Characterizing and Presenting Assessment Uncertainty

SEDAR Procedural Workshop IV

Edited by

John Carmichael

June 2010

Executive Summary

The Reauthorized Magnuson-Stevens Act requires Council Scientific and Statistical Committees (SSC) to provide fishing level recommendations, including Acceptable Biological Catch (ABC), that prevent overfishing. Revised National Standard 1 extensively cites the obligation of the SSCs to account for uncertainty in data and assessment methods when making fishing level recommendations that prevent overfishing. National Standard 1 also directs the Councils, in consultation with their SSCs, to develop ABC control rules that account for uncertainties. As a result, there is renewed interest in methods of evaluating and reporting assessment uncertainty from all sources. Particular attention is being devoted to methods that can present assessment uncertainty in terms of the probability of overfishing occurring.

Given this increased focus on uncertainty, the SEDAR Steering Committee recommended convening a procedural workshop to address uncertainty in SEDAR stock assessments. The goal of this workshop is to develop guidance that will enable SEDAR assessments to provide the information SSCs require to apply their ABC control rules, improve evaluation of uncertainty in SEDAR assessments, and increase consistency in treatment of uncertainties across SEDAR assessments.

Participants included SSC members and Council staff from around the Southeast region, University researchers, and agency staff from around the country. Dr. James Berkson of the Southeast Fisheries Science Center chaired the workshop. Presentation topics ranged from specific ABC control rules and common approaches for addressing uncertainty in the different regions to the latest, cutting-edge approaches for evaluating assessment uncertainty. Participants engaged in extensive discussion of the types of uncertainties and how they can be addressed. A number of suggestions were offered, both for the SEDAR process specifically and for consideration of uncertainty in general.

Finally, although this workshop addressed many aspects of assessment uncertainty and allowed invaluable information exchange between nationwide assessment experts, a simple “cookbook approach” for addressing uncertainty could not be produced. This is because selecting particular approaches for incorporating uncertainty is highly dependent upon the data, the methods that will adequately address the data, and, to some extent, the use of the results and recommendations. Therefore, it is impractical and unrealistic for a workshop like this to make specific

recommendations for how any particular assessment will or should evaluate uncertainty. Rather, the intent of much of the discussion and subsequent report text is to illustrate likely sensitivities and common ways of addressing them, and to provide guidance for future assessments on general expectation for evaluating and reporting uncertainty.

Specific Recommendations for SEDAR Assessments:

- SEDAR assessments should strive to explicitly identify the primary and most influential uncertainties at each step of the assessment process and ensure these are carried forward to subsequent steps.
- SEDAR assessments should acknowledge uncertainties in data and assessment techniques and strive to address those uncertainties within the modeling framework.
- SEDAR should demonstrate robustness in preferred models and provide advice that addresses overall assessment uncertainty and provides distributions for key output parameters.
- SEDAR should provide advice that incorporates uncertainties and considers multiple states of nature represented by alternative model scenarios when appropriate.
    - Scientists involved in an assessment should provide information for choosing among the alternative model scenarios for developing management advice and determining stock status.
    - Each assessment should provide single distributions for each of the key outputs that reflect model uncertainties and alternative states of nature. This should include a distribution about OFL that enables the SSCs to determine ABC in accordance with ABC control rules.
    - In the event panelists cannot choose a single model scenario, then the probable candidate scenarios should be carried into the management arena for consideration in developing management strategies. It is recommended that the panelists strive to develop probabilities for the candidate scenarios that reflect their relative likelihood.
    - Each step of the process should acknowledge uncertainties and how they are addressed to inform subsequent steps.
- SEDAR panelists should strive to better communicate uncertainties, including 1) the differences between data, model-intrinsic, population dynamic, and ecological uncertainties, 2) the purpose of the techniques used to estimate uncertainties, and 3) how to interpret these uncertainties in the context of evaluating management options.
- SEDAR assessments should strive to improve consistency between assessments in addressing common or typical uncertainties.
- SEDAR should strive to integrate scientific evaluation of the assessment phase and quantitative aspects of the implementation phase (including projections) in light of the dependencies and feedbacks between them.

General Recommendations regarding evaluation of uncertainty:

- Determination of the appropriate minimum level of uncertainty variance, and how it relates to data and model methods, requires further work, but is an important topic in SSC efforts to develop fishing level recommendations.
- Management systems such as the Council process should develop a feedback loop to keep track of recommendations, actions, and implementation, for consideration and evaluation in subsequent assessments.
- Incorporating uncertainty in ABC values remains a developing process that will require flexibility in management and adaptive management. This workshop is but one step in this process and in the evolution of an effective and robust system.
- Uncertainty is pervasive. Scientists do not currently have all the tools necessary to fully and adequately identify and evaluate all the uncertainties that exist, nor do they have the tools and information necessary to incorporate all the known uncertainties into the assessment process. Managers should expect changes as knowledge advances in these areas.
- It is useful to maintain consistency in model treatment of uncertainty from one assessment to the next. This is a means of gathering information on the relative level of uncertainty and evaluating changes in the level of uncertainty over time as new data or new methods become available.

