Southeast Reef Fish Survey Video Index Development Workshop

Nathan M. Bacheler and John. T. Carmichael, Editors

SEDAR59-RD03

24 April 2018

SOUTHEAST REEF FISH SURVEY
VIDEO INDEX DEVELOPMENT WORKSHOP

May 20-22, 2014

Beaufort, North Carolina

Final Report

Hosted by the:

National Marine Fisheries Service, Southeast Fisheries Science Center

and

South Atlantic Fishery Management Council

Nathan M. Bacheler and John T. Carmichael, Editors

Workshop Time and Place

The workshop was held May 20 - 22, 2014, at the National Marine Fisheries Service Laboratory in Beaufort, NC.

Workshop Objectives

1. Identify a comprehensive suite of modeling approaches that should be considered when developing video-based abundance indices for SAFMC-managed species.
2. Provide justification for the various predictor variables included in models, use of MeanCount as a response variable, use of a calibration factor between cameras, treatment of reconnaissance samples, and treatment of 2010 video data (which were collected with a different camera and were spatially and temporally restricted).
3. Review the analytical approaches used to develop video-based indices of abundance and recommend the most appropriate technique.

Document List

Conn, P. B. 2011. An Evaluation and Power Analysis of Fishery Independent Reef Fish Sampling in the Gulf of Mexico and U. S. South Atlantic. NOAA Tech. Memorandum NMFS-SEFSC-610.

Schobernd, Z. H., N. M. Bacheler, and P. B. Conn. 2014. Examining the Utility of Alternative Video Monitoring Metrics for Indexing Reef Fish Abundance. CJFAS. 71:464-471.

Figures List

Participants

Workshop Panel

Nate Bacheler..SEFSC
Joey Ballenger.. SC DNR
Jeff Buckel ....SAFMC SSC
John Carmichael...SAFMC
Lew Coggins ....SEFSC
Walter Ingram ....SEFSC
Burton Shank ....NEFSC
Dave Somerton... ASFSC
Ted Switzer ....FL FWCC
Erik Williams ....SEFSC

Other Attendees

Julia Byrd .... SEDAR
Rob Cheshire...SEFSC
Kevin Craig...SEFSC
Mike Errigo ....SAFMC
Todd Kellison..SEFSC
Nikolai Klibanski……………………………………………….. SEFSC
Kevin Purcell……………………………………………………SEFSC
Christina Schobernd..SEFSC
Zeb Schobernd ....SEFSC
Kyle Shertzer ....SEFSC
Kate Siegfried ....SEFSC

Introduction

This workshop was convened to discuss the use of video data collected through expanded reef fish survey efforts in the U.S. South Atlantic, and consider appropriate methods for evaluating video data as a measure of population abundance. Focal species, Red Snapper and Gray Triggerfish, were selected based on interest in including video-based indices in Southeast Data, Assessment, and Review (SEDAR) stock assessments of these stocks that will be conducted through SEDAR 41. The workshop was hosted by National Marine Fisheries Service (NMFS) Southeast Fisheries Science Center (SEFSC) Beaufort Laboratory and the South Atlantic Fishery Management Council (SAFMC). Additional participants on the workshop panel included representatives from NMFS SEFSC Pascagoula Laboratory, South Carolina Department of Natural Resources (SC DNR) Marine Resources Assessment and Monitoring Program (MARMAP), Florida Fish and Wildlife Commission (FLFWC), NMFS Alaska Fisheries Science Center, NMFS Northeast Fisheries Science Center, and the SAFMC Scientific and Statistical Committee (SSC).

The workshop began with a presentation on sampling methods, including changes in sampling approaches and coverage as the program developed. Next, participants considered initial exploratory methods to develop indices of abundance from the video data and explored additional approaches for developing abundance indices. Discussion on workshop Terms of Reference and final workshop recommendations are documented through this report. Actual abundance index values for consideration during stock assessments that result from the methods recommended here are beyond the scope of this effort, and will be documented separately during the appropriate assessment processes.

History of video sampling in the US South Atlantic

The Marine Resources Monitoring, Assessment, and Prediction (MARMAP) program has conducted most of the historical fishery-independent sampling in the U.S. South Atlantic (North Carolina to Florida). MARMAP has used a variety of gears over time, but chevron traps are one of the primary gears used to monitor reef fish species and have been deployed since the late 1980s. In 2009, MARMAP began receiving additional funding to monitor reef fish from the SEAMAP-SA program. In 2010, the SouthEast Fishery-Independent Survey (SEFIS) was initiated by NMFS to work collaboratively with MARMAP/SEAMAP-SA using identical methods to collect additional fishery-independent samples in the region. Together, these three programs are now called the Southeast Reef Fish Survey (SERFS). In 2010, video cameras were attached to some traps deployed by SERFS, and beginning in 2011 all traps included video cameras.

Overview of SERFS Survey methods

The SERFS survey currently samples between Cape Hatteras, North Carolina, and St. Lucie Inlet, Florida (Figure 1). This survey targets hardbottom habitats between approximately 15 and 100 meters deep. SERFS began affixing high-definition video cameras to chevron traps on a limited basis in 2010 (Georgia and Florida only), but since 2011 has attached cameras to all chevron traps as part of their normal monitoring efforts (Figure 1).

