C, included in the 250) that yielded thresher hits were choseas priors. We noted that, when lined up, these sites can be approximated by a N-S straight line. These sites were placed in increasing order of their decimal latitude equivalent. To get a sense of the north-south distance between sites, we assumed that latitude varies linearly, a realistic assumption, given the proximity to the equator, and used the conversion scale of 5 degrees=380 miles to convert the sites into absolute distances.

We observed the distribution of priors along the entire coast and used some intuition to define boundary lines to define different strata. We considered that sites such as Marina Del Ray need to be part of a stratum with a larger number of priors. Little bias was added by our selection of these boundaries, since natural clusters seemed to emerge. In this way, we classified the entire California coast into 8 strata (See Figures 1 & 2 and Table 1). It is noted here that this number is chosen based on catch sample data from the CRFS database to identify regions in which sampling units are as 'similar' to each other as possible (our beliefs in their respective catch activity levels) and from the annual longline survey conducted by the NOAA Southwest Fisheries Science Center that suggests clustered distributions of thresher shark presence based on longline captures of neonates. However, as discussed in the October 21st Carlsbad meeting, if there are compelling regions to believe that regions to the north of San Francisco have lower catches, depending on levels of funding, lower number of strata could suffice.

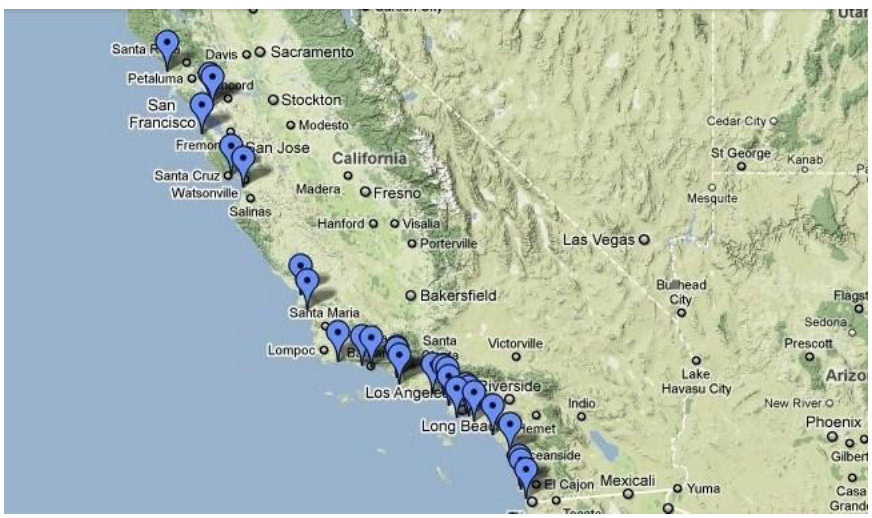


Fig 1: Distribution of prior sites from the Bay Area to the Mexican Border
The latitude of Imperial Beach Pier in the south is 32° 35' 2" N. The latitude of Santa Rosa in the north is 38° 26' 26" N.

Fig 2: Location of selected strata and the associated priors in each of them
The latitude of Imperial Beach Pier in the south is 32° 35' 2" N. The latitude of Bodega Harbor in the North is 38°19′ 25″ N.

Table 1: Strata, Associated Sites and Priors in each stratum

Stratum Number	Site Numbers	Number of Sites	Number of priors	Approx. Miles
1	1-40	40	3	42.03
2	41-95	55	7	43.5
3	96-120	25	2	15.9
4	121-162	42	7	36.21
5	163-180	18	3	106
6	181-200	20	3	64.85
7	201-225	25	2	20
8	226-275	50	1	38.61

It is worth noting that certain strata are much larger than others. For example, stratum 5 has a much larger N-S spread than other strata with a relatively low number of priors. Intuitively, this means that this stratum may not be expected to be a high yield region, so less sampling effort can be assigned to this stratum. More sampling effort should be concentrated on observing thresher catch in expected higher yield strata.

Estimation of per stratum sample size

As suggested in the literature, there is a debate about whether the sampling fraction should be kept a constant so that the number of samples to be taken within a stratum fall naturally equallyin place, or

whether the sampling fraction should follow some other pattern based on perhaps the stratum size, density of sampling sites, and/or expected catch. We decided against a constant sampling fraction, even though the analysis is simpler. This is undesirable because if a constant sampling fraction is used, the information on priors and the length of each stratum is not utilized. The number of sample units to be sampled within stratum i is given by n_i below. The selected sample units will be composed of all the prior sample units in that stratum plus additional randomly selected sample units.
Thus, we propose the following metric, by looking at the data:

n_i = Number of sites to be sampled within stratum i
= Ceiling$[(P_i) + D\ (S_i/D_i)]$

where

P_i = Number of prior sample sites in stratum i
D = Expected number of miles between sites, across all strata (a constant value approximately 1.335)
S_i/D_i = Number of sample units per mile in stratum i.

We assume that the number of priors in each stratum is a lower bound for the number of sites (=sample units) in that stratum. That is, the sample size for a stratum must be greater than or equal to the number of priors. Note that the equation for n_i is a "density", since it is the number of sample units per stratum. The following are potential choices of prior + random sites per stratum:

Stratum 1:- 3+1 = 4 sites
Stratum 2:- 7+2 = 9 sites
Stratum 3:- 2+2 = 4 sites
Stratum 4:- 7+2 = 9 sites
Stratum 5:- 3+1 = 4 sites
Stratum 6:- 3+1 = 4 sites
Stratum 7:- 2+1 = 3 sites
Stratum 8 :-1+2= 3 sites

It is important to note here that the above distribution of sites ('priors'+ random sites) is purely based from a geographical standpoint. While deriving estimators and variance of obtained estimates, it will be assumed that all the priors are pooled into a separate collection and sampling is performed amidst the priors to decide the number of sites to sample among the priors for each sample day. The estimates obtained from this collection along with the network formed by this collection will yield total population estimates that is exclusive only for this collection +associated networks. The reason behind this is that this collection of priors are believed to yield higher catches and hence while deriving population parameters, the expansion of mean estimators obtained for the collection of priors will be different from the other random units sampled. A set of random sample units are always chosen to allow for diversification of sample units selected.

Temporal and Spatial Considerations in Sampling Design

Assume the sampling experiment is performed for a period of time, say, 8 days (t) for the months April to October, when most recreational thresher fishing is expected from the CRFS database. Assume the possible number of sample days is 1 month (τ). Clearly, on each sampling day, samplers can only go to a finite possible number of sample units (n_s) amidst a total possible number of sample units (N_s) in each stratum. Note that here the term 'sample units' refers to a combination of public landing ramps, piers, other man made landing sites.

In any one stratum if, e.g., one boat returns with a thresher catch on day 1 then, on day 2, a second sample with two samplers is taken in the expanded neighborhood of the sample unit, one sample unit to the right and one to the left. If no captures occur on day 1, no neighborhoods are constructed. On day 2, samplers are sent to sample priors plus another set of randomly selected sample units. To save costs, there could also be reallocation of effort from strata which had no hits to strata which had a thresher hit.

Apparently, this is a zero sum game where a total amount of money is available and it must be allocated relative to the number of strata sampled, the intensity of sampling AND the time factor for number of days over which sampling must occur.

Derivations of Estimators

Estimates of total effort for the pooled sites described in the section 'Distribution of Sample Units' are calculated using the total number of boats sampled during the sampling , for each day-type (weekday or weekend). We propose two separate effort estimates: thresher as a target and thresher as a non-target catch. This recommendation is a result of analyses of the past 5 years of data where there is significant thresher non target catch in trip types in contrast to catch where thresher was targeted (Appendix A has an illustration of sample data extracted from the CRFS database). The two different estimators of effort are based upon data collected using the newly designed Catch and Effort Sampling Form that was developed. (Appendix D). It is noted that Highly Migratory in the database is the same as "targeted thresher" given that it is the lowest resolution available for a trip type targeting threshers.

