CIE Independent Peer Review Report

on

SEDAR 22 Gulf of Mexico Yellowedge Grouper and Tilefish Data Workshop

Prepared by

Yong Chen

Professor of Fisheries Population Dynamics
School of Marine Sciences
University of Maine
Orono, ME 04469

Contents

I. Executive Summary

The SEDAR 22 Data Workshop (DW), held in Tampa, Florida from March 15-19, 2010, was the first step in the assessment of two important fish species, Yellowedge grouper (*Epinephelus flavolimbatus*) and tilefish, *Lopholatilus chamaeleonticeps*, in the Gulf of Mexico (GOM). It was aimed to identify, evaluate and compile information on their life history parameters, abundance indices, and commercial and recreational catch based on fishery-dependent and fishery-independent data sources.

I would like to commend the great efforts of all the participating scientists, managers and fishermen in the SEDAR 22 DW in the identification, evaluation and compilation of the information on life history, fishery-dependent and fishery-independent abundance indices, and landings in the commercial and recreational fisheries for yellowedge grouper and tilefish (i.e., golden tilefish), and blueline tilefish in the GOM. I am impressed by the breadth of expertise and experience on the panel, openness of discussion for considering alternative approaches/suggestions, and constructive dialogs in each working group and at the plenary meetings throughout the workshop. All the comments, whether they were from scientists, managers, or fishermen, were fully considered and discussed. In particular, I commend SEDAR's inclusion in the Data Workshop of fishermen who provided insights on the quality and quantity of the fishery data, in particular for historical fisheries data. I observed on many occasions constructive interactions and dialogs between scientists/mangers and representatives of the industry in the Workshop.

In general, I consider the information identified and compiled to represent the best efforts given all the limitations. I consider that the approaches used in developing life history parameters, fisheries landings, and abundance indices are sound. However, there are still some issues I believe that need to be addressed to improve the processes and outputs of the DW. In general, I believe that all the DW working papers should be made available prior to the DW, rather than during the DW. The scientific names for the targeted fish species should be listed to avoid confusion. The time period covered by the stock assessment should be determined prior to the DW or determined during the DW (in which case this task should be included in the TORs). Because the quality and quantity of data tend to vary greatly among different stock assessment models, I recommend that a data workshop start with an introduction of potential models that may be used in stock assessment so that all the DW panelists understand the key requirements of the stock assessment model, which can help them in identifying and compiling the stock assessment data.

My specific recommendations/comments on the data analysis and compilation of this DW include (1) comparative studies of data from different monitoring programs to evaluate the coherence of the programs in quantifying the dynamics of the GOM yellowedge grouper and tilefish stocks; (2) comparative studies of data from WGOM (West GOM) and EGOM (East GOM) to identify possible spatial variability and evaluate needs for a separate analysis for the WGOM and EGOM for the two species; (3) estimation of uncertainty for all the key life history parameters for the development of different scenarios in stock assessment; (4) development of scenarios for evaluating impacts of using different Spawning Stock Biomass (SSB) values on stock assessment; (5) evaluation of reliability of different abundance indices and determination of their relative importance in stock assessment; (6) evaluation of reliability of estimated landing data and development of plausible scenarios for sensitivity analysis of potential errors in estimated landings for stock assessment; (7) adding key habitat variables (e.g., depth and sediment) and use of general additive models (GAM) in CPUE and abundance index standardization; (8) an evaluation of sampled length compositions with respect to their representation of population size compositions; (9) development of age-length keys to convert length composition data to age composition data; and (10) a systematic evaluation of effectiveness of sampling program for commercial catch. More specific comments and recommendations can be found in Sections IV and V of this report.

II. Background,

Quantitative fisheries assessment models, based on biological theories and empirical observations, are defined by parameters that characterize the population dynamics of the stocks (Quinn and Deriso 1999). Reliable estimation of these parameters is a central issue in fisheries stock assessment and management (Chen and Paloheimo 1998; Walters 1998). Typically, parameters are estimated by fitting stock assessment models to data collected from the studied fish stock (Hilborn and Walters 1992). The quality of parameter estimation, which determines the quality of stock assessment, can be affected by many factors (NRC 1997, 1999; Chen and Fournier 1999). Two of the most important factors are the quality and quantity of stock assessment data.

The SEDAR 22 Data Workshop (DW), held in Tampa, Florida from March 15-19, 2010, was the first step in the assessment of two important fish species, Yellowedge grouper (*Epinephelus flavolimbatus*) and tilefish, *Lopholatilus chamaeleonticeps*, in the Gulf of Mexico (GOM). The purpose was to identify, evaluate and compile the information on their life history parameters, abundance indices, and commercial and recreational catch based on data from fishery-dependent and fishery-independent sources. The quality and quantity of the data were evaluated for the assessment of these two fisheries in the DW.

The yellowedge grouper (YG) are present from North Carolina to South Florida (Huntsman, 1976), and throughout the GOM. They are typically distributed from 90 to 365 meters in various habitats. Juvenile YG are usually found in inshore shallow waters and the adult population in deeper waters in the GOM. YG are protogynous, beginning life as females and transitioning into males with increased age/size. YG supports an important commercial fishery and a small scale recreational fishery in the GOM. The fishery, which uses longlines and vertical lines (droplines), started in the 1970s, and landings have fluctuated since the 1980s with the majority of recent landings from the longline fishery.

The tilefish (often referred to as golden tilefish, but here the official American Fisheries Society name is used) is also a deep-water fish ranging from Nova Scotia to the Gulf of Mexico (Dooley 1978). They can reach maximum ages of up to 40 years, grow slowly, exhibit sexually dimorphic growth with males having large sizes (Grimes and Turner 1999), and mature at relatively large size and old age (Palmer et al. 2004). The tilefish have a unique burrowing behavior and strong habitat preference (Able et al. 1982). Although not conclusive, evidence suggests that tilefish tend to be protogynous.

Various fishery-dependent and fishery-independent monitoring programs have been developed to collect data for characterizing the GOM YG and tilefish populations and fisheries. The fishery-independent monitoring programs include the SEAMAP groundfish bottom trawl survey (for YG) and NMFS bottom longline survey (for both YG and tilefish), and the fishery-dependent program includes port sampling and longline logbooks (for both YG and tilefish).

The first comprehensive assessment for the YG stock in the GOM was conducted in 2002 (Cass-Calay and Bahnick, 2002). No formal stock assessment has been done prior to SEDAR 22 for the GOM tilefish stocks. The SEDAR 22 DW represents a major effort to identify, evaluate and compile all the information, necessary for the upcoming assessment of the GOM YG and tilefish stocks.

This report includes an executive summary (Section I), a background introduction (Section II), a description of my role in the review activities (Section III), my comments on each item listed in the Term of References (TORs, Section IV), summary of my comments and recommendations (Section V), and references (Section VI). The final part of this report (Section VII) includes a collection of appendices including the Statement of Work (SoW), a list of participants and their roles in the DW, and the schedule of the DW.

III. Description of the Individual Reviewer's Role in the Review Activities

As stated in the SoW, "SEDAR 22 will be a compilation of data, a benchmark assessment of the stock, and an assessment review conducted for Gulf of Mexico yellowedge grouper and tilefish under the SEDAR (Southeast Data, Assessment and Review) process." My role in the DW was to "participate in the Data Workshop and to provide impartial and independent reviews on the process and products of data compilation for the assessment of the GOM yellowedge grouper and tilefish stocks."

Prior to the DW, I attended the online SEDAR 22 Data Scoping webinar held from 11 am – 1 pm, February 19, 2010, and received some background information about these two fish species and their fisheries in the GOM. A small number of working papers was also made available to me prior to the Workshop.

Three separate working groups were formed in the SEDAR 22 DW: Life History Group (LHG), Indices Group (IG), and Landing Statistics Group (LSG) (see the participant name list and their roles in Appendix VII-3). The LHG is responsible for evaluating and compiling key life history parameters including age, growth, maturation, and spawning/fecundity data; the IG is responsible for developing abundance indices from both fishery-dependent and independent sources; and the LSG's role is to develop a time series of catch statistics for both commercial and recreational fisheries. Three species were considered by each working group: yellowedge grouper, golden tilefish, and blueline tilefish (*Caulolatilus microps*). The three working groups had meetings concurrently every day, and convened two times every day to report their main findings and concerns at plenary meetings for comments and recommendations from panelists of all three groups. The DW concluded that data for blueline tilefish were deemed to be inadequate for its stock assessment. This report's reference to tilefish regards only the golden tilefish, the official common name according to the American Fisheries Society.

