



## Mid-Atlantic Fishery Management Council

800 North State Street, Suite 201, Dover, DE 19901  
Phone: 302-674-2331 | FAX: 302-674-5399 | www.mafmc.org  
Michael P. Luisi, Chairman | G. Warren Elliott, Vice Chairman  
Christopher M. Moore, Ph.D., Executive Director

# MEMORANDUM

**Date:** April 17, 2018  
**To:** Dr. Chris Moore, Executive Director  
**From:** Tilefish Survey Review Committee, Council Staff  
**Subject:** Report of the Pilot Tilefish Survey Review

In January 2017, the Council funded a fisheries-independent pilot survey out of SUNY Stony Brook for golden tilefish (GTF) and blueline tilefish (BLT) from Georges Bank to Cape Hatteras. The goals and objectives put forth by the survey are as follows:

1. Establish a comprehensive fishery-independent bottom long-line survey for golden and blueline tilefish along the Atlantic coast
2. Quantify the number of individuals and size-structure of the two species
3. Determine the spatial distribution of both species and identify preferred depth strata across size range
4. Evaluate the role of environmental variables in driving the observed spatial distribution patterns
5. Evaluate proposed sampling intensity and statistical power

Following publication of the final report in December 2017, a Pilot Tilefish Survey Review Committee (Committee) was established to peer review the report and its findings and provide recommendations regarding future tilefish research and survey implementation. The Committee met via webinar on April 16, 2018 with the following Committee members in attendance: Paul Rago (MAFMC SSC), John Carmichael (SAFMC Staff), George Sedberry (SAFMC SSC), Marcel Reichert (SAFMC SSC), Nate Bacheler (SEFSC), Dave McElroy (NEFSC), Matthew Seeley, Brandon Muffley, and José Montañez (MAFMC Staff).

The goals of the meeting were to respond to the terms of reference (TORs) that address the survey objectives and provide recommendations on next steps/future directions for the survey. The meeting began with a welcome and introduction from Council staff followed by an overview of the TORs. The Committee then provided comments to address each TOR.

## Pilot Tilefish Survey Review Terms of Reference

1. *State if the final report addressed the goals and objectives stated in the request for proposals.*

The Committee determined that the report addressed all goals and objectives identified in the survey proposal and request for proposal. The requirement in the request for proposals identifying the need for a survey to sample the full range of GTF and BLT from the northern extent of their range to Cape Hatteras was met. The design, execution, and analysis were appropriate, however, the catches, especially for BLT were too low to develop a reliable index with sufficient precision for use in stock assessments. The principal investigators (PIs) adequately demonstrated the feasibility of field methods and provided strong analyses of the results, despite the low sample size of BLT.

2. *Evaluate the appropriateness and robustness of the survey design and methodology. Were the results of the pilot survey clearly interpreted?*

The survey design was robust and conducted in collaboration with all stakeholders, but given the low catches, in particular for BLT, the design may have to be re-evaluated (potentially by increasing the number of stations) to reduce uncertainty. The implementation protocols appeared to be feasible and the interpretation of the data was appropriate and valid given the effective post hoc analyses, which contained good recognition of the limitations.

### Comments:

- Bait size should be relative to hook size instead of standardizing bait size across all hook sizes.
- Consider use of Smith (2016) methodology for hook saturation bias.
- Frequency of zero catch (any species) do not cause concerns about 1 gear saturation unless the zero catches are the result of baitless hooks. The overall catch rate was only 5% catch rate (30,000 hooks with 1,300 fish caught, Supplemental Table 1) and about 2.5% for tilefish. However, if the hooks are baitless upon haulback then other species or invertebrates may be stripping the bait, thereby reducing potential catches of the target tilefish species.
  - Provide information on leading hook with bait or not; and if a baited hook came back empty (no catch and no bait)
- Need to have a more consistent soak time. Look into standardizing the soak time with the South Atlantic surveys.
- Look at species composition and bycatch species relative to soak time.
- Update and clarify the supplemental figure that shows total catch relative to soak time as there may be species-specific differences relative to the soak time due to differences in behavior, for tilefish and other species.
- Note bait presence or lack of, on a per hook basis to assist in identifying an appropriate soak time.
- Provide additional information as to when sets were made and how many were before and continued until after sunset.