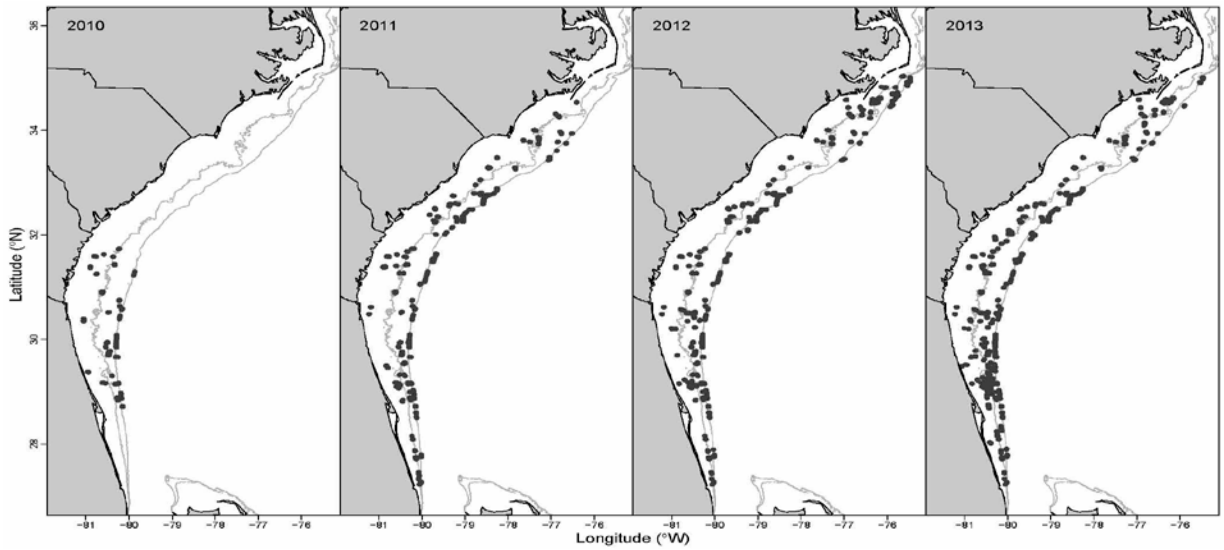

Figure 1. Annual spatial distribution of underwater video samples collected by SERFS, 2010-2013.

Hard-bottom sampling stations were selected for sampling in one of three ways. First, most sites were randomly selected from the SERFS sampling frame that consisted of approximately 3,000 sampling stations on or very near hard bottom habitat. Second, some stations in the sampling frame were sampled opportunistically even though they were not randomly selected for sampling in a given year. Third, new hard-bottom stations were added during the study period through the use of information from various sources including fishermen, charts, and historical surveys. These new locations were investigated using a vessel echosounder or drop cameras and sampled if hard bottom was detected. Only those new stations landing on hardbottom habitat were included in the analyses. All sampling for this study occurred during daylight hours between April and October on the R/V *Savannah*, R/V *Palmetto*, NOAA Ship *Nancy Foster*, or the NOAA Ship *Pisces* using identical methodologies as described below. Samples are intentionally spread out spatially on each cruise (Figure 2).

Chevron fish traps with attached video cameras were deployed at each station sampled in our study. Chevron traps were constructed from plastic-coated, galvanized 2-mm diameter wire (mesh size = 3.4 cm^2) and measured 1.7 m × 1.5 m × 0.6 m, with a total volume of 0.91 m^3. Trap mouth openings were shaped like a teardrop and measured approximately 18 cm wide and 45 cm high. Each trap was baited with 24 menhaden (*Brevoortia* spp.). Traps were typically deployed in groups of six, and each trap in a set was deployed at least 200 m from all other traps to provide some measure of independence between traps. A soak time of 90 minutes was targeted for each trap deployed.

Underwater video cameras were attached to chevron traps (Figures 3, 4). In 2010, GoPro® Hero cameras were attached over the mouth of each trap, facing away from the trap, but Canon Vixia HFS-200 video camera in a Gates underwater housing were used in 2011-2013. A second high-definition GoPro Hero video or Nikon Coolpix S210/S220 still camera was attached over the nose of most traps in an underwater housing, and was used to quantify microhabitat features in the opposite direction. Cameras were turned on and set to record before traps were deployed, and were turned off after trap retrieval. Trap-video samples were excluded from our analysis if videos were unreadable for any reason (e.g., too dark, camera out of focus, files corrupt) or the traps did not fish properly (e.g., bouncing or dragging due to waves or current, trap mouth was obstructed).