Boat Trip Effort: Thresher is targeted

It is necessary to derive estimators that satisfy the unique requirements of adaptive sampling. In particular, the estimator must account for the development of networks of sampling units. In fact, the effort must even be represented where neighborhoods overlap (neighborhoods however do not cross strata boundaries, except in the case of the 'stratum' formed by the collection of priors and units sampled from them. This assumption reduces the complexity of computing the estimators, especially variance estimators, since the total variance of our estimate for total (effort or catch) across all strata is the sum of the variances of each of the estimators for each stratum.

We develop estimators along the lines of Thompson (1992)(Pg. 134-138). Note that the sample units are approximately distributed on a straight N-S line along the California coast; thus, neighborhood expansions are one and not two-dimensional (N-S, not E-W).

Boat trip effort is defined to be the number of boats returning to a sample site. This section defines the estimators that estimate the total boat effort, and its variance, from the sample data collected. There are two categories: when thresher is the targeted fish and when they are caught incidentally. This section deals with developing estimators for the case when thresher is targeted.

Let $\hat{\hat{e}}_{sh}$ = Estimate of the mean number of boat trips per sample unit s, returning with thresher as target, across all sampling days in stratum h .

Let f_{ih} = Number of boat trips observed returning, with thresher as target, across all sampling days in the i^{th} sample unit in stratum h. (NOTE: This formulation is assuming that the parameter is not an estimate but an accurate number, i.e., no returning boats are missed, which we assume is feasible given the new revised sampling form (Appendix D).

Let m_{ih} = Number of sample units in the network containing sample unit i in stratum h. Therefore, using the technique of Initial Number of Intersections for Adaptive Sampling from Thompson and Seber (1996), (Formula 4.14, Page 98) we define

$$\hat{\hat{e}}_{sh} = \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{f_{ih}}{m_{ih}} \ . \qquad [1]$$

Note that we drop the indicator variable y_i that appears in Formula 4.14 of Thompson (1992) since we let the sum go from 1 to n_s (where n_s is the number of sample units sampled within a stratum) and not to N_s (where N_s is the total number of possible sample units within a stratum).

Let $\hat{\hat{e}}_h$ = Estimate of the mean number of boat trips per day across all sampling units that target thresher, in stratum h, where t =the number of sampling days from the observed data. Thus the stratum h estimate, $\hat{\hat{e}}_h$, is,

$$\hat{\hat{e}}_h = N_{sh} \frac{\hat{\hat{e}}_{sh}}{t} \qquad [2]$$

where N_{sh} refers to the total number of possible sample units in stratum h.

Let $\hat{E}_{ph}$ = Estimate of the total number of boat trips for the overall time period τ that target thresher in stratum h. Then,

$$\hat{E}_{ph} = \tau \hat{\hat{e}}_h \qquad [3]$$

This expands $\hat{\hat{e}}_h$ in per day units to the total number of possible sampling days until the temporal stopping criterion[1] is reached.

Finally, let $\hat{E}_p$ estimate the total number of boat trips for the time period τ that target thresher across all strata. Therefore,

---

[1] *A temporal stopping criterion is a fixed period of time (e.g. 1 month) after which the experiment is repeated. This stopping criterion ensures that neighborhoods cannot expand forever This criterion also allows a re-design of the sampling framework and choosing prior sampling units that can vary for each period in time. This is necessary, since sample units that are expected to yield high catch/effort in a particular period of time may not remain the ideal choice for the collection of priors in a different period of time.*

$$\hat{E}_p = \Sigma_{h=1}^{H} \hat{E}_{ph} \quad . \qquad [4]$$

Note the number of strata H includes the hypothetical strata of sample of priors that we assume. In computing the estimator, it is not essential to invoke these hypothetical strata, however,while computing the variance estimators, this becomes particularly useful.

The estimate for the variance of the estimators obtained in equations [1], [2], [3], and [4] are found in equations [5],[6],[7] and [8], respectively (for detailed proofs, see Appendix B):

$$Expt\ [Var\ \ (\hat{\hat{e}}_{sh})\ ] = \left(1 - \frac{n_s}{N_s}\right) \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{(e_{ih} - \hat{\hat{e}}_{sh})^2}{n_s - 1} \qquad [5]$$

Since $\hat{\hat{e}}_h = N_s \dfrac{\hat{\hat{e}}_{sh}}{t}$, it follows that

$$Expt[Var(\hat{\hat{e}}_h)] = \frac{{N_s}^2}{t^2} \left(1 - \frac{n_s}{N_s}\right) \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{(e_{ih} - \hat{\hat{e}}_{sh})^2}{n_s - 1} \qquad [6]$$

Therefore, the estimate of the variance for the total boat trip effort across stratum h,
where $\hat{E}_{ph} = \tau \hat{\hat{e}}_h$ is

$$Expt[Var(\hat{E}_{ph})] = \tau^2 \frac{{N_s}^2}{t^2} \left(1 - \frac{n_s}{N_s}\right) \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{(e_{ih} - \hat{\hat{e}}_{sh})^2}{n_s - 1} \qquad [7]$$

where, $e_{ih} = \dfrac{f_{ih}}{m_{ih}}$.

It is assumed that the variances of effort that are obtained in one stratum are independent of the effort estimates in another stratum. This is especially obvious when strata are geographically separated. Conceivably, there may be some edge effects where effort in adjacent strata could be related. It is also assumed that sample units and the associated neighborhood expansions do not cross stratum boundaries. Since the total variance of a sum equals the sum of the individual variances it follows that the expected value of the total variance of the effort is a sum of the expected values of the variances of effort in each stratum. That is,

$$Expt[Var(\hat{E}_p)] = Expt[Var(\Sigma_{h=1}^{H} \hat{E}_{ph})] = \Sigma_{h=1}^{H} Expt[Var(\hat{E}_{ph})] \qquad [8]$$

Note that in the above value for H, a 'strata' containing a sample of the priors is assumed to separate. As already discussed, this assumption is essential, since the variance term contributed by this 'stratum' of the sample of priors includes only the variance associated with selecting a sample of priors from a set of pre-chosen priors. This selection is the only component of randomness here. For the case when all priors are chosen, there will be no variance of the estimates obtained for the spatial component, only the temporal variance will be in effect and formula [8] will be added only across H-1 strata. In this case, formula [8] is only an approximation. Hence to remain consistent, even while estimating population parameters, the collection ('stratum') of priors and their associated network has to be done separately. The above equations assume that this information is captured and is part of the built in equation.

It remains to consider, the next case where threshers were caught incidentally (non-target catch).

Boat Trip Effort: Thresher is non-target catch

As already pointed out, we consider the estimation of boat effort in two parts: Thresher targeted and thresher as a non-target. Nevertheless, the equations for the mean boat trip effort and variance for the non-target case turn out to be of the same form as target.

A thresher is considered a non-target catch if the target species of the trip(i.e., trip type) was of any other type other than the highly migratory species category. The actual data we analyzed are from CRFS for January-December for all years from 2005-2009. (See Appendix A for an illustration of the January-Dec ember 2006 dataset. All the rows are instances of a thresher catch and all the trip types not under the category of highly migratory species are considered non target.) We have no idea whether these are biased in sampling, or representative. Biases could be due to wrong readings/interpretations by the samplers, misidentification by fishermen, etc. Non target boat trip catch estimation of effort $\hat{E}_b$ is calculated for each stratum as follows:

Let $\hat{\bar{e}}_{bh_s}$ = Estimate for the mean number of boat trips across all sample days per sample unit that return with thresher as a non-target, f_{ibh} = Number of boat trips returning with thresher as a non-target across all sampling days in the i^{th} sample unit in stratum h and m_{ibh} = Number of sample units in the network containing sample unit in stratum *h*. Therefore,

$$\hat{\bar{e}}_{bh_s} = \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{f_{ibh}}{m_{ibh}} \qquad [9]$$

Let $\hat{\bar{e}}_{bh}$ be the estimate for the mean number of boat trips per day across all sampling units that return with thresher as a non-target. This implies

$$\hat{\bar{e}}_{bh} = N_s \frac{\hat{\bar{e}}_{bh_s}}{t} \qquad [10]$$

Let $\hat{E}_{bh}$ = Estimate for the total number of boat trips for the time period τ that had non target catch of thresher in stratum h . Thus,

$$\hat{E}_{bh} = \tau \hat{\bar{e}}_{bh} . \qquad [11]$$

Finally we compute $\hat{E}_b$, the estimate of the total number of boat trips for the time period τ that had non target catch of thresher across all strata as follows:

$$\hat{E}_b = \Sigma_{h=1}^{H} \hat{E}_{bh} \qquad [12]$$

Note the number of strata H includes the hypothetical strata of sample of priors that we assume. In computing the estimator, it is not essential to invoke this hypothetical strata but while computing the variance estimators, this becomes particularly useful.

