Review of Stunz, G. W., W. F. Patterson III, S. P. Powers, J. H. Cowan, Jr., J. R. Rooker, R. A. Ahrens, K. Boswell, L. Carleton, M. Catalano, J. M. Drymon, J. Hoenig, R. Leaf, V. Lecours, S. Murawski, D. Portnoy, E. Saillant, L. S. Stokes., and R. J. D. Wells. 2021. Estimating the Absolute Abundance of Age-2+ Red Snapper (Lutjanus campechanus) in the U.S. Gulf of Mexico. Mississippi-Alabama Sea Grant Consortium, NOAA Sea Grant. 303 pages.

(File entitled “GRSC Report for GMFMC SSC_filesizereduced_03152021.pdf” received 15 March 2021.)

By Mary C. Christman, Courtesy Professor in the Departments of Statistics and Biology at the University of Florida, Gainesville, FL

Date: 23 March 2021

In this review, I address the statistics aspects of the study design and estimation procedures but generally I will not consider the field methods (sampling gear types) used for collecting, processing or calibrating the data to standard units. Nor can I address the assumptions that are made throughout concerning the biology, behavior or phenology of red snapper. Hence, I will not assess whether the non-statistical assumptions are appropriate, the size or direction of the potential biases in the data, and whether the results can be apportioned among age-specific composition.

<u>Overall conclusions</u>

In general, the sampling strategy was only partially a stratified random sampling design. Instead, in some instances the data were assigned strata after sampling and assumed to be random samples when in fact that was not the case. In other situations, the sampling approach within a stratum was not specified or could not be determined by this reviewer and so it is not possible to decide if the stratified random design was followed. If the strata were identified after the fact and decided based on the distribution of the red snapper counts observed in the samples, then this post-stratification would lead to serious biases in the estimates. Presumably, post-stratification of observations occurred solely due to incomplete site information, such as bottom type or depth; if so, then the only issue is that the estimated variance is an underestimate due to the random ample sizes for the pre-determined strata classes.

The estimation procedures as described in the report are unfortunately not appropriate, at least for some of the data. A serious example is the decision to treat a three-stage cluster sampling design for UCB as stratified random. This most likely does not affect the estimate of abundance but would severely underestimate the variance of the abundance estimate for two reasons. First, the observations along a ship transect are not independent across strata and second, the cluster sampling within stratum is ignored.

Another issue with the estimation procedure which is difficult to assess with respect to whether the estimate could be considered a measure of *absolute* abundance is the variation among the methods for collecting and processing the data. Only a single calibration study is described; it compares data collected by ROVs and hydroacoustics on natural hardbottom in the FL region. The comparison seems to be incomplete (see comments below) and not actually applied during data analysis. As a result, some stratum estimates are indices (e.g. MaxN) or recognized to be underestimating the true abundance with unmeasurable bias and variance.

Another concern is how the data were manipulated after the processing required to convert video and acoustics etc to counts. The authors state on page 97 that “extraordinarily high values” were removed from analyses. As a result the abundance estimate and its variance are biased low. Instead, the authors might have considered using all of the data in an alternative estimator (e.g. the ratio of means) that would be refractory or robust to data that are skewed right. This would likely reduce the bias in both estimators.

One final major point is that the validation study (pg. 83-90) is of no help in determining whether the main design-based estimation procedures were correct and appropriate. The validators used different design-based estimators (e.g use of the ratio of means rather than the mean of ratios) with further subclassification within established strata but again was based on the assumption that all data were collected according to a stratified random sampling design, which they were not. Hence both sets of estimates suffer from the same issues. Instead, it would have been more helpful if the authors had compared the approach used in the main estimation procedure against the estimates and their variances that would have been obtained under the sampling strategies that were actually implemented.

Concerning the report itself, it would have been helpful if the authors had provided a description of the actual implementation and stratification in sufficient detail that a reader would have been able to reconstruct the methodologies had they been given the data. It also would have been very helpful to have been provided the estimates at every level of the stratification rather than only the region × habitat combinations.