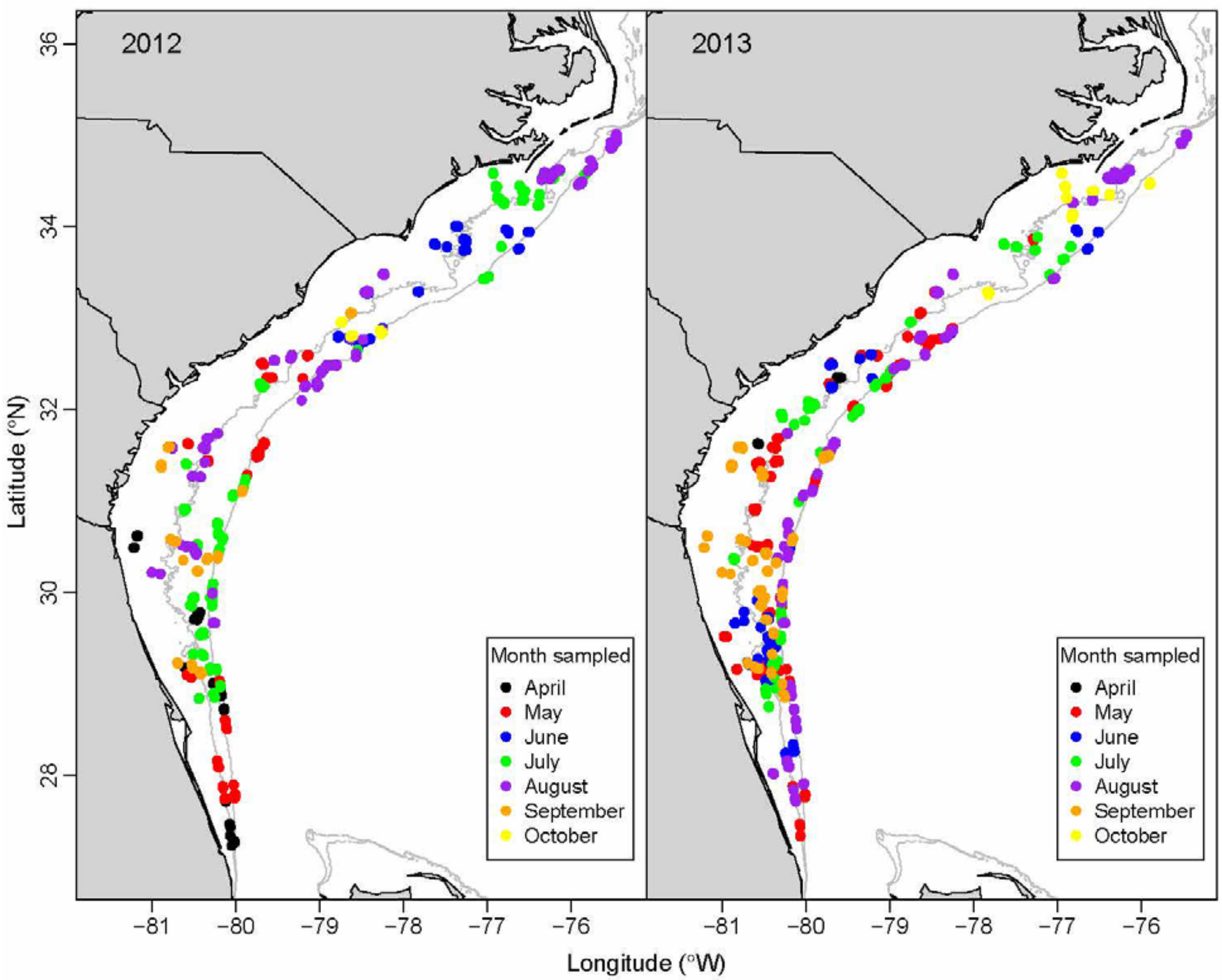

Figure 2. Seasonal distribution of SERFS video sampling, with 2012 and 2013 shown for example.

SERFS employs many video readers to count fish on videos. There is an extensive training period for each video reader, and all videos from new readers are re-read by fish video reading experts until they are very high quality. After that point, 10% or 15 videos (whichever is larger) are re-read annually by fish video reading experts. Video readers also quantify microhabitat features (percent of bottom that is hardbottom, maximum substrate relief, substrate size, coverage of attached biota, predominant biotic type, and maximum biotic height), in order to standardize for habitat types sampled over time. Water clarity was also scored for each sample as poor, fair, or good. If bottom substrate could not be seen, then water clarity was considered poor, and if bottom habitat could be seen but the horizon was not visible, water clarity was considered fair. If the horizon could be seen in the distance, water clarity was considered to be good. Including water clarity in index models allowed for a standardization of fish counts based on variable water clarities over time and across the study area. A CTD cast was also taken for each simultaneously deployed group of traps, within 2 m of the bottom, and water temperature from these CTD casts was available for standardization models.

Relative abundance of reef fish on video was estimated using the MeanCount approach (Conn 2011; Schobernd et al. 2014). MeanCount was calculated as the mean number of individuals of each species over a number of video frames in the video sample. Video reading time is limited to an interval of 20 total minutes, commencing 10 minutes after the trap landed on the bottom to allow time for the trap to settle. One second snapshots are read every 30 seconds for the 20-minute time interval, totaling 41 snapshots read for each video. The mean number of individuals for each target species in the 41 snapshots is the MeanCount for that species in each video sample. Some modeling approaches considered require count data instead of continuous data like MeanCount. These analyses used a response variable called SumCount that is simply the sum of all individuals seen across all video frames. SumCount and MeanCount track exactly linearly with one another when the same numbers of video frames are used in their calculation. Therefore, SumCount values were only used from videos where 41 frames were read (~99% of all samples).