The estimate for the variance of the estimators obtained in equations [9], [10], [11], and [12] are found in equations [13],[14],[15] and [16], respectively (for detailed proofs, see Appendix B):

$$Expt\,[Var\ (\hat{\hat{e}}_{bh_s})] = \left(1-\frac{n_s}{N_s}\right)\frac{1}{n_s}\Sigma_{i=1}^{n_s}\frac{(e_{ibh}-\hat{\hat{e}}_{bh_s})^2}{n_s-1} \quad [13]$$

Since $\hat{\hat{e}}_{bh} = N_s\frac{\hat{\hat{e}}_{bh_s}}{t}$, it follows that

$$Expt[Var(\hat{\hat{e}}_{bh})] = \frac{{N_s}^2}{t^2}\left(1-\frac{n_s}{N_s}\right)\frac{1}{n_s}\Sigma_{i=1}^{n_s}\frac{(e_{ibh}-\hat{\hat{e}}_{bh_s})^2}{n_s-1} \quad [14]$$

Therefore, the estimate of the variance for the total boat trip effort across stratum h,
where $\hat{E}_{bh} = \tau\hat{\hat{e}}_{bh}$ is

$$Expt[Var(\hat{E}_{bh})] = \tau^2\frac{{N_s}^2}{t^2}\left(1-\frac{n_s}{N_s}\right)\frac{1}{n_s}\Sigma_{i=1}^{n_s}\frac{(e_{ibh}-\hat{\hat{e}}_{bh_s})^2}{n_s-1} \quad [15]$$

where, $e_{ibh} = \frac{f_{ibh}}{m_{ibh}}$.

Again like the previous section, we assume variances of effort that are obtained in one stratum are independent of the effort estimates in another stratum. Therefore an approximation for the variance of the estimators across all strata,

$$Expt[Var(\hat{E}_b)] = Expt[Var(\Sigma_{h=1}^{H}\hat{E}_{bh})] = \Sigma_{h=1}^{H}Expt[Var(\hat{E}_{bh})] \quad [16]$$

Note that the non-target catch effort depends directly on the non-target as opposed to effort associated with target catch which is independent of the catch. Therefore, it is expected that the non target catch effort has many zeroes and that neighborhoods are not easily formed. The neighborhood could overlap with the sample unit(s) associated with the targeted effort, and vice-versa. Since the sampling form addresses each of the issues separately, targeted thresher and non target thresher catch can be considered separately with the data on the one form.

Note that in the questionnaire, the question of whether an angler knows if a thresher was released dead or alive, in case it was released, is addressed. In the case of non target thresher catch, this becomes particularly important, since one could expect a large portion of anglers to release their catch instead of keeping it, if thresher is not a target. So, if the answer is 'don't know', it should be included as a positive 'hit' in estimating the effort. This would also reduce the zero hits and is a more precautionary approach.

Angler Effort Estimation

Mean angler effort is a better measure of effort than just the boat trips, since for any one boat trip the number of anglers per boat and hence the number of angler days is likely not equal to one. In the following, subscripts associated with mean angler effort are not shown. All of the following expressions are per stratum and the total effort across all strata is obtained using an expression similar to Equation [4].

Let $\hat{A}$ estimate the total angler effort per day be given by

$$\hat{A} = \hat{a}\hat{\bar{e}} \qquad [17]$$

where $\hat{a}$ is the estimate for the mean number of angler days per boat and $\hat{\bar{e}}$ is the estimate for the mean number of boat trips per day across all sampling units that target/non target thresher.

Therefore, the total angler effort for thresher (target or by-non target catch) (with appropriate subscripts) is:

$$\hat{A}_T = \tau \hat{a}\hat{\bar{e}} \ . \qquad [18]$$

It is necessary to derive mean and variance estimators for the random variable $\hat{a}$, as well as $\hat{A}$. The expected value and variance of $\hat{A}_T$ are given respectively below

$$Expt(\hat{A}_T)$$

$$Var(\hat{A}_T) = \tau^2 Expt(Var(\hat{a}\hat{\bar{e}}))$$

where τ = total number of possible sample days (universe of days, e.g. 30)

In the previous survey [Ref [3],Page 9] it was assumed that $\hat{a}$ and $\hat{E}$ are independent and Random Variables (RV). Though not explicitly stated, the expression used indicates this assumption. In other words, an increase in the mean number of angler days per boat (directly proportional to the mean number of anglers per boat) could be accompanied by an increase in the number of boat trips. This can be explained by the following: a boat can hold only a certain number of anglers; on a day when the thresher bite frequency is high, a larger number of anglers may be expected and, as a consequence, the number of boat trips increases. This dependence between the estimate for the mean number of anglers per boat and the estimate for the number of boat trips per day kicks in only after a threshold for the expected number of anglers per boat is reached. It is assumed that if every boat trip targeting threshers holds n anglers, then that boat trip goes out with n anglers. The threshold is formed in this way. It follows that as the mean number of anglers per boat increases; it implies more (and bigger) boat trips are occurring. That is a dependency. By not assuming independence the computation becomes more complicated. Thus, unlike the preceding report [Ref. 3], this report is conservative and assumes dependency between the estimate for the mean number of anglers per boat and the boat trip estimate per day.

Also note, we do not assume that as the number of boat trips increases, the number of anglers per boat increases. So the covariance term applies only in one direction.

The covariance between two random variables X and Y is given by:

$$Cov[X,Y] = Expt[XY] - Expt[X]Expt[Y] \qquad [19]$$

$$Expt[XY] = Cov[X,Y] + Expt[X]Expt[Y] \qquad [20]$$

However, since only sample data will be available, a sampling covariance is used. Further, the sampling covariance must be one directional between the mean number of angler days per boat and the number of boat trips per day. For example,

$$Cov[\hat{\hat{e}}, \hat{a}] = \frac{1}{n-1} \Sigma_{i=1}^{n} (e_i - \hat{\hat{e}})(a_i - \hat{a}) \qquad [21]$$

To repeat, it is assumed that the mean number of angler days per boat does depend on the boat trip effort, but not vice versa. We develop explicit relationships for $\hat{\hat{e}}$ and $\hat{a}$ in the following section, so that $\hat{a}$ is the independent variable and $\hat{\hat{e}}$ is the dependent variable. In Appendix F, we also introduce another possible interpretation for $\hat{a}$ that is independent of $\hat{\hat{e}}$, so that we obtain similar estimators as obtained in the CRFS 2009 [Ref. 3] survey report.

Angler Effort Estimation: The preferred interpretation

The interpretation of mean number of angler days assumes that the angler effort on the j^{th} sample day is:

$$a_j = \Sigma_{i=1}^{e_j} \frac{(NAng)_i (hfs)_i}{24} \qquad j = 1, \ldots t \qquad [22]$$

where $NAng_i$ is the number of anglers on boat i, and hfs_i is the number of hours fished on boat i,and e_j is the number of sampled boats returning on the j^{th} sample day with a target or non-target catch of thresher. (N.B. $\frac{1}{24}$ converts number of hours to number of days). It is important to point out that e_j varies from day to day, both in terms of the number of boats sampled (what is physically observed) and the number of sample units (e.g. marinas) where sampling occurs. Note that the number of sample units observed varies from day to day based on the neighborhood expansion. However, we are trying to estimate the mean number of angler days per boat, $\hat{a}$. The fraction $\frac{a_j}{e_j}$(see [22]) is an unbiased estimator of $\hat{a}$, the mean number of angler days per boat. This is true for any arbitrary number of observed sample units $n_s < N_s$, on sample day j. Note that the number of sample units to be sampled on day j is specifically the "samplemeister's" observation on the sample day before, day (j-1) of the boat trip effort.