- There appears to be inconsistent use of “effort” across the report in terms of CPUE (e.g., compare Fig 2 vs Fig 8). This should be defined within each graph or be applied consistently across the report.

Recommendations:

- According to BLT data collected from the MARMAP SEAMAP-South Atlantic Long Bottom Longline survey south of Cape Hatteras, BLT bottom substrate preference may differ from GTF. The shallowest sampled strata were 75 meters, so the survey may have missed BLT in shallow waters (~ <50 meters).
  - Use only one hook size (small or medium) may be more appropriate in future. The small hooks seem to have overall higher catch rates and an increase in the proportion of undersized fish. The current assessment model provides little evidence of incoming recruitment and would therefore be improved with such information.
    - If continued as is, need to think about how a multi-hook survey could be used in an assessment. Either separate indices would be developed or a standardized weighting approach would need to be developed. This may add unnecessary complexity to the relative abundance index without adding much to the assessment. Separate hook-specific abundance indices would have higher variances and proper estimates of the covariance among catch rates for different sizes would be difficult to compute.
    - The pilot survey seems to have clarified the hook selectivity issues for Golden Tilefish. Using the small or medium hook size may be best for moving forward. Analyzing effect of bait size and hook size effects would require another pilot study.
    - Clarify that the same hook brand was used and standardize it with South Atlantic surveys.
  - It would be very useful to have information from hook timers, but the reviewers understand the difficulties associated with them.
    - If an appropriate approach to use hook timers can be developed, data collected from hook timers may only be needed for a year or two and could be stopped.
  - There was limited temporal coverage (just summer), which does not provide much information on temperature preference – need additional sampling to occur.
    - The survey’s timing in mid-summer might have helped to lower spiny dogfish bycatch
3. *Could this survey design and methodology be used to develop an index of abundance and stock dynamics for tilefish?*

If the survey was continued as conducted in the pilot, an index of abundance could be developed for GTF. Due to the low encounter rates for BLT, many aspects of the survey design would need to be modified (depth strata, samples per strata, hook size, bait size, etc.). Furthermore, the Committee stated it is premature to make these recommendations given the magnitude of interannual variability is unknown. The Committee suggested that

the survey may be more effective if the targeted species (and associated habitat/location) was alternated to every other year.

Comments:

- Consider adding a table of hook size (as columns) vs fate (caught, baited, empty) as rows to demonstrate potential effects of gear saturation.
- Consider how rates for different hook sizes would be handled as tuning indices in stock assessment models.

Recommendations:

- Consider a multi-year option with increased sampling intensity; or one targeted species per year with specific design and the other in the next year with a specific design
    - Likely only 1-year break – lose the information on recruits into the fishery (smallest fish caught (30-40 cm) are ~3 years old and are retained by fishery at ~4 years old)
  - Modify strata in future surveys to cover the shallower BLT habitat.
4. *Could the survey design and methodology presented in the final report (or a modification of it) be coupled with fishery-independent surveys conducted by SEAMAP-South Atlantic?*

The Committee concluded that modifications are necessary to make the surveys directly compatible. Survey compatibility would only apply to BLT due to the one-unit stock's extensive range. GLT are separate stocks, so the development of one comprehensive survey index would not be helpful to the assessments for GLT in the Southeast.

The MARMAP SEAMAP-South Atlantic Long Bottom Longline is the most compatible survey. A detailed description of this report is available in Carmichael et al. (2016). The main differences are the strata and depth sampled, number of hooks, hook size (one versus 3), bait (whole squid vs 1"x1"), and sampling season. (The survey in the SA is conducted and funded as a collaborative effort by SEAMAP-SA and MARMAP, both housed at SCDNR).