I divided my time each day among the three working groups and attended all the plenary meetings. I was actively involved in the discussion by (1) questioning and asking for clarifications on monitoring/sampling program design, data collection, statistical analysis, interpretations; (2) making observations of the process; and (3) making comments and suggestions for alternative approaches and more analyses. However, because I could not attend all the three working groups at the same time, I might have missed some important discussions within each group. However, I did have chances to cover these issues at the plenary meetings.

IV. Summary of Findings

My detailed comments on each item of TORs are detailed under the respective subtitles of TORs (see below).

IV-1. Gulf of Mexico Yellowedge Grouper Data Workshop

IV-1-1. Characterize stock structure and develop a unit stock definition. Provide maps of species and stock distribution.

There were some discussions about this TOR in the Workshop, but no evidence to support or reject the unit stock definition.

The YG population in the GOM was implicitly considered a unit stock in the DW, and was assumed to be distinct from those in the Atlantic. However, limited information is available to support/oppose this unit stock definition. The spatial and temporal coverage of fishery-dependent monitoring programs and fishery-independent survey programs was limited, and a small number of samples were taken in the fishery and surveys. Although size-dependent inshore-offshore movement is evident for YG, there is limited understanding of spatial dynamics of YG in the GOM and information on spatial variability in key life history parameters and genetic work is also limited, making it difficult to evaluate if local populations are present.

Tagging studies are almost impossible for this species, and cannot be used in identifying spatial dynamics of this stock. A comparison of size distributions between the SEAMAP groundfish bottom trawl survey, which covers coastal waters of GOM, and the NMFS bottom longline survey, which covers deeper waters, suggests that small YG are mainly found in shallow waters and large YG are mainly found in deep waters. The Δ14C analysis of otoliths collected in deep waters suggests that YG inhabit inshore waters when they are young, indicating that YG experience size-dependent inshore-offshore distribution (Cook et al, 2009). However, no evidence suggests that YG experience any long-distance migration associated with their life history processes (e.g., spawning, feeding migrations). Instead, evidence suggests that YG have limited spatial movement, which raises the possibility of existence of local populations in the GOM. I recommend that more studies (e.g., genetic study, spatial variability in life history) be conducted to improve understanding of spatial dynamics and variability of the YG stock in the GOM and the spatial scales on which the key life history processes of the GOM YG operate. This could provide suitable guidance in determining the spatial scale of the YG stock assessment and management.

Maps were provided for the YG stock distribution on the US side of the GOM, but the distribution of YG in the entire GOM was not presented, which may have resulted from lack of knowledge of YG on the Mexico side of the GOM. I suggest contact with relevant Mexican agencies to have a better understanding of the YG fishery on the Mexico side of the GOM, which could provide information for the GOM-wide YG distribution.

IV-1-2(a). Review, discuss and tabulate available life history information (e.g., age, growth, natural mortality, reproductive characteristics);

The Data Workshop had extensive and excellent discussion about this TOR. The key life history parameters needed for the stock assessment were estimated, and the uncertainty essential for defining priors and possible sensitivity analyses were provided for most parameters that will be used in the stock assessments.

Yellowedge grouper sagittal otoliths were used for ageing using conventional means and radiocarbon (^{14}C) measurement. Radiocarbon values of YG otoliths were compared to established radiocarbon chronologies in the region to validate the age and ageing methodology. In addition, eight specimens were analyzed for $^{\Delta 14}C$ to validate age estimates for fish born prior to the ^{14}C increase (Cook et al. 2009). This approach suggests maximum longevity of 85 years in the GOM. However, most samples were from EGOM (64%), in which small and young YG were more abundant. Only 36% samples were taken from WGOM in which YG tended to be bigger and older. More samples need to be collected from the WGOM.

I commend the LHG efforts to go through an exhaustive list of methods for estimating natural mortality rates. In addition, the LHG also analyzed historical age composition data that were collected prior to the fishery and estimated natural mortality using the catch-curve method (Ricker 1975). This effort yielded a wide range of estimates of natural mortality, which helped quantify uncertainty associated with M estimates. However, I believe that the LHG did not pay enough attention to the assumptions and limits associated with each method for the M estimation. I suggest an evaluation of the key assumptions.

YG shows clear characteristics of protogyny, changing from females to males with increased age/length. The contrast in maturity and sex transition by age/size is well defined, resulting in good estimates of age/size at maturity for females and age/size at sex transition. YG are batch spawners, and quantification of the total egg production over a prolonged spawning season is difficult. SSB has been identified as a reliable proxy for the YG reproductive potential. However, two measures can be used to index SSB: female SSB and sex-combined SSB. Because males transitioning from females do not make contributions to egg production and there is no evidence of sperm limitation, the female SSB may be a better measure for the reproductive potential of the YG population in the GOM.

IV-1-2(b) provide appropriate models to describe growth, maturation, and fecundity by age, sex, or length as applicable.

The Von Bertalanffy growth function (VBGF) was used to quantify the relationship between the total length and age for YG. A logistic equation was used to describe how the proportion of mature YP changes with length/age. The GOM YG is well known for its protogyny with good contrasts for maturity and transition from females to males with increased size/age. The LHG quantified how the likelihood of female maturity changes with size/age and how the transition of females to males is related to size/age. The growth model was estimated for female YG and for sex-combined YG. The sex-combined VBGF was used to describe the growth for males because males grow from females.

1979-2009 age-length data are available from fishery-dependent and independent programs, but data from the fishery-dependent program are patchy prior to 1999. Sample size becomes large after 1999. Possible temporal variability in growth was not evaluated systematically in the DW. However, based on the data availability, the data were broken into three time periods that showed potential temporal variations in growth, indicating that such temporal variability should be considered in transferring length composition into age composition using age-length key or other methods. The spatial coverage of samples used for the ageing study is limited. Spatial variability in growth is not evaluated in the DW. I recommend subsampling the growth data (based on their sampled locations and time) to evaluate possible spatial and temporal variability in growth modeling.

Growth data were collected from different monitoring programs, which cover different areas and have different selectivities. I recommend evaluating differences in growth models derived from different monitoring programs.

The natural mortality was estimated from all the samples, sex combined. I recommend evaluating possible changes in M after females change to males.

For modeling size/age-specific life history processes, no statistical variability was estimated in the key life history parameters. A classic method or bootstrap can be used to estimate confidence intervals for all the life history parameters in modeling.

I commend the LHG efforts in analyzing the 1977-1985 length-age data (Bullock et al. 1996) and in comparing the results with data collected more recently in estimating natural mortality. I recommend that a comparative study be done to evaluate if there is a significant difference in growth between the two time periods. Such a comparative study may reveal possible temporal changes in growth and help identify potential impacts of fishing on life history. However, we need to remember that most samples in Bullock et al (1996) were data from Florida.

IV-1-2(c). Evaluate the adequacy of available life-history information for conducting stock assessments and recommend life history information for use in population modeling.

The Workshop yielded the life history information needed for the YG stock assessment. Although the quantity is sufficient, caution should be applied in using the life history information. My major concern is that samples should be representative with respect to the YG population in the GOM, but there is limited understanding of temporal and spatial variability in the key life history parameters. I recommend that an age-length key be developed for the GOM YG to transfer size composition to age composition.

The derived size-dependent growth, maturation, and sex transition provide the information necessary for the estimation of SSB in population modeling. However, limited spatial coverage (weak data on the WGOM) and small sample sizes in some cases may make the representations of the estimated parameters questionable. Uncertainty resulting from measurement and process errors in association with these key life history parameters should be considered in population modeling. Uncertainty needs to be estimated for key parameters so that some informative priors can be provided to the assessment. The uncertainty can also be used to develop sensitivity analysis scenarios. I commend the LHG for their efforts on quantifying the key life history parameters and their uncertainties.

IV-1-3. Provide measures of population abundance that are appropriate for stock assessment.

This TOR was extensively discussed in the Data Workshop. Quantity and quality of the data were discussed.