This implies that the mean number of angler days per boat is:

$$\hat{a} = \frac{1}{t} \Sigma_{j=1}^{t} \frac{a_j}{e_j} \qquad [23]$$

Again, this sample estimate of the mean number of angler days per boat is an unbiased estimator of the mean number of angler days per boat for the population. It must be noted that [23] is an atypical ratio estimator used in sampling studies, compared to the typical ratio estimator indicated in Appendix F.

The estimate for variance of the estimate of the number of angler days per boat is a similar expression to the mean boat trip effort expression in Equation [13]. That is,

$$Expt(Var(\hat{a})) = \left(1-\frac{t}{\tau}\right)\frac{1}{t}\Sigma_{j=1}^{t}\frac{((\frac{a_j}{e_j})-\hat{a})^2}{t-1} \qquad [24]$$

Return now to the covariance expression [20]. Take the expectation on both sides of [17] and substitute in [20]. This yields:

$$Expt[\hat{A}] = Expt[\hat{a}\hat{\hat{e}}] = \frac{1}{t-1}\Sigma_{i=1}^{t}(e_i - \hat{\hat{e}})(\frac{a_i}{e_i}-\hat{a}) + \hat{\hat{e}}\hat{a} \qquad [25]$$

where $\hat{\hat{e}}$ is the estimate for the number of boat trips per day as obtained in Equations [2] and [10].

Note that we have dropped the subscript h for convenience. It must nevertheless be noted that all of the above estimators are per stratum and the total angler effort is obtained by taking the sum of estimates of the angler effort per stratum h.

Since the angler effort estimator is the product of two estimators, $\hat{a}$ and $\hat{\hat{e}}$, and since it was explained above why it is unreasonable to consider them independent random variables, the variance of the product (Goodman, Equation [21]) should include consideration of the covariance relationship between $\hat{\hat{e}}$ and $\hat{a}$. This is different from the obtained expression for the variance in the CRFS Methods -(Page 7,Ref. [3]), where independence between the estimate of the angler days per boat and the estimate for the mean boat trip estimate per day were assumed and thus no covariance term was involved.

$$\begin{aligned}
&tVar(\hat{A}) = tVar(\hat{a}\hat{\hat{e}}) = \\
&\hat{a}^2Var(\hat{\hat{e}}) + \hat{\hat{e}}^2Var(\hat{a}) + 2\hat{\hat{e}}\hat{a}Cov[\hat{a},\hat{\hat{e}}] + \\
&2\frac{\hat{a}}{t(t-1)}\Sigma_{i=1}^{t}[(\frac{a_i}{e_i}-\hat{a})(e_i-\hat{\hat{e}})^2] + 2\frac{\hat{\hat{e}}}{t(t-1)}\Sigma_{i=1}^{t}[(e_i-\hat{\hat{e}})(\frac{a_i}{e_i}-\hat{a})^2] + \\
&\frac{Var(\hat{a})Var(\hat{\hat{e}})}{t} + \frac{1}{t^2}\frac{1}{t-1}\Sigma_{i=1}^{t}[(e_i-\hat{\hat{e}})^2(\frac{a_i}{e_i}-\hat{a})^2] - \frac{Var(\hat{a})Var(\hat{\hat{e}})}{t^2} - \frac{1}{t^2}[Cov(\hat{a},\hat{\hat{e}})]^2 \qquad [26]
\end{aligned}$$

As in earlier expressions, there are two cases: thresher targeted and thresher as a non-target, which are both based on the generalized case of [25,26].

Angler Effort Estimation: Thresher is targeted

The estimate for the angler effort per day for targeted thresher in stratum h is

$$\hat{A}_h = \hat{a}_h\hat{\hat{e}}_h = \frac{1}{t-1}\Sigma_{i=1}^{t}(e_{ih} - \hat{\hat{e}}_h)(\frac{a_{ih}}{e_{ih}}-\hat{a}_h) + \hat{\hat{e}}_h\hat{a}_h \qquad [27]$$

where

$\hat{A}_h$ - Angler effort per day on boats that target thresher in stratum h
$\hat{a}_h$ - Mean angler effort per boat for boats targeting thresher in stratum h (Equation 24)
$\hat{\hat{e}}_h$ - Mean number of boats returning with thresher as target in stratum h (Equation 2)
e_{ih} - Number of sampled boats observed with thresher as target on sample day i.
t - Number of days sampled

Thus if $\hat{A}_{hT}$ is the estimate for the total angler effort in stratum h on boats that target thresher, then it is given by the expression

$$\hat{A}_{hT} = \tau \hat{A}_h \quad [28]$$

Thus $\hat{A}_T$ which is the total angler effort across all possible sample days across all strata H on boats that target threshers is given by

$$\hat{A}_T = \Sigma_{h=1}^{H} \hat{A}_{hT} \quad [29]$$

The variance of these estimators are given by

$$\begin{aligned} tVar(\hat{A}_h) &= tVar(\hat{a}_h \hat{\hat{e}}_h) = \\ &\hat{a}_h^2 Var(\hat{\hat{e}}_h) + \hat{\hat{e}}_h^2 Var(\hat{a}_h) + 2\hat{\hat{e}}_h \hat{a}_h Cov[\hat{a}_h, \hat{\hat{e}}_h] + \\ &2\frac{\hat{a}_h}{t(t-1)} \Sigma_{i=1}^{t} [(\frac{a_{ih}}{e_{ih}} - \hat{a}_h)(e_{ih} - \hat{\hat{e}}_h)^2] + 2\frac{\hat{\hat{e}}_h}{t(t-1)} \Sigma_{i=1}^{t} [(e_{ih} - \hat{\hat{e}}_h)(\frac{a_{ih}}{e_{ih}} - \hat{a}_h)^2] - \\ &\frac{Var(\hat{a}_h)Var(\hat{\hat{e}}_h)}{t} + \frac{1}{t^2}\frac{1}{t-1} \Sigma_{i=1}^{t} [(e_{ih} - \hat{\hat{e}}_h)^2 (\frac{a_{ih}}{e_{ih}} - \hat{a}_h)^2] - \frac{Var(\hat{a}_h)Var(\hat{\hat{e}}_h)}{t^2} - \frac{1}{t^2}[Cov(\hat{a}_h, \hat{\hat{e}}_h)]^2 \end{aligned} \quad [30]$$

And thus

$$Var(\hat{A}_{hT}) = \tau^2 Var(\hat{A}_h) \quad [31]$$

$$Var(\hat{A}_T) = Var(\Sigma_{h=1}^{H} \hat{A}_{hT}) = \Sigma_{i=1}^{H} Var(\hat{A}_{hT})$$

Angler Effort Estimation: Thresher is non-target catch

The estimate for the angler effort per day for non-target thresher in stratum h

$$\hat{A}_{bh} = \hat{a}_{bh} \hat{\hat{e}} b_{bh} = \frac{1}{t-1} \Sigma_{i=1}^{t} (e_{ibh} - \hat{\hat{e}}_{bh})(\frac{a_{ibh}}{e_{ibh}} - \hat{a}_{bh}) - \hat{\hat{e}}_{bh} \hat{a}_{bh} \quad [32]$$

where

$\hat{A}_{bh}$ - Angler effort per day in boats that have an incidental thresher catch in stratum h

$\hat{a}_{bh}$ - Mean angler effort per boat for boats that have thresher as non-target catch in stratum h (Equation 24)
$\hat{\hat{e}}_{bh}$ - Mean number of boats returning with thresher as non-target catch in stratum h (Equation 2)
e_{ibh} - Number of sampled boats observed with thresher as non-target catches on sample day i.
t - Number of days sampled