Following are comments that relate to specific TORs where I felt comfortable reviewing.

TOR 1: <u>STUDY DESIGN AND SAMPLING APPROACHES</u>

Assess the sufficiency of spatiotemporal sampling by study strata

I address the spatial aspects but not the temporal. Throughout the document, the authors state that the sampling design is a “stratified random design (Figures 2, 3, 4; pgs 15, 31, 37, 48, 80, 96, 99). Hence, I will address whether and how this design was implemented. The issues I address include: development of strata, implemented stratification, definitions of sampling frames and sampling units, sample sizes, and site selection methods.

Before considering the details, I wish to point out that a stratified random sampling design with varying methods for collecting data within each stratum is not inherently a “biased” approach to estimation. Biases occur when the design is not followed but assumed to have been during estimation and when the collected data are processed or calibrated inappropriately.

Development of Strata

The researchers used several classification variables to develop strata for this study. The region and depth zone class variables are straightforward and reasonable given the likely distribution of red snapper. The habitat class variable is well described and covers likely habitat for red snapper: known artificial reef, known natural hardbottom, known pipelines, and all other bottom areas assigned to the UCB (unknown bottom types and undocumented other). The difficulty, of course, for sampling is that there is limited mapping of the bottom in the northern Gulf which implies that large portions of the bottom in the 10 – 160 m depths may belong to the artificial reef or natural classes but must be allocated appropriately to the UCB. This does lead to issues with classifying sites within the northern gulf to strata but the researchers made a herculian effort to do as much as possible given available data. The question of whether the lack of information for a selected site as to which stratum it belongs to is adequately handled is difficult to assess as there is no description of how that was handled by the researchers responsible for the estimation. Stuntz (pers. comm.) indicated that if the site had been assigned to the UCB it was not re-stratified into a different stratum when the bottom type was observed. This is the appropriate approach for the strata definitions.

There appears to be some effort to further sub-stratify the UCB using an additional classification variable (pg. 29) namely the categorical variable based on probability of occurrence that was developed using a Random Forest (RF) model of occurrence of red snapper in the Gulf (Figures 22, 23; Table 4; pgs. 15, 27, 29, 31, 37, 39, 79; Appendix D). On the other hand, use of the RF results as described in the estimation section (pg. 79) indicates that the RF categories were the first spatial classification variable and that these strata areas were then further subdivided by region and depth. Presumably, this was for the UCB habitat class only although that was not explicitly stated here. Instead, the authors state that mapped known hardbottom was removed and not included in the high probability class for all regions except FL. Hence, the RF classes appear to have been used for both UCB and hardbottom, at least in the FL region.

The random forest model assumes that the relationships of the explanatory variables to presence of red snapper are invariant over the entire northern Gulf and further that they are sufficient for estimating the probability of occurrence. These assumptions appear to be debunked by the cross-validation of the model using only the observations off Florida (pg. 185-186 of the report). A question for this validation effort is whether the cross-validation of the FL region was done by removing the FL data from training dataset used in the model. Another possible issue is that the RF stratification appears to match to some extent the already identified depth × habitat strata and so I am not sure that it provides much additional usefulness for controlling variance of the estimators. This is not an issue for estimation but does add additional estimation work and possibly affects variance of the estimators.

Implemented Stratification

The stratification scheme that was implemented in the study appears to have varied from the one described in the early parts of the report. For example, Figure 4 indicates that the stratification scheme used in the sampling effort involved 48 strata based on a factorial cross of 4 regions (TX, LA, AL/MS, FL) by 3 depth zones (10-40m, 40-100m, 100-160m) by 4 habitat types (large artificial, small artificial, natural, and UCB). Yet, the final estimation procedure used 54 strata (pg. 79) plus a pipeline stratum that appears to have been sampled separately. It appears for example that in the FL region only 2 depth strata were used (pg. 32) and the latitudinal extent was divided into subregions. Unfortunately, the text in the report does not match the results in the tables or figures; for example, Florida was purported to be divided into 3 latitudinal regions (pg. 79) although Figure 5 (pg. 32) shows 5 regions, each with a sample size provided.