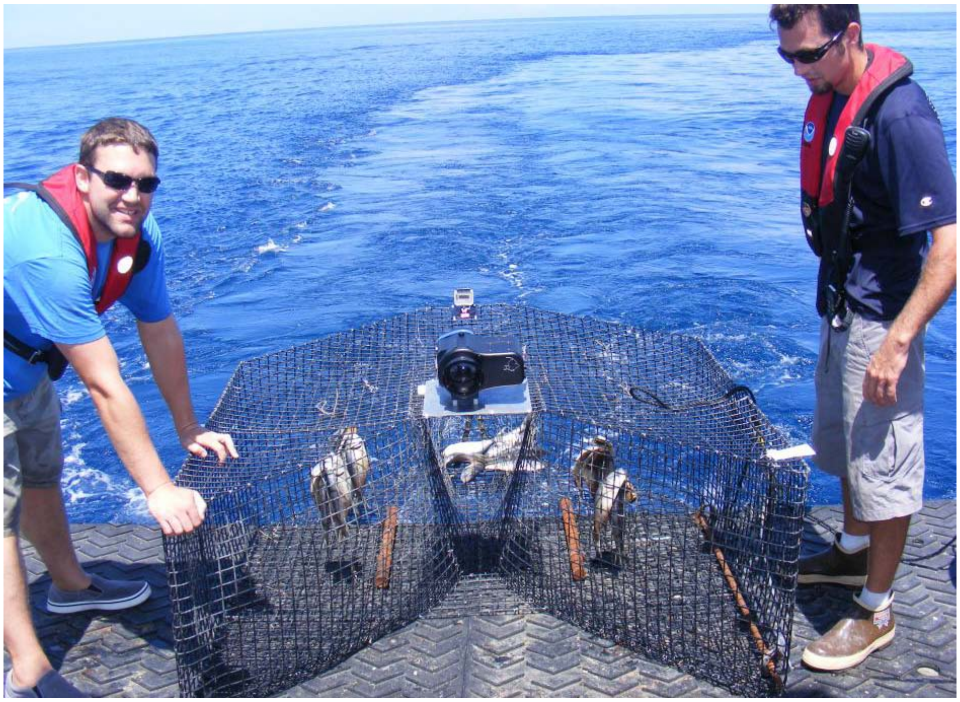

Figure 3. Chevron trap used by SERFS, with attached underwater video cameras.

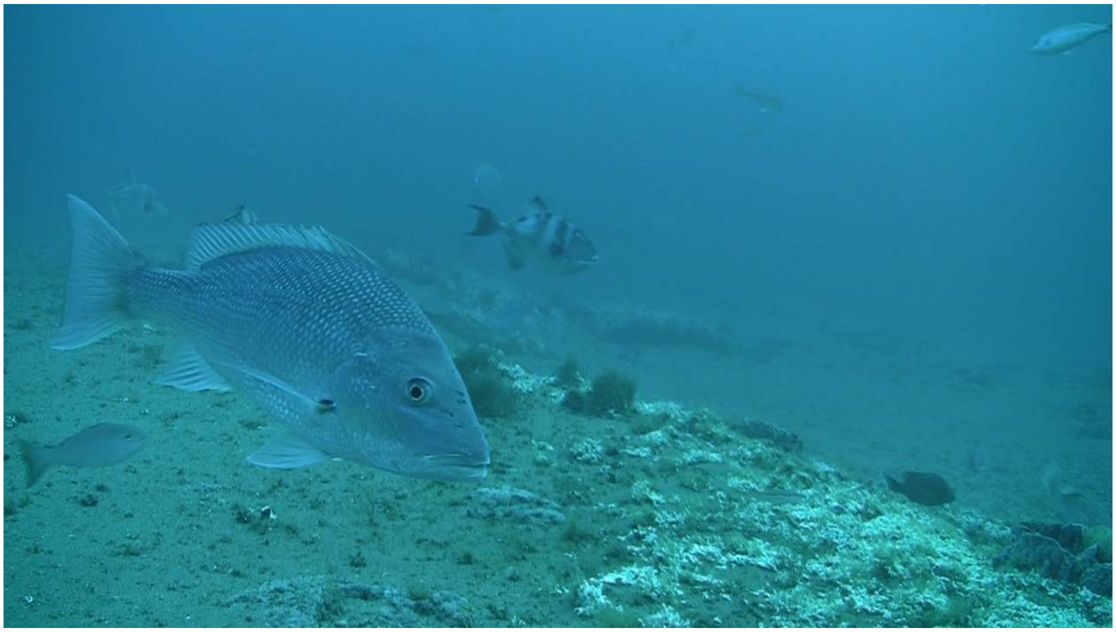

Figure 4. Still frame from an underwater video taken by SERFS off northern Florida in 2011.

Panel Discussions and Recommendations

Objective 1

Identify a comprehensive suite of modeling approaches that should be considered when developing video-based abundance indices for SAFMC-managed species, as well as potential strengths and weaknesses of each approach.

Nearly all previous fishery-independent and fishery-dependent indices of abundance in the region have used delta-generalized linear modeling (delta-GLM) approaches. Two additional analytical approaches were also considered: delta-generalized additive models (delta-GAM) and zero-inflated models. Each of these modeling approaches is considered in detail below.

Delta-GLM models contain two separate submodels: one modeling the probability of obtaining a zero (i.e., presence-absence submodel) and a model describing the mean and variance when the species was caught or seen (positive catch submodel). Predictor variables can be included as linear (or polynomial) effects, but often predictor variables are converted to categorical variables containing a variety of levels. Lognormal and Gamma error distributions are the most often used error distributions considered for positive catch submodels. The primary benefits of delta-GLM models are their simplicity and ease of use. The main drawbacks of delta-GLM models are: (1) they assume the response data is continuously distributed, but most fishery-independent data tends to be discrete count or catch data, (2) they do not fit positive catch or count data very well for species whose site-specific abundance is relatively low, and (3) they do not fit highly nonlinear relationships between the response and predictor variables very well.

Delta-GAM models are similar to delta-GLM models, with one key difference being that delta-GAM models better fit nonlinear relationships between response and predictor variables. This is useful when response observations are related to predictor variables in complex, nonlinear ways. Delta-GAM models are more challenging to fit than delta-GLM models, and the first two drawbacks of delta-GLM models described above similarly apply to delta-GAM models (Figure 5). Both delta-GLM and delta-GAM models also assume that the species of interest is never missed if truly present at a site (i.e., no "false zeroes").

Zero-inflated models assume that zero observations occur via two processes – either the species is not present at a site, or the species is present but not observed. Therefore, zero-inflated models can account for false zeroes. Furthermore, using SumCount as the response variable, zero-inflated models can account for discrete positive catch or count data using Poisson or negative binomial error distributions (Figure 6). Both of these are major advantages of zero-inflated models over delta-GLM or delta-GAM models. The downsides of zero-inflated models are that nonlinear relationships cannot be fit as easily as in a delta-GAM approach, and model selection is not very straightforward.