Thus, if $\hat{A}_{bhT}$ is the estimate for the total angler effort that do not target thresher in stratum h, then it is given by the expression

$$\hat{A}_{bhT} = \tau \hat{A}_{bh} \quad [33]$$

Thus $\hat{A}_{bT}$ which is the total angler effort across all possible sample days across all strata H that non target catch thresher is given by

$$\hat{A}_{bT} = \Sigma_{h=1}^{H} \hat{A}_{bhT} \quad [34]$$

The variance of these estimators are given by

$$tVar(\hat{A}_{bh}) = tVar(\hat{a}_{bh}\hat{\hat{e}}_{bh}) =$$
$$\hat{a}_{bh}^2 Var(\hat{\hat{e}}_{bh}) + \hat{\hat{e}}_{bh}^2 Var(\hat{a}_{bh}) + 2\hat{\hat{e}}_{bh}\hat{a}_{bh} Cov[\hat{a}_{bh}, \hat{\hat{e}}_{bh}] +$$
$$2\frac{\hat{a}_{bh}}{t(t-1)}\Sigma_{i=1}^{t}[(\frac{a_{ibh}}{e_{ibh}} - \hat{a}_{bh})(e_{ibh} - \hat{\hat{e}}_{bh})^2] + 2\frac{\hat{\hat{e}}_{bh}}{t(t-1)}\Sigma_{i=1}^{t}[(e_{ibh} - \hat{\hat{e}}_{bh})(\frac{a_{ibh}}{e_{ibh}} - \hat{a}_{bh})^2] +$$
$$\frac{Var(\hat{a}_{bh})Var(\hat{\hat{e}}_{bh})}{t} + \frac{1}{t^2}\frac{1}{t-1}\Sigma_{i=1}^{t}[(e_{ibh} - \hat{\hat{e}}_{bh})^2(\frac{a_{ibh}}{e_{ibh}} - \hat{a}_{bh})^2] - \frac{Var(\hat{a}_{bh})Var(\hat{\hat{e}}_{bh})}{t^2} - \frac{1}{t^2}[Cov(\hat{a}_{bh}, \hat{\hat{e}}_{bh})]^2 \quad [35]$$

And thus

$$Var(\hat{A}_{bhT}) = \tau^2 Var(\hat{A}_{bh}) \quad [36]$$
$$Var(\hat{A}_{bT}) = Var(\Sigma_{h=1}^{H}\hat{A}_{bhT}) = \Sigma_{i=1}^{H} Var(\hat{A}_{bhT}) \quad [37]$$

Estimation of Thresher Catch

Estimates for total catch of the primary sites are calculated using the total number of boats sampled during the month, for each day-type (weekday or weekend). Using the revised sampling form, two separate catch estimates (thresher as a target and thresher as a non-target catch) are proposed. This recommendation is a result of analyses of the past five years of data where there is significant thresher non target thresher catch in different trip types, in contrast to catch where thresher was targeted. The total catch estimate is a sum of trips where threshers are a target and occur as a non-target catch. To estimate catch, data will be obtained from the newly designed Thresher Shark Catch and Effort Sampling form found in Appendix D. For example, an answer of 'NO' to the question: *Were you fishing for thresher* *sharks today and a 'YES' to the question "did you catch any threshers?"* represents a sample for non target thresher catch.

Catch Estimation: Thresher is targeted and Thresher as a non-target

In the preceding sections, estimates for effort were based on an adaptive sampling design where data were collected from the re-designed sampling form (Appendix D). Catch is also estimated with the same adaptive design from data from the re-designed survey form, used for estimating effort.

This section requires the interface of the estimating equations below, and the sampling design. In this section, estimators for vessels that catch a thresher are given. Similar expressions can be obtained for vessels with non-target by replacing the subscript p with subscript b.

Catch estimates for each stratum are calculated as follows:

$$\hat{\bar{c}}_s = \frac{1}{n_s} \Sigma_{i=1}^{n_s} \frac{r_i}{q_i} \qquad [38]$$

Where

$\hat{\bar{c}}_s$ - Estimate for the mean catch per sample unit returning with thresher as target across all sampling days.

n_s - Number of Sampled units in stratum h

r_i - Number of threshers caught, as target, across all sampling days in the network containing sample units i.

q_i - Number of sample units in the network containing sample unit i.

Let $\hat{\bar{c}}$ = Estimate for the mean catch per day across all sampling units that target thresher. Note that this estimate is an unbiased estimator of the catch per sample unit for the whole population of thresher catch, which is given by:

$$\hat{\bar{c}} = N_s \frac{\hat{\bar{c}}_s}{t} \qquad [39]$$

Where

N_S - Number of possible sample units in stratum h

t - Number of Sampled days

Finally, let $\hat{C}_p$ estimate the total catch per stratum in the time period τ (the number of possible sample days, e.g.30) on trips that target thresher. This is given by:

$$\hat{C}_p = \tau \hat{\bar{c}} \qquad [40]$$

Adding all the per stratum estimates over all strata gives the total catch estimate over the study area across the California coast is easy to see that the adaptive design feature**.** Recall that the subscripts that refer to random sampling from the universe of the prior sample units and from the universe of non-priors must be carried out independently. Of course, once the sampling has been carried out the population parameters are estimated independently for sampling amongst priors and sampling amongst non-priors. These two total estimates are then added together with their variances.

The estimate of variance of the total catch estimate obtained above, per stratum, is then obtained as: (the derivation is very similar to total boat trip effort estimation [7])

$$Expt(Var(\hat{C}_p)) = \tau^2 \frac{N_s^{\ 2}}{t^2}\left(1-\frac{n_s}{N_s}\right)\frac{1}{n_s}\Sigma_{i=1}^{n_s}\frac{(c_i-\hat{\hat{c}}_s)^2}{n_s-1} \qquad [41]$$

The variance estimate of the total catch estimate is then obtained as the sum of individual per stratum estimates of variance of catch estimates, since we assume that the estimates obtained across every stratum are mutually independent.

CPUE (Catch per Unit Effort)

The estimate for CPUE for thresher as target is as follows:

$$\hat{T}_{cpue_p} = \frac{\hat{\hat{c}}}{\hat{A}} \qquad [42]$$

and the CPUE for thresher as a non target catch is:

$$\hat{T}_{cpue_b} = \frac{\hat{\hat{c}}(bycatch)}{\hat{A}_b} \qquad [43]$$

Where $\hat{\hat{c}}$ is given in [39] and $\hat{A}$ is given in [28]. The per day estimators are used in [42] and [43] to compute the CPUE without amplification. As shown, the effort estimate is a product of two dependent random variables for the preferred interpretation. The estimate for CPUE in equations [42] or [43] is the simple numerical quotient.

To estimate the expected value of the variance of the estimate of the CPUE, it is noted that there does not exist a closed form equation to evaluate the exact expression for the variance of a quotient of two random variables that are not independent of each other. Cochran (1977) provides an expression that is generally accepted for use with Simple Random Sampling data. Of course, the CPUE from [42] and [43] appear in this equation which is recast in terms of this report as Equation [44]. This equation is an approximation that assumes the independence between the estimate of angler days per boat and estimate for number of boat trips which target/non target catch threshers, but without the covariance terms in Equation [30], for the variance of the angler effort estimate. In principle, an equation for the expected value of the variance, adapted for cluster sampling exists with correction terms (Thompson, 1992, Chapter 12: Pg 113-116) but these corrections are ignored here. Based on this assumption, we approximate the expected variance of the CPUE estimate as the SRS Ratio Estimator:

$$Expt(Var(T)_{cpue_p}) = K \frac{1-\frac{n_s}{N_s}}{n_s(\hat{A})^2} \Sigma_{i=1}^{n_s} \frac{(C_i - \hat{T}_{cpue_b} A_i)^2}{n_s - 1} \quad [44]$$

constant K is the appropriate amplifying constant.

A similar expression for the non-target CPUE Variance is obtained with subscript b.
It is worth noting that the estimate of the CPUE per day is a ratio of two estimators (random variables), the numerator representing the catch per day amplified across N sample units and the denominator representing the effort as the amplified angler effort per day. Thus, the sampling design used is not explicitly represented. Further, on one hand, in a sport fishery one might argue that these random variables are independent. On the other hand, the premise of adaptive sampling, as used here, is that catch detected in sampling may lead to additional boats going fishing. However, obtaining covariance estimators would be more than what is possible with available data.