A final comment concerning the implementation of the stratification is that at least in some instances the stratification scheme was not followed during the field sampling. I am not referring to post-stratification of sampling units due to misclassification of habitat within a sampling unit but instead referring to the design itself being ignored at the time of sampling. One example is the sampling approach used for the UCB in the TX region. Figures 16 – 18 indicate that sampling was conducted using systematic stations strung along ship transects that appear to cross 2 depth strata (10-40m, 40-100m).

The lack of clarity concerning the stratification that was implemented when selecting sampling locations implies that either it was ignored or possibly ignored and mis-used in the estimation of abundance.

<u>Definition of a Sampling Unit</u>

Since the report describes the sampling design as “stratified random” presumably the sampling unit is well-described and understood by the researchers when sampling. This does not appear to be the case in at least some instances. The definition of a sampling unit within a stratum differs between execution of sampling and estimation from the data collected. For example, the TX UCB stratum was sampled using ship transects with systematic locations sampled along each transect (Figures 16-18). There were 140 locations sampled with towed camera/hydroacoustic sleds organized into approximately onshore-offshore ship transects containing about 7-10 sampling locations and these were further subdivided into 90m horizontal x 20m vertical cell grid for small spatial scale estimation (pg. 65) yet Table 6 lists 6435 samples for this stratum . These are clearly not independent random samples (see the review of the estimation procedures later). In another example, sampling in the AL/MS artificial reefs × depth zones strata, the sampling unit is a 2×2 km grid cell with a subsample of 2 artificial reefs within each of the selected grid cells. This approach was ignored though when estimating abundance since the strata sizes were listed as the total number of artificial reefs in each depth stratum (Table 3, pg. 49) implying that the sampled artificial reefs were selected according to simple random sampling within strata. In a final example, the sampling unit used in estimation is a 15-second bin along a 15 km transect of C-BASS gear on pipeline vectors or in the UCB (pg. 73).

For the other strata, the sampling units are often not described, so it is not possible to review the appropriateness of the methods. Did the report authors grid the entire northern Gulf into 3 arc-second squared sampling units as described in Ahrens et al. or some other methodology for identifying units for sampling. It would appear to not be the case given the variation in sample sizes listed in various locations within the report (see Table 1 of this review).

Sampling Frames

The sampling frames seem to differ among the various strata which is to be expected given that the background information available to define strata by habitat and the field methods used in the different habitats to collect data. There also is an indication that in some cases, the actual sampling frame is not well defined as the sample sizes reported vary significantly for some of the strata (Table 1 of this review). This implies that the frames are probably not well understood for designing the sampling effort and for later estimation. For example, the use of ship transects in UCB in the TX region for sampling is a mismatch with the stated "random sampling" unless the starting locations of ship transects were randomly selected from a frame of all possible start locations.

In another example, the sampling frame for the AL/MS artificial reefs × depth zones strata is the set of 2×2 km grid cells sampled in an earlier study (pg. 45) and found to have artificial reefs. The actual sampling frame is the list of previously censused locations off AL that have artificial reefs in them.

This variation among strata of a sampling frame and sampling unit, in itself, is not an issue for the implementation of sampling as long as it is consistent within a stratum and a full sampling frame is available, i.e. the list of possible reefs is as complete and accurate as possible given the availability and compilation of historical habitat data. Unfortunately, there is often too little information in the report for this reviewer to determine the sampling frames used in many of the strata. The description of sampling in the report generally focuses on the field methodology for obtaining data and not on providing information on statistical aspects such as the sampling frame and units.