The IG estimated abundance indices for YG from both fishery-dependent and independent data sources. Fishery-independent survey programs include (1) NMFS bottom longline surveys (2000-2009) and (2) SEAMAP groundfish bottom trawl survey (1972 – 2008); and the fishery-dependent program includes longline logbooks (1992 – 2009).

SEAMAP groundfish bottom trawl survey

SEAMAP groundfish bottom trawl survey has abundance index data from 1972-2008 for YG, but the data prior to 1987 are considered to be different from the data after 1987 because of changes in survey spatial coverage and tow duration. The survey did not cover Florida coastal waters prior to 1987 although this is the area where the majority of the YG fishery occurred. The survey is not designed to target YG but to catch as many demersal species as possible. When the YG catch was over 22.5 kg, they were subsampled, otherwise all catch was measured. The bottom trawl survey gear caught mainly small grouper, and it is possible that YG may be mis-identified as other grouper species, and other grouper may be mis-identified as YG. The YG sampled in the survey are small with ages ranging from 0- 3 years old.

This survey program pre-dates the YG fishery in the GOM. The abundance index may be considered as a recruitment index. However, because of lack of coverage of EGOM by this program, this requires an assumption that the recruitment process was similar between the EGOM and WGOM and small fish are assumed to move to offshore from inshore rather than die. The latter assumption seems to be valid as suggested by otolith analysis (Cook et al. 2009).

The major concern with this index is the extremely small number of YG caught in the survey. Over the survey time period, less than 5 YG were caught per year on average over the large survey area in the GOM. Caution should be applied for using such extremely small sample sizes in indexing the juvenile YG population dynamics in population modeling.

A two-stage modeling approach (Lo et al. 1992) was used in standardizing this abundance index, which used a binomial distribution function to model the proportion of positive trips per stratum and a general linear model (GLM) to quantify impacts of fixed factors on log(abundance index). However, some factors that are important in influencing the spatial distribution of YG were not included in the standardization. I recommend that sediment and depth be included in the standardization. The use of GLMs in the standardization implicitly assumes that the effects of environmental variables on the response variable (i.e., fish distribution) are linear, which may not be realistic. A more general model such as a general additive model (GAM), which can incorporate both linear and nonlinear interactions, may be more suitable for the standardization.

NMFS fishery-independent bottom longline survey

The annual NMFS fishery-independent bottom longline survey in the GOM has been conducted by the Southeast Fisheries Science Center (SEFSC) Mississippi Laboratories since 1995, and is designed to provide fishery-independent survey data for the assessment of as many species as possible. For SEDAR 22, only data from 2000-2009 were considered because of the shallow depths surveyed and the use prior to 2000 of J-type hooks, which had very low catch efficiency for grouper species. Circle-hooks have been used since 2000, which significantly increase catch of grouper species. Frozen Atlantic mackerel were used as bait and soak time was standardized at one hour. However, the time duration between the first hook in the water and last hook out of the water varied from 1.5-3 hours, which made soak time vary among hooks. This might introduce some minor biases.

The survey covers three depth strata, including the YG depth range (i.e., 70 – 365 m). Survey stations where YG were captured were plotted. However, spatial coverage of the survey program varied during the time series due to weather or mechanical problems. Only data from survey stations within the depth range of capture for YG were used in development of abundance indices. Year 2005 was excluded in the analysis because of Hurricane Katrina.

The delta-lognormal abundance index was developed using an approach described by Lo *et al.* (1992) in the index standardization. Various statistical procedures were employed (e.g., Q-Q plot and plot of residuals against year) for checking the validation of the distribution assumption in the standardization modeling. I commend the IG efforts to standardize the abundance index and consider the general approach sound. However, I notice that sediment which is critically important in influencing the YG spatial distribution was not included in the standardization. The use of GLM also implicitly assumes that the effect of spatial, temporal and environmental variables on the response variable (i.e., survey abundance) are log-linear, thus limiting other forms of interactions. I recommend that GAM be considered in the abundance index standardization.

Commercial bottom longline catch per unit of effort

Commercial catch per unit of effort (CPUE) data were derived from the Reef Fish Logbook (RFL) program which requires all vessels with reef fish permits to report their catch. A two-stage modeling approach (Lo et al. 1992) was used in standardizing CPUE, which uses a binomial distribution function to model the proportion of positive trips per stratum and a general linear model (GLM) to quantify impacts of fixed factors on log(CPUE). This was done separately for EGOM longline, EGOM handline, and WGOM longline. The general approach is sound, and I commend the IG efforts to remove impacts of environmental factors on CPUE.

I recommend that the GAM be considered because it is more flexible in incorporating possible nonlinear interactions between the YG abundance and fixed factors considered in the standardization. The depth and sediment were not included in the standardization, although it is a key variable influencing the spatial distribution of YG in the GOM. I suggest (1) using GAM, (2) including depth and sediment in the standardization, (3) testing model performance with cross-validation. However, I realize that measures for depth and sediment may not correspond to the catch because landings in this fishery were reported based on trip, not set.

For standardization of all the three abundance indices, the IG ran through a long checklist, which was developed in the NOAA Indices Workshop, for evaluating the quality of the abundance index standardization. I commend the IG efforts.

IV-1-3(a). Consider and discuss all available and relevant fishery dependent and independent data sources.

This TOR was discussed and addressed in the DW.

Abundance indices derived from the two fishery-independent and a fishery-dependent programs were recommended in the DW for quantifying the stock dynamics of YG in the GOM. These programs differ greatly in their designs, sampling gears, and spatial and temporal coverage. This suggests that these sampling programs target different components of YG population in the GOM, and have different selectivities and sampling efficiencies. The use of all three different abundance indices can complement each other. The IG provides a good description of these three programs (see my description above).

Fisheries CPUE may not be a good abundance indicator (need to understand spatial dynamics of the fishing fleet, spatial variability in CPUE and how it changes over time). Because commercial catch is reported per trip, not by set, the area where catch occurs may not be the same as the area where reported catch occurs. This bias in spatial distribution of catch may introduce potential errors in the index standardization.

The SEAMAP bottom trawl survey covers long time series, but not the major depth range of YG. It yields very small YG catch per year (< 5 YG per year on average). I recommend caution for using such a small catch as abundance index for the YG population in the GOM. The changes in sampling protocol (e.g., tow duration and spatial coverage) in 1987 warrant a separate analysis of the two time series of data or incorporation of a categorical variable in the standardization to indicate such changes. I also recommend that a separate catchability coefficient be assigned to these two time periods in assessment modeling.

The NMFS bottom longline survey covers all depth ranges of YG, but the time series only cover from 2000-2009 with 2005 being removed because of Hurricane Katrina, and no index can be reliably derived for years prior to 2000.

The IG discussed issues associated with each of these three abundance indices within the group and in the plenary meetings. However, I did not observe the discussion about the relative importance of the three abundance indices. Thus, there is no assignment of relative weights for these three abundance indices. I recommend that the IG do some comparative analyses of these three abundance indices (e.g., an analysis of the three abundance indices for their coherence over time and over space).

IV-1-3(b). Document all programs evaluated, addressing program objectives, methods, coverage, sampling intensity, and other relevant characteristics.

The two fishery-independent and one fishery-dependent programs were well described for their objectives, methods, spatial and temporal coverage, sampling intensity, fishing gears, and development over the survey time period. This helped one to understand differences and similarities among these programs.

IV-1-3(c). Provide maps of survey coverage.

Various maps were provided to document spatial distributions of survey stations with and without YG catch. However, sediment and bathymetric information was often not shown on such a map. I suggest that all maps of survey coverage should include such information. A year-to-year plot of survey coverage together with sediment and bathymetric information can improve our understanding of annual variability in spatial coverage of key habitats of the GOM YG in the survey.

IV-1-3(d). Develop CPUE and index values by appropriate strata (e.g., age, size, area, and fishery); provide measures of precision and accuracy.

The nominal and standardized abundance indices and fishery CPUE were derived for the GOM YG. No abundance index and CPUE were estimated by age, size, and area. This might have resulted from small YG catch in the surveys. CPUE was estimated for the GOM longline fishery, but not for vertical lines and recreational fisheries because of lack of reliable catch and effort data in these fisheries. Coefficients of variation (CV) were estimated for the standardized CPUEs and abundance indices. No measure of accuracy was provided.