RECOMMENDATIONS:

- Use the zero-inflated negative binomial model to develop video indices for use in stock assessment given better model fit and better accounting of "false zeroes".
- Delta-GAM models should used as a first step to better understand the relationships between response and predictor variables and to narrow the selection of predictor variables considered by the zero-inflated models.
- Develop delta-GAM models for regularly reporting initial CPUE values for priority stocks. Refine these models and develop ZINB models for individual stocks as necessary to support stock assessments, based on the process described in this report for developing indices for Red Snapper and Gray Triggerfish.

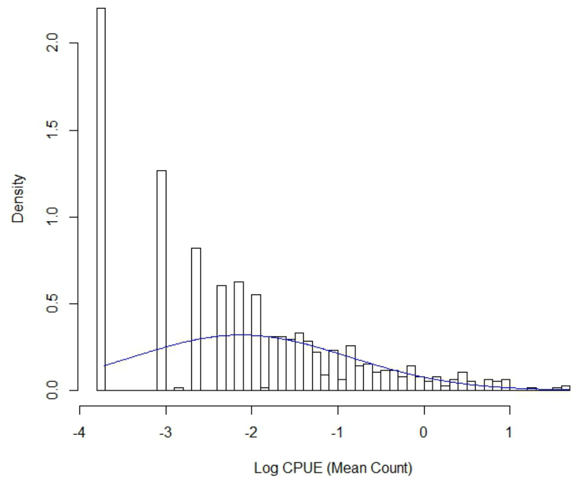

Figure 5. Model fit for the delta-GAM model, showing modeled blue line fit in relation to the distribution of Red Snapper video MeanCount data (white bars). The model fits larger MeanCount values well, but does a particularly poor job of fitting low MeanCount

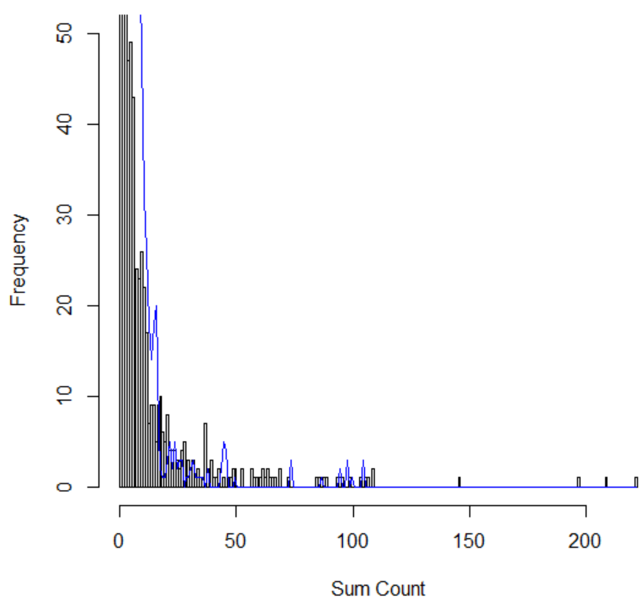

Figure 6. Model fit for the zero-inflated negative binomial model to Red Snapper video data, showing modeled blue line fit in relation to the distribution of SumCount data (black bars).

Objective 2

Provide justification for

(a) the various predictor variables included in models,

(b) use of MeanCount as a response variable,

(c) use of a calibration factor between cameras, and

(d) treatment of reconnaissance samples, and treatment of 2010 video data (which were collected using a different camera and were spatially and temporally restricted).

2 (a) the various predictor variables included in the index-development models

The panel received a thorough presentation on initial efforts to develop video index models and identify appropriate predictor variables. Many of the decisions charged to the group were overlapping. For example, the choice of predictor variables may vary with different model types as different models may handle uncertainties differently, as well as with data decisions such as whether or not to include 2010 observations. As a result, the group discussion did not always flow from one objective to the next and issues related to core decisions such as predictor variables may be mentioned throughout this report. Predictor variable selection may also be species-specific. For these reasons, specific recommendations for each stock are provided under objective 2b. The group did recognize that certain variables consistently influence survey observations and provide an appropriate starting point when developing initial indices for individual stocks.

Recommendation:

- include year, depth, latitude, water temperature, habitat, water clarity, and current direction as predictor variables in standardization models.

2 (b) MeanCount as a response variable

The panel received several references addressing potential video response variables that were considered during program design, a thorough background presentation on response variable selection, and was aware that a change in the basic response variable would require that all video samples be re-evaluated before indices could be developed. The primary variables considered were MeanCount, described above, and MinCount, commonly used in the Gulf of Mexico video surveys. MinCount is the maximum number of fish observed in a single video frame, determined by examining a 20-minute video segment.

MeanCount was used in examining the SERFS video samples based on its more linear theoretical relationship to true site abundance (Schobernd et al. 2014). It is recognized that MeanCount could result in higher variability, and as a statistic based on a number of samples (41 frame reads in this case), it has an associated measure of uncertainty. Nonetheless, the two metrics are highly related and provide observation responses from video data that are typically quite similar.

MeanCount and MinCount give similar trends and results for most species. MeanCount may underestimate positive occurrences for rare or cryptic species, due to the potential that they are present in the full video series while not appearing in one of the 41 frames read. This shortcoming could be easily resolved by reading more frames with the MeanCount variable, or developing an alternative metric for such species. MinCount values typically plateau or reach an asymptote for species of high abundance; this appears to occur at around 20 individuals in a frame. In terms of the time and costs associated with video data interpretation, MeanCount and MinCount are likely quite similar.

Recommendation:

- Support the rationale for using MeanCount, which in turn justifies the use of SumCount for methods requiring count data.