Strata Sample Sizes and Allocations

There is a manuscript by Ahrens et al. (pgs. 379-414) describing a simulation study for the Phase II proposal in which estimated sample sizes are listed (Tables 11 and 12, pg. 407) for the region × depth strata. The numbers are based on optimal allocation and vary according to the assumed proportion of habitat occupied by red snapper. In all cases n=30 was assumed for all artificial habitats in all depth × region strata. These values are based on creating a population of 3 arc-second squared sampling units (~35 million) between 10 and 160 m depths. These sampling units were then partitioned among the strata.

The final sample sizes by region and habitat are listed in Table 6 in the report but do not indicate either how these values were arrived at or how they had been allocated among the main strata variables or the RF or FL subregion substrata. Hence, I cannot assess whether the sample allocation was appropriate or adequate for the purpose of minimizing variance. In addition, the final sample sizes are not described consistently within the report so it is not possible to determine if the number of samples/stratum were sufficient for the study's goal of reducing the relative standard error (RSE called CV in the report) to less than 0.30.

Table 1. Sample sizes reported in text or Tables 6-7 or appendices

Region	Habitat	Sub class	Sample Size in Text	Page in Text	Table 6 Values	Table 7 values	Table 12 Ahrens et al.[g]
FL	Natural		749	32	505	295	180
	Artificial		65	32	84	84	180
	Natural+ Artificial		927	Fig 5			
	UCB[a]				530	453	
AL/MS	Natural[c]		32		32	32	114
	Artificial		130[f]		198	198	90
	Artificial[b]	Shallow AL	68	Table 3			
		Mid AL	45	Table 3			
		Deep AL	4	Table 3			
		MS	13	Table 3			
	UCB[a]				931	628	
TX	Natural		40	55	36	30	1071
	Artificial		18	55		22	90
	Artificial	Large			45		
		Small			4		
	UCB	10-100 m	140	65	6435	3538	
	UCB[d]		8	55			
LA	Natural		22[e]	68		656	603
	Artificial		42[e]	68		42	90
	UCB		1540[e]	68		3745	
Pipelines					27	15618	

[a] no explanation is provided in text in Regional Sections for FL, AL/MS
[b] number of reefs sampled, not number of selected grid cells
[c] number of “natural features”
[d] considered to be UCB due to habitat features noted during the survey
[e] unclear whether these are the number of samples from TX used for imputation or are the final sample sizes within the LA region
[f] sum of the 4 classses within artificial for AL/MS
[g] numbers are for number of artificial reefs or number of 3 arc-second square areas (~90 m x 90 m) for natural hardbottom

Site Selection Methods

There were several different methods used to select sampling units. The first was mentioned above, namely the AL/MS approach for collecting data on artificial reefs. For the AL/MS artificial reefs × depth zones strata the sampling frame is the set of 2×2 km grid cells sampled in an earlier study (pg. 45) and censused for artificial reefs. The sampling design was a 2-stage cluster sampling design with grid cells selected at the first stage and 2 reefs within a selected grid cell sampled randomly at the second stage. It is unclear how the grid cells were selected at the first stage but most likely the sampling was random selections from the sampling frame. This is not a stratified simple random sample and so will have a variance different (larger) than that of a stratified simple random sample of artificial reefs.

In another instance, the sampling of UCB in the TX region is not random sampling as implied by the phrase "stratified random sampling" as is implied in the phrase but instead is another 2-stage cluster sampling design with ship transects as the sampling units at the first stage and individual sites as the second stage. It is unclear how the ship transects were chosen at the first stage and it appears from the Figures that the sites are systematically located along the ship transect at the second stage. It would be inappropriate to treat this design as a stratified (by depth) random sample since it was neither depth stratified nor random. The difference in variance depends on the randomness of the spatial distribution of red snapper which I suspect is unlikely given the variation in depth along the transect.