IV-1-3(e). Evaluate the degree to which available indices adequately represent fishery and population conditions.

The SEAMAP trawl survey program does not cover major habitats for adult YG and has very limited catch per year, suggesting that the indices derived from this program may not adequately represent population condition. The NMFS bottom longline survey does cover major habitats for adult YG, but its spatial coverage between the EGOM and WGOM varies from year to year. Given the likely difference in population structure and large differences in oceanographic dynamics between the EGOM and WGOM, the quality of the indices derived from this program may be questionable in representing the overall population in the GOM. For the longline fishery CPUE, because fishermen report catch in catch per trip, not catch per set, CPUE may not correspond well with the effort, which suggests that CPUE may not be a reliable indicator of fish abundance.

Inconsistency could be seen among different programs, which can lead to reduced confidence in the stock status assessment, particularly in light of the varying spatial coverage and the lack of information about whether these surveys represent distinct populations or samples from a common population. Differences in spatial coverage among the monitoring programs and the fact that no fishery-independent survey program is specifically designed for sampling the GOM YG may also be two issues that need to be considered in using these indices in stock assessments. I encourage the continued evaluation of fishery-independent and fishery-dependent data quality and development of different scenarios for evaluating their impacts on stock assessments.

IV-1-3(f). Recommend which data sources are considered adequate and reliable for use in assessment modeling.

The DW recommended that abundance indices derived from the two fishery-independent survey programs (SEAMAP groundfish survey and NMFS bottom longline survey) and one fishery-dependent program (longline logbook) be used in assessment modeling. I concur with the recommendations, but also have the following suggestions.

Fishery-independent survey programs are not designed for capturing YG, which is likely to lead to large variability in survey catchability (and vulnerability) and may in turn result in biased representation of population structure. I recommend that data from different survey programs with similar spatial coverage be compared with key population variables/parameters (e.g., abundance, sex ratio, and size composition). Such a comparative study may provide some information on differences in sampling catchability. For each program, I also suggest that an evaluation of among-station variability be conducted, which may provide some insights about the quality of the data (i.e., highly aggregated or over a wide spatial range).

The numerous surveys were presented with equal weight even though some must give a better indication of YG abundance than others. The multiple surveys for YG sometimes give conflicting, or at least inconsistent signals, likely due to survey differences in gear, location, timing, and spatial extent. I recommend that the programs be ranked according to their relevance to YG spatial distributions and sampling efficiency. The ranking can then be considered in the assessment.

IV-1-4(a). Characterize commercial and recreational catch, including both landings and discards, in pounds and number.

This TOR was discussed extensively in the Workshop.

Fishery-dependent data were discussed in characterizing catch in commercial and recreational YG fisheries. Commercial landings of YG were estimated for 1983-85 after extensive consultation with fishermen. Discards were considered very small, and could be neglected in the stock assessment. The working group considered that it was almost impossible to reliably quantify recreational catch over the spatial and temporal range of the targeted YG population in the GOM. Catch was measured in weight (lbs), and no catches in numbers were estimated.

It is clear to me from the discussion in the DW that the quality of commercial and recreational catch is questionable. However, I believe that the SS3 assessment program considers that the catch statistics have no errors. This assumption is certainly violated for YG. I recommend that different scenarios of catch be considered for sensitivity analyses to evaluate impacts of uncertainty in catch estimates on the stock assessment.

IV-1-4(b). Provide estimates of discard mortality rates by fishery and other strata as appropriate or feasible.

Limited data shown in the DW suggest that discards were very low in the YG fishery in the GOM (low discards only occurred in the closed season). This was confirmed by all the fishermen present at the DW. The reasons behind this include no size limit regulation and no incentive for high grading. However, with the implementation of trip limits on Jan. 1, 2010, which may provide incentives for high grading, discarding may become an issue. Monitoring programs need to be developed to collect the relevant data for future assessments.

Because YG is a deep water species, most discarded YG are assumed to die.

There was no discussion about discards in the recreational YG fishery.

IV-1-4(c). Evaluate and discuss the adequacy of available data for accurately characterizing harvest and discard by species and fishery sector.

Overall, the harvest data reflect temporal variability in commercial landing in the GOM. However, I heard repeatedly in the DW that commercial landing statistics might have large uncertainty, especially in early years (the 1980s) because of species misidentification and errors in reporting. Some unrealistic temporal patterns were observed in the DW, especially for the vertical line catches in the 1980s. There are also three years (1983-85) for which the landing data are not available. These three years are the time period when fishermen moved back from offshore to inshore for YG. There were some discussions with the involvement of fishermen about how unrealistic vertical line catch data and missing landing data should be dealt with in the DW. Although consensus was reached at the end, I believe that the data compiled for these years by gears represent only a best guess.

Species misidentification may be another issue that may affect the quality of landings data. This may include YG being mis-identified or mislabeled as other species (e.g., yellowfin grouper) and other grouper species being mis-identified as YG. Such impacts may differ among years because a preliminary study suggests that species misidentification rates changed over time, tending to be severe in early years but less so in more recent years. This issue was discussed extensively, but not solved in the DW. A working group was formed to further discuss this issue after the DW.

There was limited discussion about the quality and quantity of recreational fishery data.

IV-1-4(d). Provide length and age distributions if feasible.

There was limited discussion about the quality of estimated length distribution for landed catch. Given the large spatial distribution of the fishery and landing ports, the quality of length distribution should have been discussed more extensively in the DW. This is especially true for length composition data in earlier years when port sampling efforts were limited. There were also large differences in port sampling efforts among different states in the GOM, which might result in different quality of length composition data estimated for the GOM.

There was also limited discussion on the estimation of age composition for the commercial landings. No sex ratio was estimated for catch data.

I recommend more discussion and evaluation of length composition estimates and development of alternative scenarios for different approaches used in deriving length composition estimates.

IV-1-4(e). Provide maps of fishery effort and harvest.

There was no map for the spatial distribution of fishing effort and landed catch. Because fishermen reported their catch by trip, which might last a few days, not by set, reliable spatial mapping of fishing efforts and catch is difficult. Changes in reporting from catch per trip to catch per set are recommended.

IV-1-5. Provide recommendations for future research in areas such as sampling, fishery monitoring, and stock assessment. Include specific guidance on sampling intensity (number of samples including age and length structures) and appropriate strata and coverage.

The DW had some discussion about this TOR, but the discussion was not extensive enough to lead to concrete recommendations for future research in the areas identified in this TOR.

WGOM and EGOM differ greatly in their oceanographic conditions, and there is a need to evaluate if there is a difference between the two areas in the key life history parameters. I suggest that spatial variability be evaluated in key life history processes to identify population spatial dynamics, and possible existence of local stocks.

Sampling programs for quantifying size composition and age composition of commercial catch need to be carefully evaluated. Factors such as adequate spatial and temporal coverage and sampling intensity to have high effective sample sizes should be considered. I recommend developing alternative sampling designs, developing a simulated fishery that mimics temporal and spatial variability in size and age compositions in commercial landing, applying current and alternative sampling programs to the simulated fishery, comparing the performance of the sampling programs with respect to their replications of built-in size and age compositions in the simulated fishery, and identifying a cost-effective port sampling program for quantifying size and age compositions of commercial landings.

IV-1-6. Develop a spreadsheet of assessment model input data that reflects the decisions and recommendations of the Data Workshop. Review and approve the contents of the input spreadsheet by June 1.

This TOR was not dealt with in the DW, but will be done after the DW.

I did not see a spreadsheet for assessment model input data reflecting the decisions and recommendations of the DW. Some issues surrounding commercial landings were not resolved in the DW. However, I believe life history data and abundance indices/CPUE data are ready.

IV-1-7. Prepare the Data Workshop report providing complete documentation of workshop actions and decisions (Section II. of the SEDAR assessment report). Develop a list of tasks to be completed following the workshop.

This TOR was addressed in the DW.

The final DW report was not completed in the DW, and will not be due for four weeks, two weeks ahead of the deadline when I have to submit this CIE review report. However, I received the draft reports from the LHG at the end of the DW. Thus, this CIE report only reflects my observations and comments on the DW process, discussions and preliminary reports.

The DW Coordinator (i.e., Dr. Julie Neer) clearly defined a list of tasks to be completed (and their deadlines) at the end of the DW.