2 (c) use of a calibration factor for different video cameras

As mentioned above, GoPro cameras were used for fish counts in 2010, while Canon cameras were used in 2011-2013. To calibrate fish count between these two cameras, side-by-side Canon-GoPro videos (Figure 7) were taken during the summer of 2013 and read for Red Snapper and Gray Triggerfish. Additionally, a lab experiment was conducted to quantify differences in depth of field between the two cameras. Results indicated the GoPro cameras see a much wider angle than Canon cameras (Figure 8), but the quality of GoPro videos was perhaps lower than that of Canon videos. Videos for the comparison were selected based on Red Snapper and Gray Triggerfish occurring on videos for GoPro cameras. There is no data available to examine whether the Canon cameras could view these species when the GoPro did not, which is possible given the better depth of field and clarity of the Canon cameras.

A total of 15 calibration videos included Red Snapper and 17 videos included Gray Triggerfish. Based on a regression analysis applied to the calibration video results, there were 53% (1 - the regression slope parameter) fewer Red Snapper seen on Canon cameras compared to GoPro cameras and 39% (1 - the regression slope parameter) fewer Gray Triggerfish (Figure 9).

It may appear surprising that calibration results differ between species since a primary difference between the cameras is viewing area. However, video quality also differs, with the Canon camera typically enabling species identification at greater distances. Differences in resolution and overall video quality, combined with different responses to the presence of the cameras, can lead to calibration values that are species-specific as observed here. In the case of Gray triggerfish and Red Snapper, Red Snapper tend to stay close to the trap and video cameras, while Gray Triggerfish tend to remain further away. Therefore, it can be more difficult to identify individual gray trigger in the sampling area, but further away from the trap and camera, with the GoPro. Conversely, since Red Snapper approach the trap and camera more closely, the calibration factor is similar to the difference in viewing area.

Figure 7. Canon Vixia HF S200 (left, used in 2011-2013) and GoPro Hero (right, used in 2010)

Figure 8. GoPro (left) and Canon (right) still images from the same time and video, showing the wider angle nature of the GoPro videos.

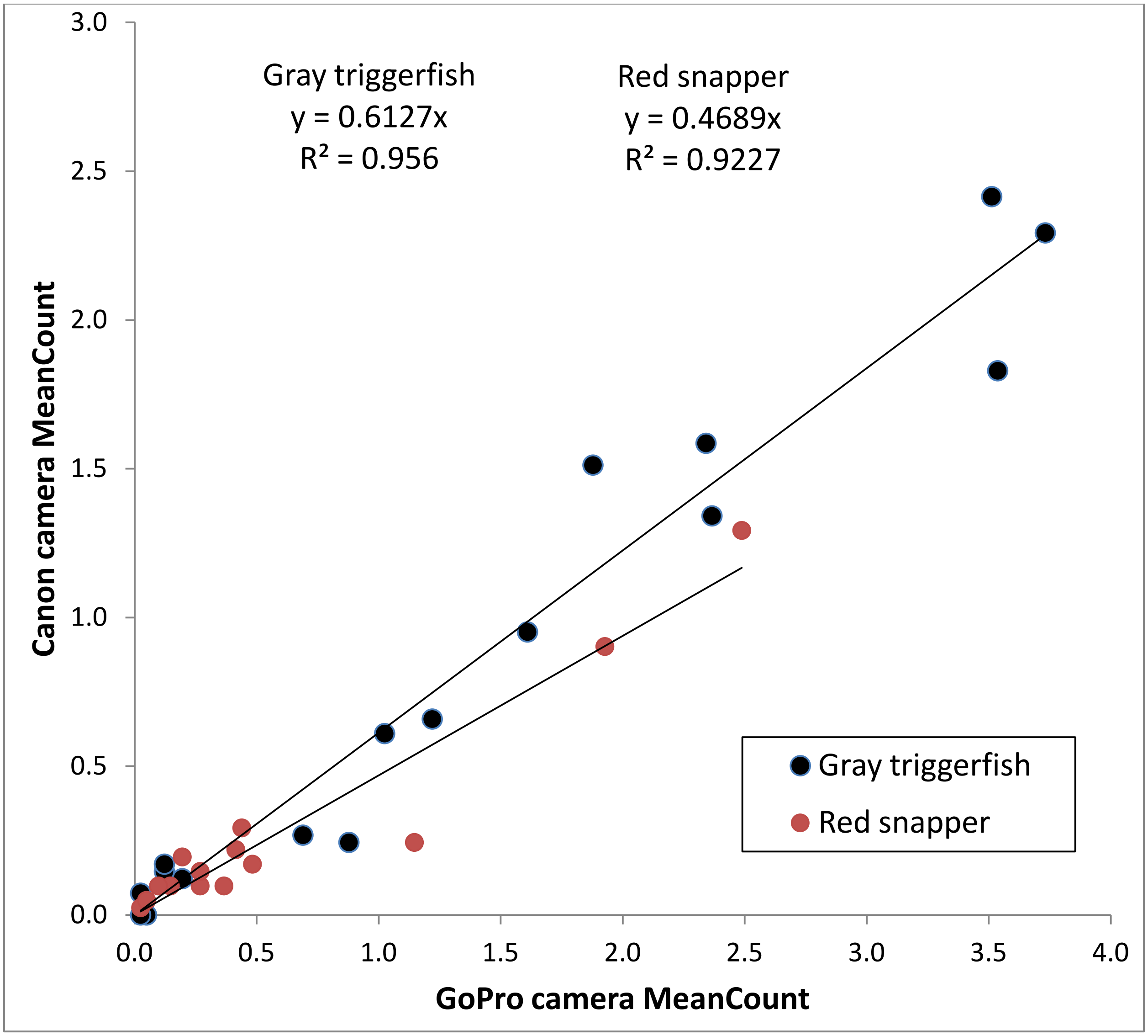

Figure 9. Red Snapper and Gray Triggerfish MeanCount values compared from two side-by-side video cameras (Canon and GoPro cameras).