Another case where the sampling was not random was for the C-BASS sampling of pipelines and UCB (pg. 74). Here the design is a 1-stage cluster sample where the cluster is a randomly selected pipeline start location where the sampling unit used in the estimation is a 15-second bin along the 15 km transect (pg. 72-73). This same design was used for the UCB as well when C-BASS was deployed.

Does heterogeneity in sampling by strata affect estimates of absolute abundance and variance around that estimate?

Yes, using different field methodologies in the different strata does affect the both the estimates and their variances but the field methods used in the different strata could not have been performed sufficiently well otherwise. The only real issue is whether abundance estimates from more than one gear type were combined within a single stratum without any calibration beforehand. For example, the C-BASS methodology was stated to have been used to generate density estimates over pipeline, UCB and natural hard bottom in all of the regions in this study (pg. 71). Yet, various other methods were also used in these regions (see each sub-section of the Regional Surveys). There was only one calibration study described in the report (FL natural hardbottom using hydroacoustics and ROVs) but not for other combinations of gears. Hence, biases could arise if these data were combined without calibration.

Evaluate assumptions and biases inherent to the design, and the directionality of those biases

I address this in my review for TOR 2, Statistics and Data Analysis

TOR 2: STATISTICS AND DATA ANALYSIS

- Evaluate the statistical methods used to analyze the data, and to construct the absolute abundance estimate and its variance.

- Is the statistical variance appropriate for habitat-specific, regional and Gulf-wide estimates?
- Are potential sources of uncertainty effectively incorporated into variance estimates?
- Are imputations made for unsampled regions appropriate, and what are the potential implications for the direction of biases in the estimates.

In the section, I review briefly the data processing steps as described in several different sections of the report and the statistical methods as described mostly in the section entitled "Final Abundance Estimates" (pg. 78 – 89).

<u>Data Processing</u>

ROVs Only
ROVs were used for estimation in FL for natural bottom and artificial reefs and in AL/MS for natural bottom. The ROVs used video to capture red snapper during deployment. In FL, the ROV was deployed using 4 orthogonal 25m long transects at natural sites and using the "point-count" method (Patterson et al., 2009) at artificial reef sites. Although not explicitly stated, it appears that ROVs were placed in a stationary location at each sampled natural bottom site in the AL/MS region (see description of deployment on reefs, pg. 45).

The data were processed differently in each of these two regions. In FL, the entire video at each sampling location was processed and all red snapper counted. This was converted to a density (#/area) (pg. 33). Assuming counts were low (which they appeared to have been), this approach is appropriate.

In AL/MS, the MaxN count was used and standardized to a density (#/m^2) using the average area (417m^2) (pg. 50). This average area was derived from calculations based on ROV deployments at artificial reefs (pg. 45). There are two issues here with the estimation procedure used to obtain a density for AL/MS. First, MaxN is the largest number of fish in a frame and so if multiplied over the total area observed on a reef would be an overestimate of density and second, the use of an average area surveyed for reefs as opposed to the actual area surveyed at each site (natural or reef). The average is only appropriate if the ROVs were always deployed in exactly the same manner in the same water clarity and for the same length of time. This aspect is not clear and so this approach could introduce a bias whose direction is unknown.

ROVs + Hydroacoustics
This combination of gears was used at natural bottom sites in the FL region to compare density estimates for possible calibration whereas in TX they were used mainly at artificial reef and natural bottom sites to obtain abundance estimates. In TX, the ROVs provided species composition for the hydroacoustic data and the proportion of fish that were age 2+ red snapper.

In FL the hydroacoustics array was deployed in a "flower pattern" of towed transects although the length of these transects is not given. In TX, the deployment of the hydroacoustics gear differed over the two habitat classes. For large artificial reefs, the data were collected in 10m depth layers and at natural bank and small artificial reef sites, surveys consisted of four 500-m transects in a radial pattern (Reynolds et al. 2018) centered on the geographic station position. At large artificial reef sites, a back-and-forth "mow-the-lawn" survey pattern was used to survey the entire artificial reef site (pg. 57).