IV-2. Gulf of Mexico Tilefish Data Workshop

IV-2-1. Characterize stock structure and develop unit stock definitions for the tilefish complex. Provide maps of species and stock distribution.

There was some discussion about this TOR in the DW, but no evidence either reject or support the unit stock definition.

Two tilefish species were considered in the DW, including golden tilefish and blueline tilefish. According to the American Fisheries Society, the tilefish is "golden tilefish". Because the DW considered that there was not enough information for blueline tilefish stock assessment, I focus this review on the GOM tilefish (i.e., golden tilefish).

The population of tilefish in the Gulf of Mexico was implicitly considered as a unit stock in the DW, and was assumed to be distinct from those in the Atlantic. However, limited information is available to support/oppose this unit stock definition. An understanding of the spatial dynamics of tilefish is limited by the spatial and temporal coverage of fishery-dependent monitoring programs and fishery-independent survey programs and by relatively small number of samples taken in the fishery and surveys. I recommend that more work (e.g., genetic studies, quantifying spatial variability in key life history parameters) be done for the evaluation of possible existence of local stocks/populations and if assessment and management on a spatial scale much finer than the Gulf-wide scale is needed for the tilefish in the GOM.

The distributional maps of tilefish in the US part of the GOM were provided, but the accuracy is unknown. No map was provided for Mexico's part of the GOM. I suggest working with relevant Mexican agencies to develop a Gulf-wide map for tilefish distribution.

IV-2-2. Review, discuss and tabulate available life history information (e.g., age, growth, natural mortality, reproductive characteristics); provide appropriate models to describe growth, maturation, and fecundity by age, sex, or length as applicable. Evaluate the adequacy of available life-history information for conducting stock assessments and recommend life history information for use in population modeling.

The Workshop had extensive and excellent discussion about this TOR. The key life history parameters needed for the stock assessment were estimated, and the uncertainty essential in defining priors and possible sensitivity analyses were provided for most parameters.

Samples used to determine the tilefish growth were mostly from the EGOM (70%). However, previous studies show that tilefish in the WGOM tend to be large and old. Thus, the growth data derived for the tilefish might not reflect spatial variability in growth in the GOM.

After an extensive discussion, the Workshop concluded that tilefish show characteristics of protogyny, changing from females to males with an increased age/length. However, the contrast in maturity and sex transition by age/size was not well defined, resulting in poor estimates of age/size at maturity for females and age/size at sex transition.

Tilefish are batch spawners, and quantification of the total egg production over a prolonged spawning season is difficult. The data collected in previous studies suggest that somatic biomass is not proportional to egg production, making it an unreliable approximation to reproductive potential. The DW suggests that the SSB-total, SSB-female, and the female gonad weight proxy be used in more sensitivity runs. I concur with the DW recommendations.

Most tilefish age-length data were not sexed. There were large differences in VBGF between females and males. However, there was a lack of small/young fish in samples, making the estimates of some VBGF parameters not well defined (e.g., unrealistic t0 value). No variability was estimated for the VBGF parameters. I suggest that bootstrapping be used in defining uncertainty in estimating VBGF parameters. I also recommend that appropriate age-length keys be developed to transform length composition data to age composition data.

I commend the LHG efforts to go through an exhaustive list of methods for estimating natural mortality rates. This effort yielded a wide range of estimates of natural mortality, which helps quantify uncertainty associated with M estimates.

Based on the discussion and previous studies, it is likely that spatial variability in life history parameters may exist for the GOM tilefish. However, limited information is available for such tests. I recommend that future sampling efforts collect data across large spatial contrasts for evaluating spatial variability in life history parameters.

IV-2-3(a). Provide measures of population abundance that are appropriate for stock assessment.

This TOR was extensively discussed in the Data Workshop.

The IG estimated abundance indices for the tilefish from both fishery-dependent and independent data sources. The fishery-independent survey program is the NMFS bottom longline surveys (2000-2009 with 2005 being removed because of Hurricane Katrina) and the fishery-dependent program is longline logbooks (1992 – 2009).

NMFS fishery-independent bottom longline survey

The annual NMFS fishery-independent bottom longline survey in GOM has been conducted by the Southeast Fisheries Science Center (SEFSC) Mississippi Laboratories since 1995, and is designed to provide fishery-independent survey data for the assessment of as many species as possible. For SEDAR 22, only data from 2000-2009 were considered because the use of J-type hooks and lack of coverage of deep depth strata in the survey prior to 2000 resulted in very low catch efficiency for the tilefish and circle-hooks have been used since 2000, which greatly increases tilefish catch. Frozen Atlantic mackerel are used as bait and soak time is standardized at one hour. However, the time duration between the first hook in the water and the last hook out of the water varies from 1.5-3 hours, which makes soak time vary among hooks.

The survey covers three depth strata. Survey stations where tilefish were captured were plotted. However, spatial coverage of the survey program varied during the time series due to weather or mechanical problems. Only data from survey stations within the depth range of capture for tilefish were used in development of abundance indices. Year 2005 was excluded in the analysis because of Hurricane Katrina. The DW suggests that the depth range covered by the program may not be deep enough to cover all depth ranges of tilefish. I recommend a comparative study of tilefish abundance index among different depth strata, if sample sizes are sufficient.

The delta-lognormal abundance index was developed using an approach described by Lo et al. (1992) in the index standardization. Various statistical procedures were employed (e.g., Q-Q plot and plot of residuals against year) for checking the validation of the distribution assumption in the standardization modeling. I commend the IG efforts to standardize the abundance index and consider the general approach to be sound. However, I notice that sediment, which is critically important in influencing the tilefish spatial distribution, was not included in the standardization. The use of GLM also implicitly assumes that interactions of spatial, temporal and environmental variables versus survey abundance are log-linear, thus limiting other forms of interactions. I recommend that GAM is considered in the abundance index standardization.

Fishery-dependent bottom longline fishery

Commercial CPUE data were derived from the Reef Fish Logbook (RFL) program, which requires all vessels with reef fish permits to report their catch. A two-stage modeling approach (Lo et al. 1992) was used in standardizing CPUE, which uses a binomial distribution function to model the proportion of positive trips per stratum and a general linear model (GLM) to quantify impacts of fixed factors on log(CPUE). This was done separately for EGOM longline, EGOM handline, and WGOM longline. The general approach is sound, and I commend that the IG efforts to remove impacts of other factors on CPUE.

However, the GAM model may be more flexible in incorporating possible nonlinear interactions between the tilefish abundance and fixed factors considered in the standardization. Sediment and depth were not included in the standardization, although they are key variables influencing the spatial distribution of tilefish in the GOM. I suggest that (1) GAM be used, (2) sediment and depth be included in the standardization, (3) model performance be evaluated with cross-validation.

For standardization of the two abundance indices, the IG ran through a long checklist that was developed in the Indices Workshop to evaluate quality of the abundance index standardization. I commend the IG efforts.

IV-2-3(b). Consider and discuss all available and relevant fishery dependent and independent data sources.

Abundance indices derived from a fishery-independent program and a fishery-dependent program were recommended in the DW for quantifying the stock dynamics of tilefish in the GOM. The two programs differ greatly in their designs, data sources, and spatial and temporal coverage. They may target different components of the tilefish population in the GOM, and have different selectivities and sampling efficiencies. The use of these two abundance indices can complement each other. The IG provided a good description of the programs. However, there were concerns in the DW that few tilefish were caught in southern Florida, which might result from the monitoring program missing the main tilefish habitat and depths, rather than the lack of tilefish in southern Florida. I concur with the concern and recommend considering covering deep depths and tilefish habitat in the area in the future.

Fisheries CPUE may not be a good abundance indicator (need to understand spatial dynamics of fishing fleet, spatial variability in CPUE and how it changes over time). Because commercial catch was reported per trip, not by set, the area where catch occurs might not be same as the area where reported catch occurred. This bias in spatial distribution of catch might introduce potential errors in the index standardization.

The IG discussed issues associated with each of the two abundance indices within the group and in the plenary of all the panelists. However, I did not observe the discussion about the relative importance of the two abundance indices. Thus, there is no assignment of relative weights for the abundance indices. I recommend that the IG do some comparative analyses of these two abundance indices (e.g., an analysis of the two abundance indices for their coherence).