5. *Identify strengths and weakness on the continuation (development) of a comprehensive tilefish survey, including comments on applicability of the survey design, and comprehensive versus single species survey approach.*

Overall, the investigators have done an outstanding job of evaluating the results to date through identifying relationships between environmental data and catch rates. This information should be used in the future to assist in refining the survey coverage.

The Committee agreed there is a clear need of a comprehensive long-term survey for tilefishes in the mid- and south Atlantic regions. This type of survey has been listed as a high priority research area in various (SEDAR) stock assessments and other reports, and has strong support from the South Atlantic and Mid-Atlantic SSCs and Councils.

Continuation of this survey in a form that will increase catches and is comparable with survey efforts South of Cape Hatteras (SEAMAP-SA/MARMAP) will be extremely useful for (region wide) BLT stock assessments. Whether the survey is conducted annually or every other year will depend highly on availability of funds and cost-benefit of conducting a survey for either or both GLT and BLT.

Sampling efforts in collaboration with the industry can be cost effective and powerful in terms of buy-in (stakeholder involvement). However, the nature of a long-term fishery independent survey requires consistency (e.g. sampling methods and seasons) and longevity. This means that it is imperative that participants are cognizant of the scientific constraints and long-term commitment requirements for participation.

Comments:

- The investigators recognize the limitations of an optimal allocation scheme, when compared to the current survey design, since it depends strongly on the magnitude of estimated variance. In many instances, optimal configurations are not stable over time. Implementation of optimal design for year  $t$  in year  $t+1$  may in fact lead to worse performance.

Recommendations:

- Consider how the hook size data would be used to either create an estimate of abundance or be used in a stock assessment model.

6. *Make recommendations to improve the survey design and implementation; e.g., sampling effort for golden and blueline tilefish, cost-benefit of changes to the survey design. Comment on potential funding sources for the implementation of future fishery-independent tilefish surveys.*

The Committee agreed that it may be highly cost effective to run the survey for a single species every other year (GTF, BLT, GTF, BLT, etc.). This will hopefully reduce fuel costs, boat time, staff effort, etc. due to not having to cover as much spatial coverage in each survey. This will allow for more stations per species and ultimately increase the overall precision of the survey.

See TOR 2 for additional specific comments on ways to improve the survey design and implementation.

Comments:

- Consider effects of bait loss and saturation on abundance estimation, using methodology of Smith 2016.
- Consider the effects of multiple hook sizes and rationale for retaining.
  - Need a more detailed consideration of size selectivity.
  - Propose table of hook size vs size composition—supplemental figure 3. But, scaled for numbers caught.
- Work up the current (flow) meter data.

- This may serve as an adjunct with a camera related system and may also help define a bait plume footprint.
- Analyses of effect of soak time was inadequate because there was not enough variation in soak time. Regression is pretty much determined by high leverage points on boundaries. A plot of confidence intervals would be helpful.
- Stratum variances can be expressed as a negative binomial with  $\text{predicted\_var} = \text{mean} + \alpha * \text{mean}^2$ , with alpha about  $\sim 1.04$ . This has important implications for precision of estimates and for future survey designs.
- Can boost the revenue slightly by only using the small hooks.

Potential funding sources: Marine Fisheries Initiative (MARFIN), NOAA Cooperative Research Program (CRP), Southeast/Northeast Fisheries Science Centers.

Recommendations:

- The survey may not have adequately sampled BLT habitat. BLT bottom type preference may differ from GTF. The shallowest sampled strata were 75 meters, so the survey may have missed BLT in shallow waters ( $\sim < 50$  meters). Alter survey strata locations to gather more informative data and thus, become more cost effective.
- There was limited temporal coverage (just summer), which does not provide much information on thermal habitat preference across seasons – need additional sampling to occur.