In FL, the hydroacoustics data were processed to extract count estimates within 5x5m "cells" along the transects. These densities were then interpolated over the total area enclosed within the transects (pg. 37) using kriging to obtain a geostatistical surface of density estimates in each 5x5m cell (note that kriging also provides estimates of the variance). "To estimate site-specific Red Snapper abundance, the proportion of Red Snapper occurrence relative to all fish species recorded in video survey data was multiplied by the sum of fish abundance calculated for all cells at the site. To estimate site specific Red Snapper density, total Red Snapper abundance counted in given survey was divided by the areal coverage of the sonar survey conducted at that sampling site" (pg. 37). These density results were then compared to density estimates obtained from the ROV deployed at the same location (Figure 7). So, from this description, the kriged densities were converted to cell abundances and then summed over all cells in the kriging model. This abundance was then converted to a density and compared to the ROV density estimates. Not surprisingly, there was a serious mismatch, likely due to the overextension of the hydroacoustics transects to areas not natural hardbottom (as indicated in the text). A different comparison could have been to clip the kriged hydroascoustics surface to the area covered by the ROV transects and compare those.

In TX, the hydroacoustics data were processed to obtain an estimate of the number of fish within a school. "Schools were scaled, and then an abundance estimate was calculated using RStudio (R Core Team version 3.5.1; RStudio Team, 2018). Schools were paired with border targets within 7 m by estimating the minimum distance between border targets and each school. The paired TS and S_v values were converted to linear densities and then to volumetric fish density using the formulae described above. Abundances by layer were produced by averaging the kriged areal densities and multiplying by the total area of the layer. We then took the abundance value and applied the proportion of age-2+ Red Snapper, as indicated by the video data, to produce a Red Snapper abundance by layer. Red Snapper abundances were summed across all layers to produce a Red Snapper abundance estimate for the site" (pg. 59).

Towed Sled Camera and Hydroacoustics
This combination of gears was used in TX region to obtain abundance estimates for red snapper in the UCB stratum. The hydroacoustics data were processed using the standard methods (pg 62) as were the videos (pg. 65). Although not stated it appears that the videos were used to determine species composition for use with the acoustics data.

C-BASS
This gear was used in the natural and UCB habitats and for pipelines (pg. 71). For pipelines, point locations were randomly generated from the universe of pipeline vectors and were used as center points of 15 km transects along the line. To obtain UCB transects the gear was deployed parallel to the pipeline transect approximately 1 nautical mile away. Some UCB locations were sampled based on "logistical constraints" (pg. 74). Red snapper density was estimated for each

15-second bin along the transect where the area covered in each 15-second bin was estimated. These densities were then averaged over all bins in a habitat stratum.

Evaluate the statistical methods used to analyze the data, and to construct the absolute abundance estimate and its variance.

Before going into details, a serious issue for this reviewer is that there are often-opposing statements made throughout the document that make it difficult to determine exactly how the stratification was used in the estimation of abundance. For example, the implication of the statement “stratified random sample design” is that stratification was *a priori* determined before any field work and sample sizes were allocated according to some procedure (proportional, optimal, Neyman, or other) and samples collected according to that allocation. On the other hand, there are several instances where it is clear that the described stratification was not followed and so it appears that there was some *a posteriori* allocation of sites selected independently of the strata. This is sometimes confirmed in later sections of the report

In most cases, the region × habitat stratification appears to have been used to allocate samples since the methods (gear) used for data collection had to vary among these strata; for example, the differences between sampling in the TX UCB vs the FL UCB. This is appropriate. On the other hand, the use of RF probability of occurrence categories is not clearly explained nor are depth strata always followed as indicated in the Estimation Section.