IV-2-3(c). Document all programs evaluated, addressing program objectives, methods, coverage, sampling intensity, and other relevant characteristics.

The fishery-independent and fishery-dependent programs were well described in terms of their objectives, methods, spatial and temporal coverage, sampling intensity, fishing gears, and development over the survey time period. This helps one to better understand differences and similarities between the programs.

IV-2-3(d). Provide maps of survey coverage.

Maps of the bottom longline survey coverage were provided.

IV-2-3(e). Develop CPUE and index values by appropriate strata (e.g., age, size, area, and fishery); provide measures of precision and accuracy.

The nominal and standardized abundance indices and fishery CPUE were derived for the GOM tilefish. No abundance index and CPUE were estimated by age, size, and area. CPUE was estimated for the GOM longline fishery, but not for vertical lines and recreational fisheries. Coefficients of variation (CV) were estimated for the standardized CPUEs and abundance indices, but not for the nominal abundance indices and CPUEs. No measure of accuracy was provided.

IV-2-3(f). Evaluate the degree to which available indices adequately represent fishery and population conditions.

The NMFS bottom longline survey is likely to cover the spatial distribution of tilefish in the GOM although it was suggested in the DW that the survey might not cover the full depth range of tilefish. The spatial coverage of this program varies from year to year. Given likely differences in population structure and large differences in oceanographic dynamics between the EGOM and WGOM, the quality of the indices derived from this program may be questionable for representing the overall tilefish population dynamics in the GOM. For the longline CPUE, because fishermen report catch in catch per trip, not catch per set, CPUE might not correspond well with the effort, which suggests that CPUE is perhaps not a reliable indicator of fish abundance.

The quality of these indices may lead to reduced confidence in the stock status assessment. Differences in spatial coverage between the monitoring programs and the fact that no fishery-independent survey program is specifically designed for sampling the GOM tilefish may also be issues needing to be considered in using these indices in stock assessment. I encourage the continued evaluation of fishery-independent and fishery-dependent data quality and development of different scenarios for evaluating their impacts on stock assessment.

IV-2-3(g). Recommend which data sources are considered adequate and reliable for use in assessment modeling.

The DW suggested that abundance indices derived from one fishery-independent survey program (NMFS bottom longline survey) and one fishery-dependent program (longline logbook) be used in the assessment modeling.

The NMFS bottom survey program is not designed for capturing tilefish and experienced changes in spatial and depth coverage and changes in fishing gear (changed from J-shape hooks to circle hooks), which may lead to large variability in survey catchability (and vulnerability) over time and result in biased representation of population structure. I recommend that data from the two programs be compared for key population variables/parameters (e.g., abundance, sex ratio, and size composition). Such a comparative study may provide some information on differences in sampling catchability. For each program, I also suggest that an evaluation of among-station variability be conducted, which may provide some insights about the quality of the data (i.e., highly aggregated or over a wide spatial range). I also recommend that the programs be ranked for their relevance to tilefish spatial distributions and sampling efficiency. The ranking can then be considered in the assessment.

IV-2-4(a) Characterize commercial and recreational catch, including both landings and discard, in pounds and number.

This TOR was discussed extensively in the DW.

Fishery-dependent data were discussed in characterizing catch in commercial and recreational tilefish fisheries. Unusually high tilefish landings from droplines were verified based on three counties' statistics, but the proportion of catch from droplines was exceptionally high in early years; 10% catch from dropline was assumed to be more reasonable according to fishermen. The catch statistics were adjusted accordingly. Discards were considered small, and could be neglected in the stock assessment. The working group considered that it was almost impossible to reliably quantify recreational catch over the spatial and temporal range of the tilefish population in the GOM. Catch was measured in weight (lbs), and no catches in number were estimated.

It was clear to me from the discussion in the DW that the quality of commercial and recreational catch is questionable. However, I believe that the SS3 assessment program considers that catch statistics have no errors. This assumption is certainly violated for the tilefish. I recommend that different scenarios of catch be considered for sensitivity analyses to evaluate impacts of uncertainty in catch estimates on stock assessment.

The quality of landings data for blueline tilefish was considered questionable.

IV-2-4(b). Provide estimates of discard mortality rates by fishery and other strata as appropriate or feasible.

Limited data shown in the DW suggest that discards are very low in the GOM tilefish fishery. This was confirmed by all the fishermen present at the DW. This might result from lack of size limit regulation and no incentive for high grading. However, with the implementation of trip limits on Jan. 1, 2010, which often provides incentives for high grading, discarding may become an issue. Monitoring programs need to be developed to collect the relevant data.

Because tilefish is a deep water species, the mortality rate of discarded tilefish is likely to be 100%.

IV-2-4(c). Evaluate and discuss the adequacy of available data for accurately characterizing harvest and discards by species and fishery sector.

Overall, the harvest data reflect temporal variability in commercial landings in the GOM. However, I heard repeatedly in the DW that commercial landing statistics might have large uncertainty, especially in earlier years. Species misidentification might affect the quality of landings data. This includes tilefish being mis-identified or mislabeled as other species and other species being mis-identified as tilefish. Such impacts may differ among years because a preliminary study suggests that species misidentification rates change over time, tending to be severe in early years but less so in more recent years. This issue was discussed extensively, but not totally resolved in the DW. A working group was formed to further discuss this issue after the DW.

IV-2-4(d). Provide length and age distributions if feasible.

There was limited discussion about the quality of estimated length distributions for landed catch. Given the large spatial distribution of the fishery and landing ports, the quality of length distribution should have been discussed more extensively in the DW. This was especially true for length composition data in earlier years when port sampling efforts were limited. There were also large differences in port sampling efforts among different states in the GOM, which might result in different quality of length composition data estimated for the GOM.

There was limited discussion on the estimation of age composition for the commercial landings. There was little discussion about the estimation of sex ratio for catch data.

IV-2-4(e). Provide maps of fishery effort and harvest.

There was no map for spatial distribution of fishing efforts and landed catch.

IV-2-5(a). Provide recommendations regarding the feasibility of conducting a benchmark assessment for each species in the tilefish complex.

This TOR was discussed in the DW and it was concluded that a formal stock assessment can be done for tilefish (i.e., golden tilefish), but not for blueline tilefish.

Given the data quality and quantity for tilefish and blueline tilefish, I believe this is an appropriate assessment.

IV-2-5(b). If the data are deemed insufficient for a benchmark assessment, provide guidance on the type of management advice that can be provided with that data (see SEDAR Caribbean Data Evaluation Workshop report).

This was not discussed in the DW.

IV-2-6. Provide recommendations for future research in areas such as sampling, fishery monitoring, and stock assessment. Include specific guidance on sampling intensity (number of samples including age and length structures) and appropriate strata and coverage.

The DW had some discussion about this TOR, but the discussion was not extensive enough to lead to concrete recommendations for future research.

WGOM and EGOM differ greatly in oceanographic conditions, suggesting a need to evaluate whether there was difference between the two areas in key life history traits. I suggest that spatial variability be evaluated in key life history processes to identify population spatial dynamics, and possible existence of local stocks.

Sampling programs for quantifying size composition and age composition of commercial catch need to be carefully evaluated. Factors such as adequate spatial and temporal coverage and sampling intensity to have high effective sample sizes should be considered. I recommend developing alternative sampling designs, developing a simulated fishery that mimics temporal and spatial variability in size and age compositions in commercial landings, applying current and alternative sampling programs to the simulated fishery, comparing the performance of the sampling programs with respect to their replications of built-in size and age compositions in the simulated fishery, and identifying a cost-effective port sampling program for quantifying size and age compositions of commercial landings.

IV-2-7. Develop a spreadsheet of assessment model input data that reflects the decisions and recommendations of the Data Workshop. Review and approve the contents of the input spreadsheet by June 1.

This TOR was not dealt with in the DW, but will be done after the DW.

I did not see a spreadsheet for the assessment model input data reflecting the decisions and recommendations of the DW. Some issues surrounding commercial landings were not resolved in the DW. However, I believe life history data and abundance indices/CPUE data are ready.

IV-2-8. Prepare the Data Workshop report providing complete documentation of workshop actions and decisions (Section II. of the SEDAR assessment report). Develop a list of tasks to be completed following the workshop.

This TOR was addressed in the DW.