7. *Could the survey design and methodology presented in the final report (or a modification of it) be coupled with other fishery-independent surveys? E.g., method for assessing blueline and golden tilefish stocks using a baited underwater video system.*

Yes, this survey design could be coupled with other fishery-independent surveys. Coordination of efforts and survey design will significantly increase the utility of the collected data for assessments and management. Coupling with additional survey methods can be useful yet, many surveys use different gear, sample at various times, target different regions, etc. The lack of consistency between surveys needs to be considered and adjusted on a survey-to-survey basis to help all variables become more consistent.

Comments:

- The use of video may be limited due to the water depth and associated low light conditions, as well as, the need to cover a much larger area.
  - This may require a light source, which may affect the survey observations and survey design.

Recommendations:

- Think in the context of what is needed for future assessments and what is actually feasible in a single survey.
  - Video surveys may be an effective approach for evaluating habitat/burrows, but there is no way to know if they are occupied/

- Is this an attempt to build a mechanistic, multi-gear estimator of abundance?
- OR, is it part of a population model that incorporates removals with fishery independent and dependent abundance indices?
- One potential linkage would be to use the bottom current measurements to develop a bait plume footprint.
- Differences in soak time, hook size, and hook spacing may be important. May need to rely on literature or conduct separate experiments.

#### 8. *Other Comments or Issues*

- How important is it to conduct a BLT survey index of abundance – considering the cost-benefit of the survey and the fishery?
- Survey experienced limited bycatch and was able to focus on tilefish.
- Commercial vessel platform – probably the best approach from a practical (set-up, crew etc.) and public relations approach; need clear protocols for captains to follow to minimize their effects and minimize leeway.
- Operational costs (\$6,000) was quite reasonable for other fishing-based platforms and when compared to the use of a scientific vessel.
- Depth and area stratification is appropriate for GTF – may want to reconsider for BLT.
- The easiest way to increase the precision is to increase the number of stations. Increasing the catch per station does very little to nothing for the precision (see additional comments on survey catch rates and variance).
- Standardize methods (including type and number of hooks, length of gangions, length of ground line, soak time, sampling season, and bait) among surveys regionally.
  - There are significant operational and analytical challenges to making the different surveys similar enough to combine data.
    - In some cases, this may not be surmountable or creates significant analytical problems (e.g. different habitat and bottom types).
- Consider (continued) use of hook timers.

#### References

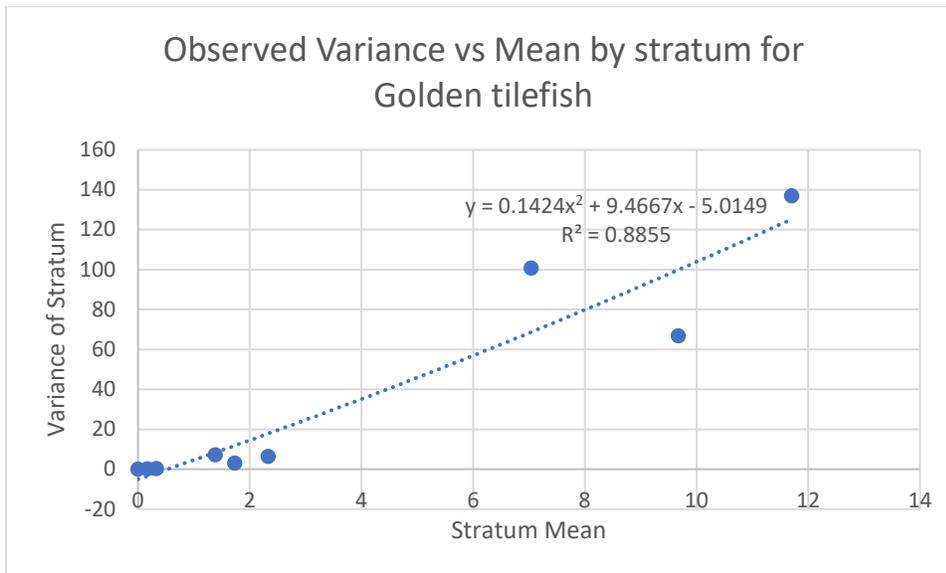
Carmichael, J, M Duval, M Reichert, N Bacheler and T Kellison. 2015. Workshop to determine optimal approaches for surveying the deep-water species complex off the southeastern U.S. Atlantic coast. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SEFSC-685. 24 p. doi:10.7289/V5GB222C