Five options to calibrate GoPro and Canon counts were considered in order to adjust 2010 GoPro observations to be comparable to 2011-2013 Canon video data:

- Multiply SumCount from GoPro camera (in 2010) by the estimated calibration slope to predict the number that would have been seen by the Canon. However, this would result in partial fish being seen by the Canon camera, which is unreasonable.
- Use a bootstrap procedure where the calibration slope is used as the success probability for a binomial random variable, and the number of draws is the number of fish seen on the GoPro camera.
- Use a bootstrap procedure where the data for the number of fish seen on the Canon camera is based on a multinomial probability vector for each number of fish seen with the GoPro camera. However, as noted above videos were selected for Red Snapper presence, resulting in biased multinomial probabilities when 0 fish were seen on GoPro cameras.
- Use the above #3 multinomial bootstrap procedure, except assume Canon cameras always see 0 fish when a GoPro camera sees 0 fish. This approach was almost indistinguishable from #2 above.
- Use a post-hoc correction factor by reducing 2010 relative abundance value by a correction factor, determined from the slope of the regression comparing GoPro to Canon video counts.

The group initially narrowed the choices, focusing on options 2 and 5. Both gave similar results for Red Snapper and Gray Triggerfish. Option 5 was given a slight edge based on ease of computation and explanation. Workshop participants recommended using #5, whereby the relative abundance estimate in 2010 was reduced by 53% to account for the fact that the camera used in 2010 likely saw more fish than the Canon camera used in 2011 – 2013. It was also recommended that the standard error of the slope be included in the calibration regression in a bootstrap procedure, to better account for camera calibration uncertainty in 2010.

Recommendation:

- Apply option 5 for camera calibration.

(d) use of reconnaissance samples

The survey uses reconnaissance (recon) samples to identify potential sites for increasing geographic coverage and the universe of sampling stations. Sites having confirmed hard bottom habitat are added to the sampling universe for the following year, and may be selected for sampling just as any other site. The issue posed to the group is whether successful (containing hard bottom habitat) sites should be considered for developing indices in the current year. The number of recon sites per year is considerable during these early years of the survey, as the researchers worked to develop a large and complete sampling universe, and is expected to decrease in future years due to fewer unexplored areas remaining and limitations on total sampling effort.

The group agreed it is important to include as much data as possible and did not identify any reasons why recon samples should not be included in the year they are identified.

Recommendation:

- Include current year recon samples in analyses.

(e) treatment of 2010 data

In addition to the change in cameras after 2010, sampling occurred slightly later in the year on average than sampling in 2011 – 2013 (Figure 10). There were also many fewer samples taken in 2010 compared to subsequent years, which resulted from reduced geographic coverage. The group considered these differences and received presentations comparing 2010 to later years. While the 2010 sampling differs from other years, the additional year may prove an important point in the overall time series given that the survey is in its early years. With only 4 years available at this time, omitting 2010 will reduce the time series by 25%.

Decisions on whether to include 2010 should be reconsidered, and may become less important, in the future. The group recommended evaluating 2010 data for each species, considering factors such as sampling coverage relative to species distribution and the availability of other survey sources. The different models considered (GLM, GAM, ZINB) varied in their ability to resolve different sample levels in 2010; this likely varies by stock depending on how 'representative' the 2010 sampling is of the entire stock.

Red Snapper

To examine the impact of survey coverage, workshop participants considered models that were limited spatially and temporally to the area covered by 2010 sampling compared with those developed from the full sampling universe. Specifically, this included restricting the data to Georgia and Florida and excluding October sampling. Resulting trends were similar for the restricted and full data sets, indicating that 2010 sampling coverage was representative of the core Red Snapper population. As further support of this conclusion, the group noted that the area chosen for sampling in 2010 was specifically selected to include known areas of Red Snapper occurrence to address concerns with the Red Snapper population and the need to obtain stock information despite restrictive regulations. In addition, survey information available for the prior assessment, based on MARMAP trap sampling alone, was considered inadequate for Red Snapper.

Another approach considered was comparison of positive samples in the various trap surveys, considered as an indication of how well the 2010 samples covered the range of the species. Of 74 positive occurrences of Red Snapper in traps coast wide between North Carolina and Florida, 65 were within the area covered by the 2010 video sampling (Georgia and Florida), suggesting that 2010 sampling likely indexed core Red Snapper abundance well..

Based on the evidence considered, workshop participants recommended including calibrated 2010 video data in the Red Snapper analyses because of its importance in the relatively short time series of video data and the need to use all the available data given the restrictive fishery regulations. The available 2010 video samples occurred in the center of the range of Red Snapper on the East Coast (Florida and Georgia), and is believed to provide reliable information on relative abundance for Red Snapper in the region.