The final DW report was not completed in the DW, and will not be due for four weeks, two weeks ahead of the deadline when I have to submit this CIE review report. However, I received the draft reports from the LHG at the end of the DW. Thus, this CIE report reflects my observations and comments on the DW process, discussions and preliminary reports.

The DW Coordinator (i.e., Dr. Julie Neer) clearly defined a list of tasks to be completed (and their deadlines) at the end of the workshop.

V. Conclusions and Recommendations

I would like to commend the great efforts of all the participating scientists, managers and fishermen in the SEDAR 22 DW in the identification, evaluation and compilation of the information on life history, fishery-dependent and fishery-independent abundance indices, and landings in the commercial and recreational fisheries for YG, tilefish (i.e., golden tilefish), and blueline tilefish in the GOM. I was impressed by the breadth of expertise and experience of the panelists, openness of discussion for considering alternative approaches/suggestions, and constructive dialogs in each working group and at the plenary meetings throughout the workshop. All the comments, whether they were from scientists, managers, or fishermen, were fully considered and discussed. In particular, I commend the inclusion in the Data Workshop of fishermen, who provided insights on the quality of the fishery data, in particular for historical fisheries data. I observed on many occasions constructive interactions and dialogs between scientists/mangers and representatives of the industry in the Workshop.

In general, I consider the information identified and compiled in the DW represents the best efforts given all the limitations associated with data quality and quantity. I consider the approaches used in developing life history parameters, fisheries landings, and abundance indices sound.

Having said that, I believe that there are large uncertainties associated with data identified and compiled in the DW, and that there is room for further improvement. I have made the following general comments and specific recommendations.

<u>General comments</u>
Although the SoW states that all the working papers and reference/background information for the workshop will be available two weeks before the workshop, only a few working papers (less than 25% of all the working papers promised) were available before the start of the workshop (not mention two weeks before the start of the workshop). Many working papers were still not ready in the middle of the workshop, which made my work difficult. The three separate working groups worked concurrently everyday, making it impossible for me, as the only CIE reviewer, to be fully involved in each group’s discussions.

I was told at the DW that Stock Synthesis 3 (SS3) will be used for the assessment of YG and tilefish. This choice of stock assessment model has direct impacts on the quality and quantity of the data that need to be evaluated and compiled in the Workshop. However, I observed that most DW panelists did not know exactly the data requirements, key assumptions, and options of the

SS3 program. I recommend that future data workshop start with the introduction of the stock assessment model that will be used in the assessment so that data workshop participants understand the information needs of the stock assessment model.

I noticed that the time period that the SEDAR 22 assessment covers had not been defined prior to the DW. I suggest that a stock assessment time period be defined prior to the DW so that working groups can focus on the defined time period, and not waste time discussing data falling outside the target stock assessment. The DW may also be a good place to discuss and make a decision about the time period the stock assessment should cover.

There is a need to include scientific names for all species covered in the TORs and SoW. The tilefish is the official name of golden tilefish in the American Fisheries Society list of fish species. However, both golden tilefish and blueline tilefish were discussed at the Workshop. This creates some confusion.

It is clear from all the discussions at this Workshop that the information for blueline tilefish is not sufficient for a formal stock assessment using an assessment model like SS3.

Specific recommendations

Although I have provided detailed comments and recommendations under each TOR, I re-iterate the following recommendations.

- Possible existence of local stocks for both species needs to be evaluated;
- More comparative studies need to be done to evaluate differences in data collected from different monitoring programs;
- More comparative studies need to be done to evaluate differences in parameters estimated using different methods to improve our understanding of the degree of uncertainty associated with these parameters;
- More comparative studies need to be done to evaluate spatial and temporal variability in key life history parameters, abundance indices and landings;
- More habitat variables need to be included in CPUE and abundance index standardization;
- General additive models need to be considered in standardizing abundance index and CPUE;
- Instead of using a point estimate as a bias correction factor in correcting potential biases in landings data, a range of correction factors can be used so that large uncertainty in landings data can be incorporated into the stock assessment;

- The quality of catch data (landings, catch size/age composition, catch sex ratio etc.) is probably the most questionable of the data available to the stock assessment for both fish species, and the stock assessment model should have an ability to incorporate uncertainty in catch data;

- A critical evaluation of fishery-independent monitoring programs should be done to identify problems associated with the current program design in quantifying population dynamics;

- A systematic mail survey/interview of fishermen who have been involved in the GOM YG and tilefish needs to be done to have a better understanding of the degree of misreporting/underreporting and to identify if there is spatial and temporal variability in underreporting;

- It appears that outliers may exist in the assessment and given the data quality concerns, I suggest that robust estimation methods be used in the assessment (although this may be the choice of the modelers, but I believe that the Data Workshop is a place to make the recommendation because this is the place to deal with data quality issues);

- Uncertainty should be considered in all life history modeling, and confidence intervals should be estimated for the key life history parameters for the GOM YG and tilefish;

- Because of the extremely small YG catch in the SEAMAP bottom trawl survey, caution should be used in applying the derived abundance index, and the change in survey protocol in 1987 calls for a separate analysis of the two time periods and two different catchabilities in population modeling;

- Different measures for SSB should be considered for both tilefish and YG in stock assessment modeling; and

- I recommend conducting a systematic evaluation of current sampling programs for quantifying size composition and age composition of commercial catch. Factors such as adequate spatial and temporal coverage and sampling intensity to have high effective sample sizes should be considered. I recommend developing alternative sampling designs, developing a simulated fishery that mimics temporal and spatial variability in size and age compositions in commercial landings, applying current and alternative sampling programs to the simulated fishery, comparing the performance of the sampling programs with respect to their replications of built-in size and age compositions in the simulated fishery, and identifying a cost-effective port sampling program for quantifying size and age compositions of commercial landings.

Finally, I strongly concur with the recommendations made by the LHG in their draft DW report regarding life history work for the GOM YG and tilefish, and I think all the issues raised in the report are critical to improve the life history data quality. The draft reports of the other two groups (IG and LDG) were not available when I prepared this report so I cannot make any comments regarding the recommendations they will list in the DW reports.

VI. References

Able, K. W., C. B. Grimes, R. A. Cooper and J. R. Uzmann. 1982. Burrow construction and behavior of tilefish, Lopholatilus chamaeleonticeps, in Hudson Submarine Canyon. Env. Biol. Fish. 7:199-205.

Cass-Calay, S. L. and M. Bahnick. 2002. Status of the yellowedge grouper fishery in the Gulf of Mexico. Sustianable Fisheries Division Contribution No. SFD-02/03-172. Southeast Fisheries Science Center, Miami, FL.

Chen, Y., and D. Fournier. 1999. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Canadian Journal of Fisheries and Aquatic Science 56:1525–1533.

Chen, Y., and J. E. Paloheimo. 1998. Can a more realistic model error structure improve parameter estimation in modelling the dynamics of fish populations? Fishery Research (Amsterdam) 38:9–19.

Cook, M., G. R. Fitzhugh and J. S. Franks. 2009. Validation of yellowedge grouper, *Epinephelus flavolimbatus*, age using nuclear bomb-produced radiocarbon. Environ Biol Fish 86:461–472 DOI 10.1007/s10641-009-9536-x.

Dooley, J.K. 1978. Systematics and Biology of the Tilefishes (*Perciformes: Branchiostegidae and Malacanthidae*), with description of two new species. NOAA Technical Report. 411: 1-78.

Grimes, C.B., and S.C. Turner. 1999. The complex life history of tilefish, *Lopholatilus chamaeleonticeps*, and vulnerability to exploitation. In Life in the slow lane: Ecology and conservation of long-lived marine animals (J. A. Musick, ed.), p. 17-26. American Fisheries Society Symposium 23. Bethesda, MD.

Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York.

Lo, N. C. H., L.D. Jacobson, and J.L. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49: 2515-1526.

NRC (National Research Council). 1997. Improving fish stock assessments. National Academy Press, Washington D.C.

NRC (National Research Council). 1999. Sustaining marine fisheries. National Academy Press, Washington D.C.

Palmer, S.M., Harris, P.J., and P.T. Powers. 2004. Age, growth and reproduction of tilefish, *Lopholatilus chamaeleonticeps,* along the southeast Atlantic coast of the United States, 1980-

1987 and 1996-1998. SEDAR04-DW-18. SEDAR/SAFMC, 1 Southpark Circle #306, Charleston, SC. 21 pp.