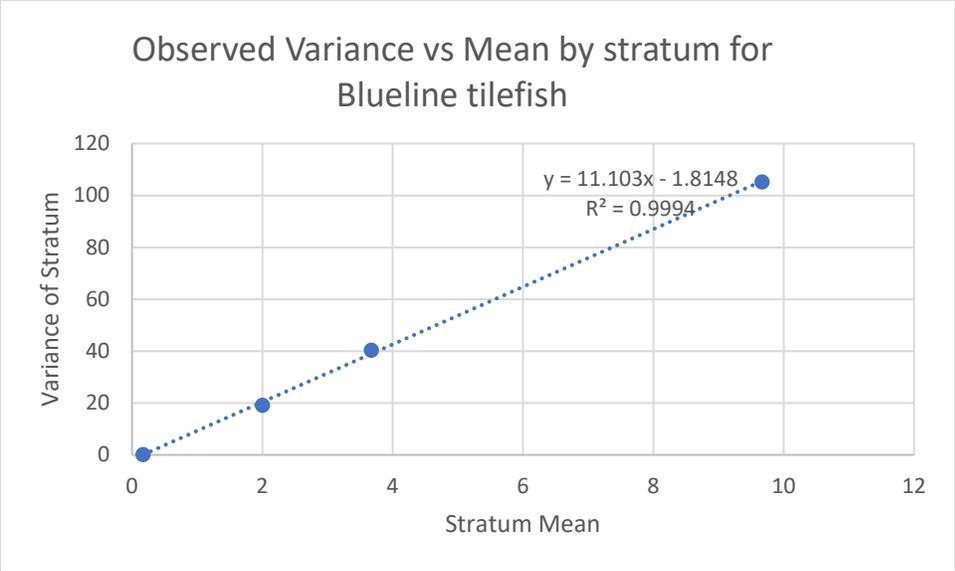
Smith, S. J. 2016. Review of the Atlantic Halibut longline survey index of exploitable biomass. Can. Tech. Rep. Aquat. Sci. 3180: v + 56 p

## Appendix

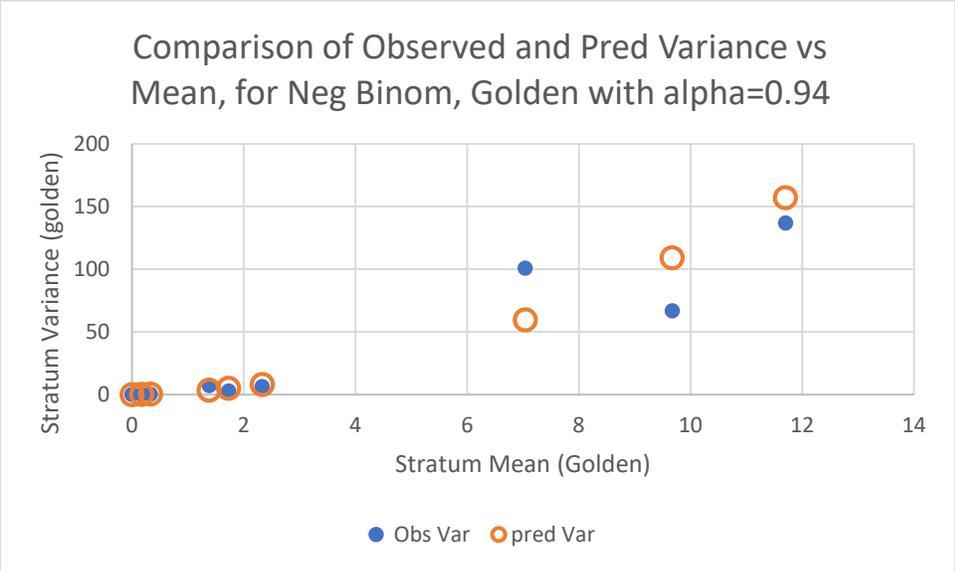
### Additional comments on survey catch rates and variance.