Another confusing issue is that the sample sizes by major strata were provided in several places throughout the document but often did not match (Table 1 of this report). It appears that is some cases, pseudo-replication has occurred, i.e. observations are treated as independent (in a design sense) when in fact they are a cluster of observations along a transect or within a grid cell. This causes underestimation of variance of the estimators (see below). Because of the issues with sample sizes, any question of whether the number of samples are adequate to produce unbiased and efficient estimates cannot be clearly answered due to the difficulty of 1) determining how sample sizes were used, 2) whether samples were allocated according to a stratified design, and 3) in some instances what the actual sample sizes were.

Estimation procedures are described by habitat class in the section on Abundance Estimates and I will address these in the same manner.

<u>Estimation of Totals and Variances</u>

In those areas where data were collected, “density observations were post-hoc aligned with the appropriate strata” where the density was estimated as described in other sections of the text. These values were treated as a random sample within the stratum for purposes of estimation of the abundance as well as its variance. In reality, the data were collected neither by the defined strata nor at random for many instances already noted. Although not explicitly stated, I assume that the variance of the total was the product of the variance of the mean times the area or size of the sampling frame was conducted.

Hardbottom

For FL, the mean density (#/100m^2) based on ROV transects in high occurrence habitats was calculated and expanded using the total number of sampling units of 100 m^2 grid cells (pg. 80). So, this implies that the total area of known hardbottom was gridded into cells, stratified and then individual cells randomly sampled within the high occurrence strata. Further, the stratified mean ($\bar{y} = \sum_h w_h \bar{y}_h$) was calculated and then adjusted to obtain a total where the strata over which the mean is based are the high occurrence strata only. This approach overestimates the abundance of red snapper on hardbottom in Florida if the high occurrence habitat was correctly identified and the red snapper are more prevalent there. There is no way to know this from the report, because results at the individual stratum level were not provided.

For the remaining regions, the sampling design is unclear and what the estimation procedure is based on except as described in earlier portions of the text. That is, it is unclear how stratification was used during sampling and how sampling units and sampling frames were developed.

In AL/MS, ROVs were deployed and the MaxN count obtained. Presumably, these data were obtained from stationary ROVs as opposed to transects although that is not indicated. For AL/MS, it appears that these data were converted to densities using the average area surveyed for reefs. Presumably this was then modified to be density as count/100m^2 area for further calculations. I assume that the stratum weights were the proportions of total area within the strata. If not, the stratified mean estimate would be biased (direction unknown). If the data were post-stratified into these pre-defined strata then it should be recognized that there could be additional variance unaccounted for in the estimation of variance due to the random sample sizes.

In TX, the sampling employed ROVs and hydroacoustics. Like the AL/MS region, a stratified mean density (#/100m^2) was calculated and then expanded using the total area. It is unclear again how the stratification was used during sampling.

It appears that the LA mean density estimate was based on 22 samples from the eastern and central TX hardbottom region since there appears to have been no data collected in LA. The authors give several explanations for why these data that were selected are adequate for imputing the missing information for LA. I cannot address whether this was appropriate.

UCB

In the TX region, UCB was sampled according to a 2 or 3-stage cluster design where the first stage was the selection of ship transects, the second stage was the selection of stations along the ship transect and the third stage was the 15-sec video segments. It is unclear whether every 15-sec segment was used in the estimation for the ROV and hydroacoustics data or whether the TX scientists used the systematic selection of every 40th 15-sec video for processing as was done for C-BASS data. The choice of every 40th 15-sec segment appears to have been determined by "estimate[ing] the degree of autocorrelation" but the specifics are not given. As a result, the authors treated the data as stratified simple random sampling and not as data nested within cluster stages; the results is that the sample size listed in Tables 6 and 7 are much too large. As a result, the estimated variance used in the calculations is an underestimate of the likely true estimated variance. I propose that the authors do not post-stratify the 15-sec videos but instead assume a single depth and RF zone stratum and calculate the estimated density using equations appropriate for 3-stage cluster sampling.