Another related consideration, is that, in this formulation, the effort random variable is the product of two other random variables where independence is not assumed in [30].

Sampler Cost

The dollar cost of sampling in this type of design is more complicated to estimate than it is for the more common sampling methods. This occurs because adaptive sampling may add additional sampling effort depending on the outcome of the immediately preceding days. Thus, the “samplemeister” can only estimate the number of samples to be taken in the first stage. After that, additional samples will be taken based on the results from the first stage by building the neighborhoods around the hits in the initial sample.

Nevertheless, some of the traditional design questions also arise: for example, the number of days to sample and in which months. Based on the CRFS data, most of the activity is known to occur during the months of April to October. As noted, this design includes a criterion involving prior sample units called primary sites. The “samplemeister” will choose the primary sites and randomly select other sample units, and how many, in each case. These decisions will be based upon both the specification of priors from the last five years with the “old” design (Appendix C) and the number of randomly chosen “new” sets. Further, the “samplemeister” also must specify the numbers of days before neighborhoods are no longer sampled. In other words, the “samplemeister” will need to determine how to distribute sampling effort spatially and temporally. Ultimately, the number chosen will depend upon the total budget and the salary of the samplers.

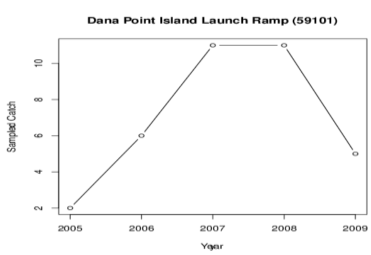
Remember, when "hits" occur and neighborhoods are built, samplers do not revisit but instead move to the neighborhood sample units. The “samplemeister” will need to construct a stopping rule for the number of successive days that sampling occurs in a neighborhood (again primarily a question of cost).

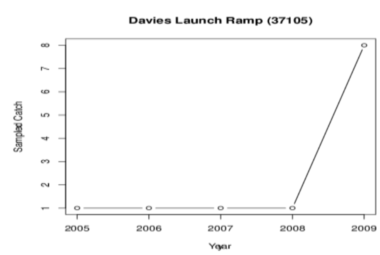
Analysis of Extant Data

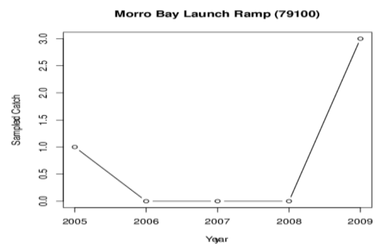
In this section, information is provided on the sample data available from 2005-2009 at hand that could potentially aid in selecting possible sites to sample and to expand neighborhoods. This data was obtained from the CRFS database *['Download New CRFS Estimates' Ref. [5]]*. These data were examined in search of trends in the total sampled catch for all thresher cases (caught and kept, released alive and released dead) from 2005-2009.

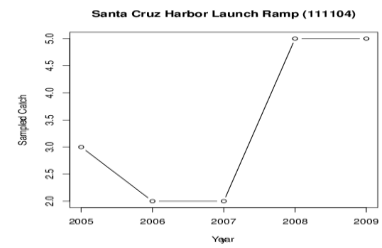
Figure 3 is a set of sample graphs for PR1 sites representing the data at hand. Clearly, the catch levels do not follow any trends that could be fitted to a mathematical equation for any of these sites. At best, one can predict a general trend for some sites, such as Dana Basin (generally increasing) and Marina Del Rey (generally decreasing). However, the lack of sufficient historical data does not warrant relying on any trend based prediction of catch behavior that can be modeled mathematically. Nevertheless, it is potentially useful to see how catch has changed year to year for selected prior landing sites (Fig.3). One might look for constancy, trends etc. In addition, as was noted, different sample units (e.g., marinas, landing ramps), while equally likely to be selected could be weighted differently. These representative catch time series could be assigned weights based on the number of slips per marina or some other criterion. This section also tends to justify the use of a clustered adaptive sampling design because atleast in southern California, some of these sample units can be thought of as forming clusters. For a mathematical criterion regarding the efficiency of Adaptive Sampling vs. Simple Random Sampling, refer to Appendix F.

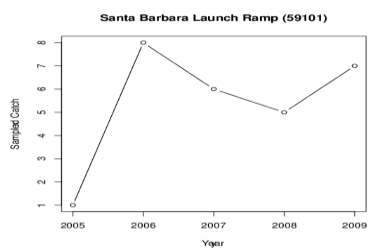
Fig 3: Thresher Shark Catch Data Analysis, showing the catch from sampling over 2005-2009, for 10 of the priors

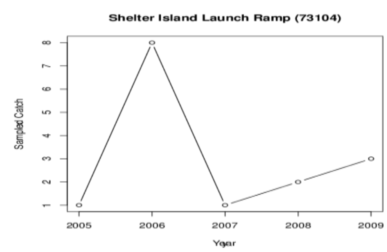


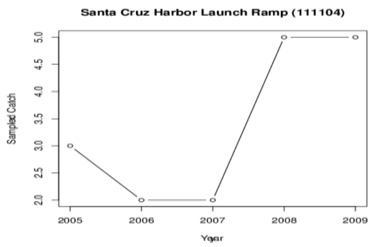


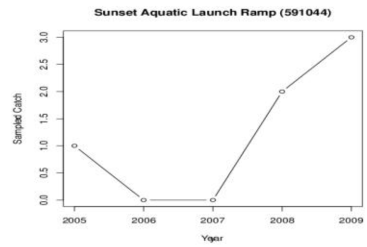

Morro Bay Launch Ramp (79100)
Sampled Catch
Year

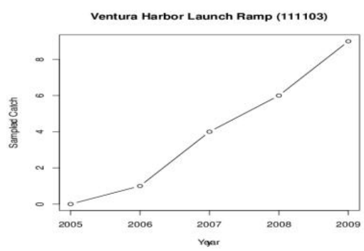

Santa Cruz Harbor Launch Ramp (111104)
Sampled Catch
Year

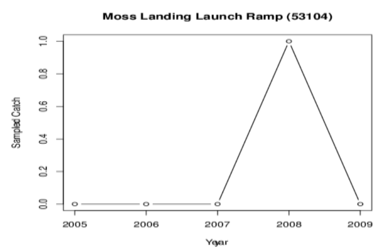

Santa Barbara Launch Ramp (59101)
Sampled Catch
Year


Shelter Island Launch Ramp (73104)
Sampled Catch
Year









Selected Assumptions of the Sample Design

- It is assumed that all sample units are unweighted. That is, they are of size 1. This design is flexible to take into account both weighted and un-weighted cases.
- Some 'Thresher boats' probably return to sites during hours when samplers are not present. In this case, an additional survey is needed, e.g. a telephone or mail survey. This is not considered in this report
- Sites to the north of the San Francisco Bay area could be discarded and not be a part of n_s and N_s .
- Since this adaptive design is extended spatially over most of the California coast and temporally over several days each month, and since sample unit selection is not straightforward simple random sampling, it maybe useful to introduce some statistical training for would be 'samplemeisters'.

7 Bibliography

[1] W.G. Cochran. *Sampling Techniques*. Wiley; 3rd edition (January 1, 1977).

[2] L.A. Goodman. *On the exact variance of products*. *Journal of the American Statistical Association*, 55(292):708-713, 1960

[3] California Department of Fish and Games. *"California Recreational Fisheries Survey Methods, 2009"*.

[4] S.K. Thompson, S*ampling*. Wiley, New York, 1992

[5] CRFS Database, Download New 2 CRFS Estimates, (*http://www.recfin.org/forms/dcrfs3.html*)

[6] S.K. Thompson, G.A.F. Seber, and OG Okogbaa. *Adaptive sampling*. Wiley New York, 1996.