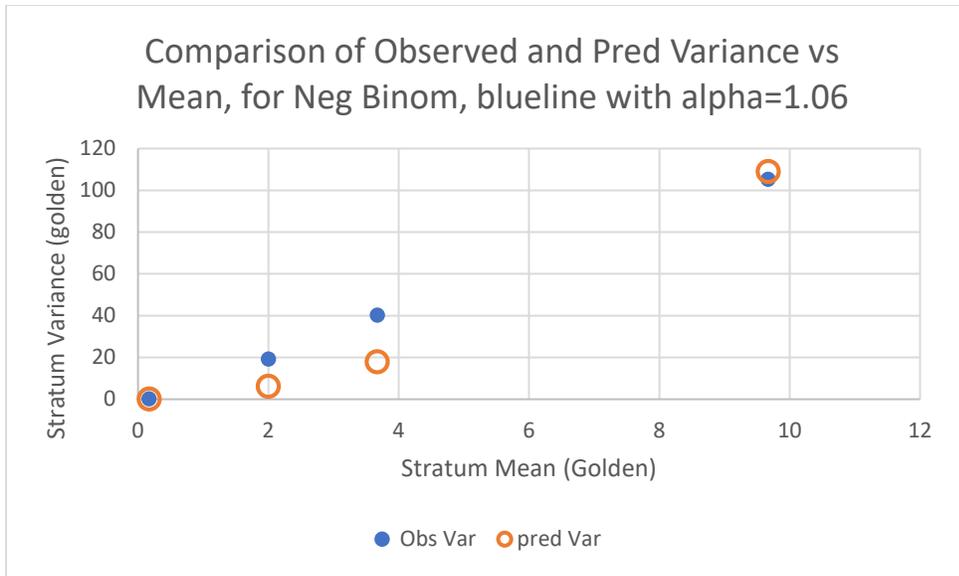
At several times during the discussion last week, we noted the low catches of blueline tilefish in particular, and the difficulties that posed for developing an abundance index. Such concerns are warranted for the collection of biological samples. Moreover, we are generally more comfortable stating that relative abundance has declined 50% when catch rates go from 10 per set to 5, rather than 0.1 to 0.05 per set. However, increasing the total average catch per set should not be the primary determinant of that comfort level. Instead it is the relative precision of the estimate that should give us comfort, not the magnitude per se. To examine this, I first looked at the relationship between the variance and mean catch rates per stratum. If fish are distributed in patches, then theory suggests that the catches should follow a negative binomial model wherein the variance is a function of the mean plus the mean squared. In a Poisson model the variance will equal the mean. Using the data in the report, (Table 7, Table 8, Table 9) I plotted the variance vs the mean for Golden, Bluelines and combined and fitted a negative binomial model as  $\text{Var} = \text{mean} + \alpha * \text{mean}^2$ . The results are shown below:





Negative Binomial parameterization of Mean variance relationship for Golden Tilefish





The negative binomial model seems plausible for golden but less so for blueline. To examine the effects of increased catch per set, I used the above fit for the negative binomial to predict the variance for an increase in the mean of 10X. Since the variance increases with the square of the mean, you might expect that very little gain in precision occurs. The computations are given below.