In addition, UCB data were collected on the TX shelf by USF using C-BASS methods. It is unclear whether these data were combined with the ROV/hydroacoustics data from TX into a single stratum estimate or whether the C-BASS data were even used in the estimation. Nor is it clear whether the two different methods for data collection were calibrated if the data were combined. The C-BASS data were collected in the UCB by displacing the gear into the UCB about 1 km away from randomly selected pipeline sampling. These sites could be considered to have been randomly selected but the sampling frame from which they were selected is smaller than the entire UCB. This would have an impact on the estimation but neither the direction of possible bias nor changes in variance are predictable without additional information about he biology of red snapper.

For the other regions, it is unclear how the data were collected but it appears to have been collected in the C-BASS program run by USF. In that case, the same caveats concerning the data analysis are relevant here as for that described for TX in that there should be some determination if the choice of transects parallel to pipelines is truly random for the entire UCB or not.

Artificial Structures
In all regions, the estimators assume direct simple random sampling of structures by strata and for some regions by size. In at least one region, AL, this is not true; there a 2-stage cluster sampling design was employed where the first stage was the grid cell and the second stage was 2 artificial structures within a grid cell. At each stage random sampling was used. Hence a stratified 2-stage cluster sampling estimator and its variance should have been used and the stratum weights modified to indicate the actual sample size and sampling frame size at the first stage. This will necessitate estimating totals within each stratum and then sum across strata rather than using a stratified mean expanded with the number of features. For the other regions, the actual sampling routine for selecting samples is unknown.

Pipelines
It appears that the pipeline data were treated as random cluster samples based on the sample size in Table 6 and the use of bootstrapping of r sub-sampling. Additional information on how the bootstrapping was applied would be helpful to determine whether the bootstrap samples were selected according to the actual methods used or if the 15-sec segments were treated as a simple random sample. The variance estimate (CV = 31%) indicates that it was done following the actual design assuming that the data themselves are not highly variable. The only concern is whether it was appropriate to post-stratify the pipeline widths during the estimation. If the data were reviewed after collection in order to make that decision and lower variance, then it is inappropriate.

<u>Imputations</u>

Imputations were done for several strata: FL northern mid-depth low occurrence UCB, FL northern deep-depth high occurrence UCB, AL/MS shallow strata UCB, AL/MS deep-depth high probability of occurrence, LA shallow strata UCB, LA hardbottom, and LA artificial reefs. In the case of the artificial reefs in LA, previous data collected in LA on the proportion of fish that were red snapper and applied to the mean data from the hydroacoustics was a good approach to estimating abundance on artificial reefs in the LA region.

Insufficient information is provided to determine 1) if the imputations are reasonable from a biological perspective and 2) whether a variance estimate was also carried forward for each mean density carried forward. In the case of (2), it should have been.

<u>Sources of Variance/Uncertainty Unaccounted for in the Estimation</u>

In several places through the report, the authors indicate that uncertainty in some values was assumed to be zero. These statements were often associated with the determination of area of habitat classes, assumptions concerning behavior of the fish observed, and the number of fish observed or estimated for a location and method.

All of these contribute to the variance of the estimates provided in Tables 6 and 7 and are mostly unknowable but in a few instances the uncertainty was estimated (e.g. AL/MS) or could have been (proportion of buried pipeline). These estimates of the variability of habitat estimates should be incorporated into the estimates of variance (standard error) for abundance. For examples such as these, that variance can be incorporated into the estimated variance of the density/abundance following Goodman (The Variance of Products, Journal of the American Statistical Association, Vol. 55, No. 292 (Dec., 1960), pp. 708-713).

The authors also provide an extensive list of other sources of uncertainty that limit confidence in the results by this reviewer. It is important that it be recognized that the uncertainty associated with the results is much larger than that indicated in the standard errors and CVs given and so the estimate of abundance has much more uncertainty than is captured in the CV that is provided.