[7] Pollard JH, Palka D, Buckland ST *Adaptive line transect sampling*. *Biometrics*. 2002 Dec;58(4):862-70

[8] Smith, D.R., Brown, J.A., Lo, N.C.H. *Application of adaptive sampling to biological populations. In* Sampling rare or elusive species: concepts, designs, and techniques for estimating population Parameters. Island Press, 2004

Appendix A : Example Illustration of Non Target effort and catch from the Jan 2005- Dec 2005 CRFS Dataset

Table A1: Illustrates the frequency of thresher non target catch when thresher is not targeted. Highly migratory means thresher targeted. All other categories refer to thresher not targeted. Refer to Table B2 in the California Recreational Fisheries Survey 2007 (*http://www.dfg.ca.gov/marine/pdfs/crfs2007review.pdf*) for examples of species falling under various categories. Note that in 2005, 12 threshers were targeted catch and 15 threshers were non target catch. Wd and We stand for week day and weekend day, respectively.

cntysite	kod	Triptype	COMMON	month	year
37010	Wd	Inshore	thresher shark	2	2005
73113	We	Inshore	thresher shark	3	2005
73113	We	Highly migratory	thresher shark	4	2005
73113	We	Inshore	thresher shark	4	2005
37010	Wd	Inshore	thresher shark	5	2005
37010	We	Highly migratory	thresher shark	5	2005
73204	We	Highly migratory	thresher shark	5	2005
37010	We	Bottomfish	thresher shark	6	2005
37010	We	Highly migratory	thresher shark	6	2005
59101	We	Highly migratory	thresher shark	6	2005
73204	We	Highly migratory	thresher shark	6	2005
37010	We	Bottomfish	thresher shark	7	2005
37010	We	Highly migratory	thresher shark	7	2005
37010	We	Inshore	thresher shark	7	2005
37010	Wd	Highly migratory	thresher shark	8	2005
37010	We	Bottomfish	thresher shark	8	2005
37010	We	Highly migratory	thresher shark	8	2005
73104	Wd	Inshore	thresher shark	8	2005
37010	Wd	Inshore	thresher shark	9	2005
37010	We	Highly migratory	thresher shark	9	2005
37010	We	Inshore	thresher shark	9	2005
37010	We	Anything	thresher shark	10	2005
37010	We	Inshore	thresher shark	10	2005
59101	Wd	Highly migratory	thresher shark	10	2005
37010	We	Highly migratory	thresher shark	11	2005
37105	We	Anything	thresher shark	11	2005
59104	We	Inshore	thresher shark	12	2005

Appendix B : Variance of estimate for the mean number of boat trips per sample unit returning with thresher as target across all sampling days per stratum

The proof is along the lines of Thm 2.2 and Thm 2.4 in Cochran (1977).

Let $\overline{E_s}$ - Population (True) Mean for the number of boats that target thresher = $\frac{1}{N_s}\Sigma_{i=1}^{N_s}\frac{f_i}{m_i}$. This cannot be found directly. Hence we use the sample mean $\hat{\bar{e}}_s$.(we drop the subscript h for convenience since this estimate is per stratum). It can also be shown that $Expt[\hat{\bar{e}}_s] = \overline{E_s}$ which makes this an unbiased estimator. If $(\hat{\bar{e}}_s - \overline{E}_s)$ is the error between the estimate of the mean number of boat trips per sample units across all days and the actual mean number of boat trips per sample unit across all days, then $n_s \times$ (the error estimate) is:

$$n_s(\hat{\bar{e}}_s - \overline{E_s}) = (e_1 - \overline{E_s}) + (e_2 - \overline{E_s}) + (e_{n_s} - \overline{E_s}).$$

$$\Rightarrow n_s{}^2 Expt[(\hat{\bar{e}}_s - \overline{E_s})^2] = Expt[(e_1 - \overline{E_s})^2 + (e_2 - \overline{E_s})^2 + (e_n - \overline{E_s})^2] +$$
$$2\big((e_1 - \overline{E_s})(e_2 - \overline{E_s}) + (e_1 - \overline{E_s})(e_3 - \overline{E_s}) + n_s(n_s - 1)/2\text{terms}\big)]$$

Since $Expt[\hat{\bar{e}}] = \overline{E},$ it follows that

$$\frac{1}{n_s} Expt(\Sigma_{i=1}^{n_s} e_i) = \frac{1}{N_s}(\Sigma_{i=1}^{N_s} e_i)$$

Therefore,

$$Expt[e_1 + e_2 .....e_{n_s}] = \frac{n_s}{N_s}[e_1 + e_2 + ...e_{N_s}]$$

Subtracting $n_s\overline{E_s}$ from both sides, squaring both sides and grouping common terms, we see that

$$Expt[(e_1 - \overline{E_s})^2 + (e_2 - \overline{E_s})^2 + ...(e_{n_s} - \overline{E_s})^2] = \frac{n_s}{N_s}[(e_1 - \overline{E_s})^2 + (e_2 - \overline{E_s})^2 + ...(e_{N_s} - \overline{E_s})^2]$$

and

$$Expt[\big((e_1 - \overline{E_s})(e_2 - \overline{E_s}) + (e_1 - \overline{E_s})(e_3 - \overline{E_s}) + ...n_s(n_s - 1)/2\text{terms}\big)] =$$
$$\frac{n_s(n_s - 1)}{N_s(N_s - 1)}[\big((e_1 - \overline{E_s})(e_2 - \overline{E_s}) + (e_1 - \overline{E_s})(e_3 - \overline{E_s}) + N_s(N_s - 1)/2\text{terms}\big)]$$

Therefore

$$n_s^2 Expt[(\hat{\bar{e}}_s - \overline{E_s})^2] = \frac{n_s}{N_s}\Big([(e_1 - \overline{E_s})^2 + (e_2 - \overline{E_s})^2 + ...(e_{N_s} - \overline{E_s})^2] +$$

$$2(\frac{n_s - 1}{N_s - 1})[\Big((e_1 - \overline{E_s})(e_2 - \overline{E_s}) + (e_1 - \overline{E_s})(e_3 - \overline{E_s}) + ...N_s(N_s - 1)/2\text{terms}\Big)]\Big) =$$

$$\frac{n_s}{N_s}\Big((1 - \frac{n_s - 1}{N_s - 1})([(e_1 - \overline{E_s})^2 + (e_2 - \overline{E_s})^2 + ...(e_{N_s} - \overline{E_s})^2] +$$

$$\frac{n_s - 1}{N_s - 1}\Big((e_1 - \overline{E_s}) + (e_2 - \overline{E_s}) + ...(e_{N_s} - \overline{E_s})\Big)^2\Big)$$

Note that the second term in the above sum is exactly equal to 0.

Since $Var(\hat{\bar{e}}_s) = Expt[(\hat{\bar{e}}_s - \overline{E_s})^2]$, by bringing the $\frac{1}{{n_s}^2}$ to the other side from the above derivation, we get

$$Var(\hat{\bar{e}}_s) = \frac{N_s - n_s}{n_s N_s (N_s - 1)} \Sigma_{i=1}^{N_s}(e_i - \overline{E_s})^2 = \Big(1 - \frac{n_s}{N_s}\Big)\frac{1}{n_s}\Sigma_{i=1}^{N_s}\frac{(e_i - \overline{E_s})^2}{N_s - 1}$$

An unbiased estimator for

$$\Sigma_{i=1}^{N_s}\frac{(e_i - \overline{E_s})^2}{N_s - 1}$$

is

$$\Sigma_{i=1}^{n_s}\frac{(e_i - \hat{\bar{e}}_s)^2}{n_s - 1}$$

(For proof, refer Thm. 2.4 in Cochran.)

Appendix C: Sample 'Prior' Sites

The 28 sample sites that were identified from the 2005-2009 data are shown in Table A2. These occur in eight strata stretching from the U.S.-Mexico border to north of San Francisco. In each stratum, there is one or more 'prior' sample sites. Each site is a sample unit. See Fig. 2 for location of sites and Table 1 for strata location.