Quinn, T. J., II, and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191.

Walters, C. J. 1998. Evaluation of quota management policies for developing fisheries. Canadian Journal of Fisheries and Aquatic Science 55:2691–2705.

Appendix VII-1: Statement of Work for Dr. Yong Chen

External Independent Peer Review by the Center for Independent Experts

SEDAR 22 –Gulf of Mexico yellowedge grouper and tilefish Data Workshop

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. The CIE reviewer is selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. The CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.com.

Project Description: SEDAR 22 will be a compilation of data, a benchmark assessment of the stock, and an assessment review for conducted Gulf of Mexico yellowedge grouper and tilefish under the SEDAR (Southeast Data, Assessment and Review) process. This proposal is for a CIE expert to be appointed to participate as a CIE independent peer reviewer on the Assessment Panel during the data compilation and assessment processes. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

SEDAR assessments typically involve an assessment panel composed of assessment analysts named by the lead SEDAR cooperator, fishery scientists as SSC members, and fishery managers. This proposal is based in part on a recent SEDAR assessment panel recommendation that the assessment panel include an independent expert peer review person to serve as a workshop panelist during the process leading to an Assessment Review Workshop. While the independent expert will not contribute to the production of science products, he or she can be valuable by providing peer review advice regarding technical details of the methods used in SEDAR assessments and decisions related to model configuration during the workshop. In providing peer review advice during the assessment workshop, the independent expert can improved the overall assessment process by advising the analysts regarding issues that might become points of contention in the formal peer review workshop—at which time it would be too late to revise the actual assessment (assessment data decisions, assumptions, models, modifications, etc. are confined to the assessment process before the peer review workshop).

Requirements for CIE Reviewer: One CIE reviewer shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. The CIE reviewer shall have working knowledge and recent experience in the application of stock assessment, statistics, fisheries science, and marine biology sufficient to complete the task of participation in discussions of technical details of the data and methods used for this SEDAR assessment and decisions related to model configuration in compliance with the workshop's Terms of Reference. The CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: The CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Tampa, Florida during 15-19 March 2010.

Statement of Tasks: The CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewer. The NMFS Project Contact is responsible for providing the CIE reviewer with the background documents, report, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When the CIE reviewer participates during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewer who is a non-US citizens. For this reason, the CIE reviewer shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewer the necessary background information and report for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. The CIE reviewer is responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewer shall read all documents in preparation for the peer review.

By 15 October 2009, the NMFS Project Contact will provide a list of background documents and reports with estimated number of pages that the reviewer will use to prepare for the meeting.

I do not know at this time how many documents will be available for review prior to the data workshop but with two species, I would suspect at least 20 working papers.

Panel Review Meeting: The CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **The CIE reviewer serves only as a peer reviewer in accordance with the SoW, and shall not serve as an analyst during the workshop. Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** The CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewer as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Report: The CIE reviewer shall complete an independent peer review report in accordance with the SoW. The CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. The CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: The CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewer: The following chronological list of tasks shall be completed by the CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and report provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in Tampa, Florida from 15-19 March 2010, as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than 2 April 2010, the CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and David Sampson, CIE Regional Coordinator via email to david.sampson@oregonstate.edu. The CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

8 February 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
1 March 2010	NMFS Project Contact sends the CIE Reviewer the pre-review documents
15-19 March 2010	The reviewer participates and conducts an independent peer review during the panel review meeting
2 April 2010	CIE reviewer submits draft CIE independent peer review report to the CIE Lead Coordinator and CIE Regional Coordinator
16 April 2010	CIE submit CIE independent peer review report to the COTR
23 April 2010	The COTR distributes the final CIE report to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewer to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review report by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these report shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review report) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE report in *.PDF format to the COTR. The COTR will distribute the CIE report to the NMFS Project Contact and Center Director.

Key Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Julie Neer, SEDAR 22 Coordinator, NMFS Project Contact
South Atlantic Fishery Management Council
4055 Faber Place, Suite 201
North Charleston, SC 29405
Julie.Neer@SAFMC.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewer should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewer should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewer should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewer shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

SEDAR 22 –Gulf of Mexico Yellowedge Grouper Data Workshop

1. Characterize stock structure and develop a unit stock definition. Provide maps of species and stock distribution.
2. Review, discuss and tabulate available life history information (e.g., age, growth, natural mortality, reproductive characteristics); provide appropriate models to describe growth, maturation, and fecundity by age, sex, or length as applicable. Evaluate the adequacy of available life-history information for conducting stock assessments and recommend life history information for use in population modeling.
3. Provide measures of population abundance that are appropriate for stock assessment. Consider and discuss all available and relevant fishery dependent and independent data sources. Document all programs evaluated, addressing program objectives, methods, coverage, sampling intensity, and other relevant characteristics. Provide maps of survey coverage. Develop CPUE and index values by appropriate strata (e.g., age, size, area, and fishery); provide measures of precision and accuracy. Evaluate the degree to which available indices adequately represent fishery and population conditions. Recommend which data sources are considered adequate and reliable for use in assessment modeling.
4. Characterize commercial and recreational catch, including both landings and discard, in pounds and number. Provide estimates of discard mortality rates by fishery and other strata as appropriate or feasible. Evaluate and discuss the adequacy of available data for accurately characterizing harvest and discard by species and fishery sector. Provide length and age distributions if feasible. Provide maps of fishery effort and harvest.
5. Provide recommendations for future research in areas such as sampling, fishery monitoring, and stock assessment. Include specific guidance on sampling intensity (number of samples including age and length structures) and appropriate strata and coverage.
6. Develop a spreadsheet of assessment model input data that reflects the decisions and recommendations of the Data Workshop. Review and approve the contents of the input spreadsheet by June 1.
7. Prepare the Data Workshop report providing complete documentation of workshop actions and decisions (Section II. of the SEDAR assessment report). Develop a list of tasks to be completed following the workshop.

SEDAR 22 –Gulf of Mexico Tilefish Data Workshop

1. Characterize stock structure and develop unit stock definitions for the tilefish complex. Provide maps of species and stock distribution.
2. Review, discuss and tabulate available life history information (e.g., age, growth, natural mortality, reproductive characteristics); provide appropriate models to describe growth, maturation, and fecundity by age, sex, or length as applicable. Evaluate the adequacy of available life-history information for conducting stock assessments and recommend life history information for use in population modeling.
3. Provide measures of population abundance that are appropriate for stock assessment. Consider and discuss all available and relevant fishery dependent and independent data sources. Document all programs evaluated, addressing program objectives, methods, coverage, sampling intensity, and other relevant characteristics. Provide maps of survey coverage. Develop CPUE and index values by appropriate strata (e.g., age, size, area, and fishery); provide measures of precision and accuracy. Evaluate the degree to which available indices adequately represent fishery and population conditions. Recommend which data sources are considered adequate and reliable for use in assessment modeling.
4. Characterize commercial and recreational catch, including both landings and discard, in pounds and number. Provide estimates of discard mortality rates by fishery and other strata as appropriate or feasible. Evaluate and discuss the adequacy of available data for accurately characterizing harvest and discard by species and fishery sector. Provide length and age distributions if feasible. Provide maps of fishery effort and harvest.
5. Provide recommendations regarding the feasibility of conducting a benchmark assessment for each species in the tilefish complex. If the data are deemed insufficient for a benchmark assessment, provide guidance on the type of management advice that can be provided with that data (see SEDAR Caribbean Data Evaluation Workshop report).
6. Provide recommendations for future research in areas such as sampling, fishery monitoring, and stock assessment. Include specific guidance on sampling intensity (number of samples including age and length structures) and appropriate strata and coverage.
7. Develop a spreadsheet of assessment model input data that reflects the decisions and recommendations of the Data Workshop. Review and approve the contents of the input spreadsheet by June 1.
8. Prepare the Data Workshop report providing complete documentation of workshop actions and decisions (Section II. of the SEDAR assessment report). Develop a list of tasks to be completed following the workshop.

Annex 3: Tentative Agenda

SEDAR 22 –Gulf of Mexico yellowedge grouper and tilefish Data Workshop

Tampa, Florida during 15-19 March 2010

Point of contact for reviewer security & check-in - NA