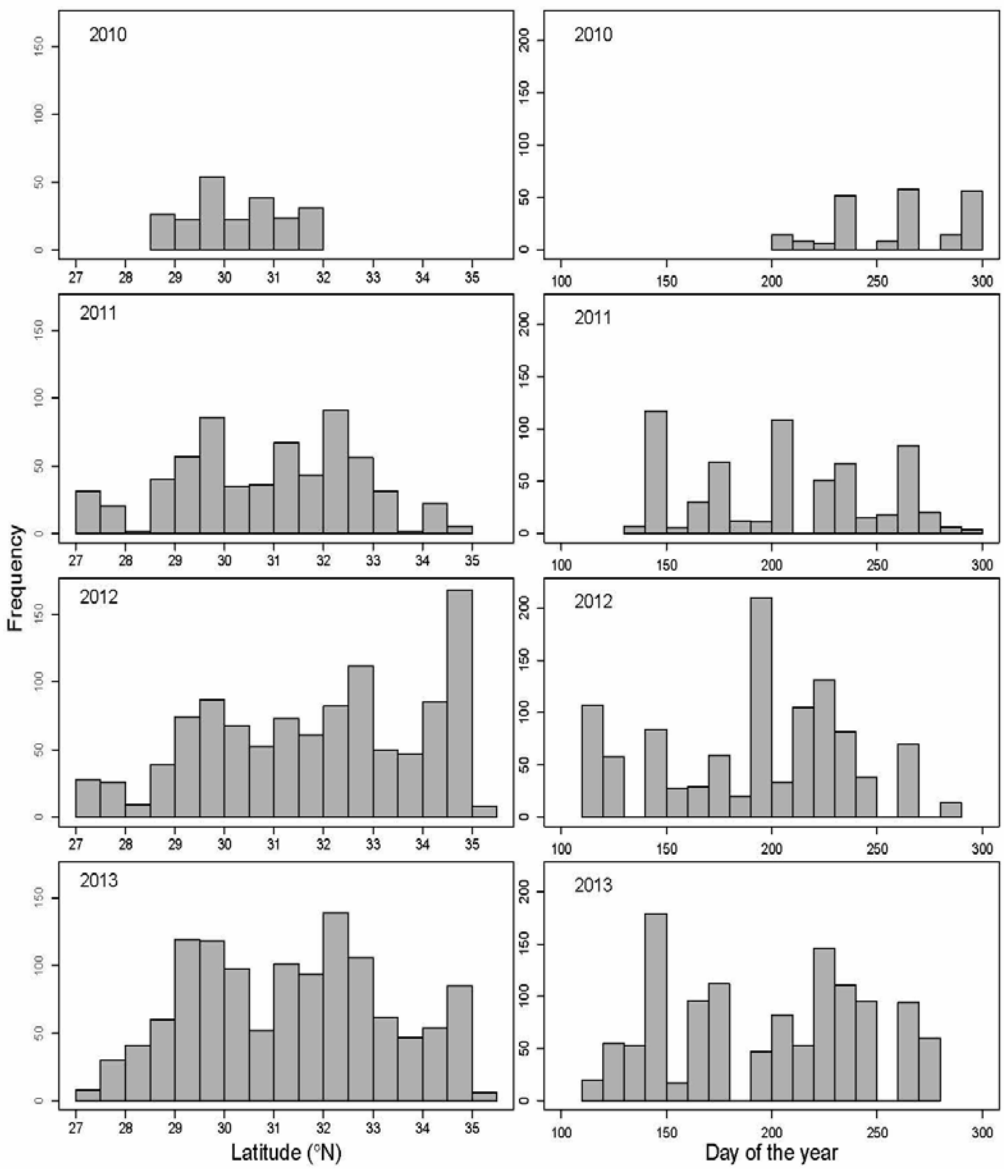

Figure 10. Histograms of usable SERFS video samples across latitude (left column) and day of the year (right column) in 2010 – 2013.

Gray Triggerfish

The group considered Gray Triggerfish using analyses similar to those prepared for Red Snapper. Gray Triggerfish are not subject to recent restrictive management changes. Comparing full trap sampling to the area of limited 2010 coverage suggested that Gray Triggerfish have a much broader range than red snapper. Trends in relative abundance in the reduced area were not consistent with those for the full survey area. Therefore, workshop participants concluded that 2010 samples did not occur in the center of Gray Triggerfish distribution on the East Coast and recommended excluding 2010 video data for Gray Triggerfish.

Recommendation:

- Examine 2010 data on a species by species basis.
- Include 2010 video data for Red Snapper.
- Exclude 2010 video data for Gray Triggerfish.

Objective 3:

Review the analytical approaches used to develop video-based indices of abundance and recommend the most appropriate technique.

The group reviewed a range of model configurations for the three model types considered (GLM, GAM, ZINB). Discussion centered around specific considerations such as model type selection, predictor variables, predictor variable data types, error modeling, and interactions. It was generally recognized that the ZINB performed best for providing index values, while the delta-GAM models were most efficient for exploring predictor variables and how they should be handled within the model. Further model explorations included interaction between depth and latitude, retention of 2010 data, polynomial error structures, number of categories for the season and latitude variables, and trap catch as a covariate.

Red Snapper and Gray Triggerfish Recommendations:

- Use the ZINB model to develop indices for Red Snapper and Gray Triggerfish.
- Include year, depth, latitude, water temperature, habitat, water clarity, and current direction as predictor variables in standardization models.
- Use 8 levels of latitude and season (i.e., “octiles”) in the zero-inflated model for Red Snapper to better account for the nonlinear relationships suggested by the delta-GAM models.
- Use 8 levels for season and 4 levels for latitude in the ZINB for Gray Triggerfish, since relationships between Gray Triggerfish and latitude were not as obviously nonlinear.

Other species recommendations:

- Index development for other species should consider, but not be limited to, the methods used for Red Snapper and Gray Triggerfish. In general, an approach of using exploratory data analysis (e.g. GAM models, multivariate analyses, visualization methods) to inform statistical distribution and predictor variable selection for subsequent GLM-type modeling is encouraged.

Research Recommendations

- Recommend that video index CPUE based on initial GAM modeling and the predictor variables suggested here be developed, updated annually and results reported in the annual SERFS monitoring report.
- Explore a paired index that uses the trap and video indices to inform each other.
- Recommend further examination of gear selectivity, to determine whether video and trap data have different selectivities and, if so, examine impact this could have on size and age composition estimates.