Table A2: The sample 'prior' sites and their associated strata

Stratum Number	Prior Site
1	Mission Bay West
1	Shelter Island launch Ramp
1	Imperial Beach pier
2	Marina Del Ray Sport Fishing
2	Dana Point launching ramp
2	Oceanside Launch ramp
2	Sunset Aquatic Launch Ramp
2	Cabrillo Beach launch Ramp
2	Huntington Beach (MM)
2	Hermosa Beach pier
3	Venice Pier
3	Santa Monica pier
4	Santa Barbara Harbor
4	PR Ventura Marina
4	Marine Stadium east launch ramp
4	PR Channel Islands
4	Goleta pier - Ward M
4	Malibu Beach pier

4	BB San Buenaventur
5	Avila
5	Morro Bay
5	Gaviota State Beach
6	Santa Cruz harbor
6	Moss Landing
6	Princeton
7	Berkeley Marina
7	Richmond
8	Bodega West

Appendix D: Sampler’s form

CRFS THRESHER SHARK ADAPTIVE DESIGN SURVEY FORM

(Note: Completion time should be less than 10 mins)

Sampler ID:
Date:
Time:
Site Code:
Boat Number:
Whether allowed to visually inspect?

1. Were you fishing for thresher sharks today?

2. Were you fishing for any other HMS Shark species today?

3. How many thresher sharks did you catch today?
   3(a) Number Kept? 2(c): Released Alive?
   3(b) Released Dead? 2(d): Don’t Know?
   3(e) Number of sharks lost during fight?

4. Weight of each thresher on board: Fork length of each thresher on board:

5. Did you leave for fishing this morning?
   If no goto Question 4(a) else goto question 5.
   4 (a) How many days have you been fishing?

6. Day and hour you left for fishing?

7. How many anglers are in the boat?

8. Total number of fish you caught today?

9. What kind of fishing hooks did you use and what was the hooking location on the shark? (e.g. J-hook or circle hook; tail hooked or mouth hooked).

10. Type of gear used during this trip: (troll lures, live bait, dead bait, combo)

11. Are you part of a club that has/will have record the number of threshers caught/kept/released alive/dead?

12. Zip Code of each angler

13. Provide an estimate of the number of boats missed during this interview .

Appendix E : Angler Effort Estimation: Another possible interpretation

In contrast to [23], if we assume that the mean number of angler days is given by:

$$\hat{a} = \frac{1}{\Sigma e_j} \Sigma_{j=1}^{\Sigma e_i} (a_i)'. \quad i = 1,.,t \qquad [1]$$

where $(a_i)'$ is the angler-days on boat I, which depends on $(\text{Nang})_i$ – The number of anglers on boat i and $(\text{hfs})_i$ – The number of hours fished in boat i.

$$(a_i)' = \frac{(NAng)_i * (hfs)_i}{24} \qquad i = 1,....,\Sigma_{i=1}^{t} e_i$$

and $\Sigma_{i=1}^{t} e_i$ is the total number of boat trips have thresher as target or thresher as non target catch, depending on the context, across ALL Sample days t. The variance of $\hat{a}$ is similar to [24], that is:

$$Expt(Var(\hat{a})) = \left(1 - \frac{\Sigma_{i=1}^{t} e_i}{\Sigma_{i=1}^{\tau} e_i}\right) \frac{1}{\Sigma_{i=1}^{t} e_i} \Sigma_{i=1}^{\Sigma_{i=1}^{t} e_i} \frac{((a_i)' - \hat{a})^2}{(\Sigma_{i=1}^{t} e_i) - 1} \qquad [2]$$

It remains to estimate $\frac{\Sigma_{i=1}^{t} e_i}{\Sigma_{i=1}^{\tau} e_i}$, which has a population term in the denominator.

Note that the true boat trip effort per day, $\frac{\Sigma_{i=1}^{\tau} e_i}{\tau}$, can be estimated using the sample boat trip effort per day $\frac{\Sigma_{i=1}^{t} e_i}{t}$. That is,

$$Expt\ \left(\frac{\Sigma_{i=1}^{t} e_i}{t}\right) = \frac{\Sigma_{i=1}^{\tau} e_i}{\tau} \qquad [3]$$

Thus, cross multiplying both sides, we find that $\left(\frac{\Sigma_{i=1}^{t} e_i}{\Sigma_{i=1}^{\tau} e_i}\right)$ can be estimated as $\frac{t}{\tau}$. Plugging this result in [2] results in Equation [4] for estimating the variance of the estimator in [1].

$$Expt\{(Var(\hat{a}))\} = \left(1 - \frac{t}{\tau}\right) \frac{1}{\Sigma_{i=1}^{t} e_i} \Sigma_{i=1}^{\Sigma_{i=1}^{t} e_i} \frac{(a_i - \hat{a})^2}{(\Sigma_{i=1}^{t} e_i) - 1} \qquad [4]$$

A possible consequence is that the mean anglers per boat is independent of the number of boat trips per day since the mean number of anglers per boat is independent of the number of sample days. Whereas, the boat trip estimate varies from one sample day to another. So, due to the independence of one of the random variables on the number of the sample day and the dependence of the other on the number of the sample days, these two random variables are independent.
Thus

$$\hat{A} = \hat{a}\hat{\hat{e}}$$

and

$$Expt(Var(\hat{A})) = \hat{a}^2 Expt(Var(\hat{\hat{e}})) + \hat{\hat{e}}^2 Expt(Var(\hat{\hat{a}})) - Expt(Var(\hat{a}))Expt(Var(\hat{\hat{e}}))$$

This estimator and the estimate of the variance of the product is NOT recommended but it is interesting to note that that the estimator uses a more typical ratio estimator. The expression for the obtained variance estimate is also the same expression as obtained in the CRFS Manual (2009), Page 7.

Appendix F : Adaptive Cluster Sampling vs. Simple Random Sampling

In this section, the condition required for lower variance of estimators obtained using adaptive sampling vs. simple random sampling is discussed. The corresponding theorem and derivations are shown in Thompson (1992), Page 275 and page 284 respectively.

In the following, the condition under which adaptive sampling will have a lower variance is stated and the reader is encouraged to refer to Thompson (1992) for further details.

Consider an adaptive cluster sampling design with an initial sample of n sample units chosen. Note that this isn't true in this design, since only a subset of sites are chosen randomly, the other set of sites are 'priors', chosen with probability 1. Hence for the initial sample day, the variance of estimates of catch and effort obtained in these prior sites is 0. Note that this design isn't a case of Bellweather sampling where only representative sample units are chosen, sub-sampling amongst priors is also performed along with the set of random sites.

Let there be a simple random sample experiment involving sampling a fixed set of sampling units n^* on each of the sample days. The very nature of Adaptive Sampling is such that, we can never predict the number of sample units that will be sampled beforehand. The following inequality assumes that in the adaptive design, the n sample units are chosen randomly.

The very nature of adaptive sampling is such that, we can never predict the number of sample units that could be sampled beforehand unlike, simple random sampling.

Based on these assumptions, adaptive sampling will have a lower variance than simple random sampling only if,

$$\left(\frac{1}{n}-\frac{1}{n^*}\right)\sigma^2 < \frac{N-n}{Nn(N-1)}\Sigma_{k=1}^{K}\Sigma_{i\in A_k}(y_i-w_i)^2$$

Where

N=Total number of possible sample units per stratum
n=Number of initial number of sample units chosen for adaptive sampling
n^*= A fixed set of sample units for the simple random sample experiment.
K= Number of networks in the population
A_k = Number of sample units in network k
w_i = Average of the observations in the network which includes the i^{th} sample unit
y_i= Catch/Effort observed in the i^{th} sample unit
σ^2 = Finite population variance for a simple random sample

Thus adaptive sampling will be better than simple random sampling if the within-network variance of population is sufficiently high and occurs in patches. This is an assumption made regarding the population distribution of threshers for the choice of an adaptive sampling design. The variance of adaptive sampling estimators is always lesser.

Implications:

1. If $n=n^*$, then the variance of estimators obtained using adaptive sampling is always lower than the variance obtained using SRS. This is because the LHS becomes 0 and the RHS is always greater than 0.
2. If $n>n^*$, then the LHS is a negative number and the variance of estimators obtained using adaptive sampling is ALWAYS lower.
3. SRS will require a larger fixed sample for obtaining the same precision as adaptive sampling, that is, $n^*>n$.