### Baseline Scenario

Species	stratum	n_h	Wh	ybar_h	sd_h	var_h	var_h/n_h	Wh^2
Golden	3.2	10	0.157	0.17	0.41	0.1681	0.01681	0.024649
Golden	3.3	26	0.215	7.04	10.04	100.8016	3.876984615	0.046225
Golden	3.4	3	0.012	0.33	0.58	0.3364	0.112133333	0.000144
Golden	4.2	10	0.147	1.38	2.67	7.1289	0.71289	0.021609
Golden	4.3	20	0.172	9.67	8.17	66.7489	3.337445	0.029584
Golden	4.4	3	0.016	2.33	2.52	6.3504	2.1168	0.000256
Golden	5.2	6	0.084	0	0	0	0	0.007056
Golden	5.3	22	0.184	11.7	11.7	136.89	6.222272727	0.033856
Golden	5.4	3	0.014	1.73	1.73	2.9929	0.997633333	0.000196

y-strata 5.62465

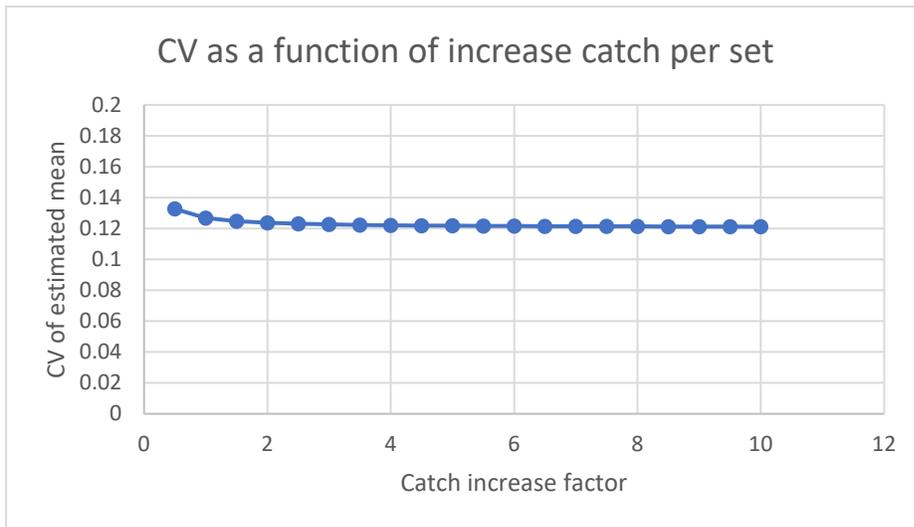
Var\_stratified 0.505182626  
 SD stratified 0.710762004  
 CV\_stratified 0.126365552

The 10X scenario is below

## 10X Catch Scenario

				Factor 10	Pred SD	Pred Var				
<i>Species</i>	<i>stratum</i>	<i>n<sub>h</sub></i>	<i>Wh</i>	<i>ybar<sub>h</sub></i>	<i>sd<sub>h</sub></i>	<i>var<sub>h</sub></i>	<i>var<sub>h</sub>/n<sub>h</sub></i>	<i>Wh<sup>2</sup></i>		
Golden	3.2	10	0.157	1.7	2.169239	4.7056	0.47056	0.024649		
Golden	3.3	26	0.215	70.4	72.28282	5224.806	200.9540923	0.046225		
Golden	3.4	3	0.012	3.3	3.824343	14.6256	4.8752	0.000144		
Golden	4.2	10	0.147	13.8	14.55533	211.8576	21.18576	0.021609		
Golden	4.3	20	0.172	96.7	99.10411	9821.626	491.08128	0.029584		
Golden	4.4	3	0.016	23.3	24.24676	587.9056	195.9685333	0.000256		
Golden	5.2	6	0.084	0	0	0	0	0.007056		
Golden	5.3	22	0.184	117	119.8063	14353.56	652.4345455	0.033856		
Golden	5.4	3	0.014	17.3	18.12627	328.5616	109.5205333	0.000196		
y-strata				56.2465				Var_stratified	46.44781339	
								SD stratified	6.815263267	
								CV_stratified	0.121167775	

Note that the CV is almost the same. I examined the predicted CV over a range of multipliers in the following graph.



The obvious take-home message is that increasing the catch, when the catches follow a negative binomial distribution, does not have much effect on relative precision.

In contrast, sampling theory suggest that the biggest gains in precision come when you can increase the number of stations. To examine this effect, I looked at a range of increases in the number of stations.

## Sampling Effort Increase Scenario

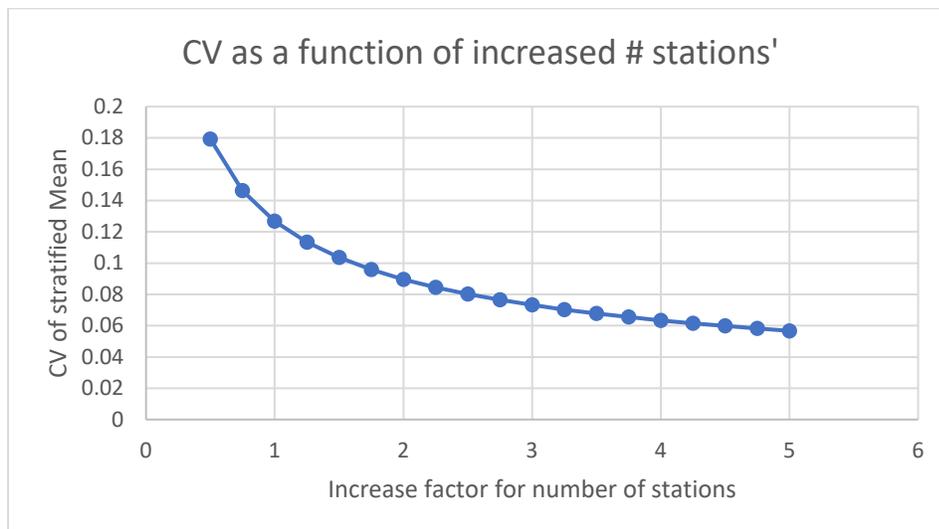
		Sample Size Adj Factor 2		Mean Adj Factor 1	Pred SD	Pred Var		
<i>Species</i>	<i>stratum</i>	<i>n_h</i>	<i>Wh</i>	<i>ybar_h</i>	<i>sd_h</i>	<i>var_h</i>	<i>var_h/n_h</i>	<i>Wh^2</i>
Golden	3.2	20	0.157	0.17	0.447276	0.200056	0.0100028	0.024649
Golden	3.3	52	0.215	7.04	7.654023	58.58406	1.126616615	0.046225
Golden	3.4	6	0.012	0.33	0.665775	0.443256	0.073876	0.000144
Golden	4.2	20	0.147	1.38	1.833187	3.360576	0.1680288	0.021609
Golden	4.3	40	0.172	9.67	10.34018	106.9193	2.6729814	0.029584
Golden	4.4	6	0.016	2.33	2.824191	7.976056	1.329342667	0.000256
Golden	5.2	12	0.084	0	0	0	0	0.007056
Golden	5.3	44	0.184	11.7	12.41232	154.0656	3.501490909	0.033856
Golden	5.4	6	0.014	1.73	2.200594	4.842616	0.807102667	0.000196

y-strata 5.62465 Var\_stratified 0.254088446

SD stratified 0.504071866

CV\_stratified 0.089618352

In this example a two-fold increase in sampling stations reduces the CV from 0.126 to 0.090. Over a range of sample size increases the effects are even more pronounced.



Of course the costs of increased sampling stations vs longer strings are not equal, but it is clear that increases in average catch per se will not do much to increase the precision (ie reduced the variance).