Table of Contents

Preface

The Council-Federal cooperative SEDAR process provides stock assessments for fisheries resources of the National Marine Fisheries Service's Southeast Region. As part of ongoing efforts to improve the efficiency of the assessment process and quality of the assessment products, SEDAR occasionally holds procedural workshops devoted to topics of concern. This, the fourth of such workshops, is devoted to the issue of uncertainty. Workshop objectives include identifying primary uncertainties in the assessment process, recommending Best Practices for addressing such uncertainties, and providing guidance to enable future SEDAR assessments to provide the information SSCs require to develop fishing level recommendations that account for uncertainty.

SEDAR convened this procedural workshop from February 22 - 26, 2010 in Charlotte, North Carolina, to discuss issues related to uncertainty in stock assessments. Participants included representatives from Federal agencies, territorial governments, nongovernmental organizations, Council technical and constituent advisors, and university researchers. The charge to the workshop was specified in Terms of Reference developed from the overall SEDAR Steering Committee approved objectives by a workshop Organizing Committee. Presentations were provided on typical SEDAR approaches to uncertainty by Southeast Region population analysts, on approaches used around the Nation by representatives of other NMFS Regional Science Centers, on advances in approaches by researchers active in the field, and on the SSCs control rules by representatives of the Gulf and South Atlantic SSCs.

This report documents the findings of the SEDAR Procedural Workshop on Uncertainty and addresses the workshop Terms of Reference. It also contains recommendations and Best Practices to be considered in future SEDAR assessments. The report is structured to best present the discussion and recommendations, and therefore does not always follow the order of the Terms of Reference as is typical of such reports. However, through the presentations, group discussions, and final recommendations, all Terms of Reference are addressed and report sections note which Terms of Reference are addressed therein. The first section of this report documents workshop presentations and subsequent group discussion on the topic addressed by the presentations. It includes abstracts from the presentations and summaries of the group discussion. The second section documents discussion and recommendations pertaining to four key areas of uncertainties that emerged from initial workshop discussion. The final section addresses overarching recommendations and conclusions and captures key statements made throughout the workshop.

The Workshop was organized around four themes: (I) past and current characterization techniques, (II) new and evolving techniques, (III) management implications, and (IV) recommendations. To provide context, the workshop opened with presentations on control rule development from representatives of the Gulf of Mexico and South Atlantic Fishery Management Council SSCs. Discussion of Theme I began with representatives from the Southeast Fisheries Science Center presenting overviews of the approaches typically used in SEDAR assessments to address uncertainty, and continued with presentations on techniques used in other regional science centers by representatives of the Pacific Islands, Northeast, Southwest, and Alaska centers. Several new approaches were presented during Theme II, including some with applicability in data poor situations. A presentation on progress of the ACL working group led off discussion for Theme IV and concluded the general presentations.

Participants divided into four working groups during the next phase of the workshop to develop straw man recommendations on key topics derived from the previous discussions. Topics selected for further development included: 1) Data and Input Uncertainties, 2) Assessment and Model Process Uncertainties, 3) Model Outputs and Developing Recommendations, and 4) Implementation and Feedback. Over the remaining days of the workshops there were several cycles of workgroups presenting results to the full panel and gathering responses and suggestions for refinement. By Friday morning each group had a set of final recommendations for review by the full panel and consideration of Best Practices. The final section of this report addresses the recommendations for each key topic area.

Finally, the workshop panel reviewed the recommendations of each subgroup within the context of the workshop Terms of Reference, and while considering the overall Federal fisheries management goal to prevent overfishing of fisheries resources. This allowed participants to consider linkages between each uncertainty area and ensure consistency across the overall process.

Workshop Time and Place

SEDAR Procedural Workshop IV – Uncertainty was held February 22 – 26, 2010 in Charlotte, North Carolina.

Terms of Reference

Review and discuss past approaches to characterizing and presenting uncertainty in stock assessments and projection analyses conducted under the SEDAR process.

Review and discuss alternative approaches to characterizing and presenting uncertainty in stock assessments and projection analyses, including those utilized in other regions. Discuss the sources of uncertainty which require consideration in the stock assessment process. Sources may include implementation uncertainty, within model uncertainty, inter-model uncertainties. Make recommendations on which sources of uncertainty to consider for future SEDAR stock assessments and projection analyses.

Recommend approaches for representing uncertainty in the assessment documentation and expressing confidence in estimated parameters. Discuss both inter- and intra-model uncertainty. Include guidance for different model classes.

Review and discuss uncertainty estimates needed for management, specifically addressing the needs for each council's ABC control rules and ACT determinations. Make recommendations on which uncertainty estimates should be included in future SEDAR stock assessments and projection analyses.

Provide recommendations on best use of uncertainty characterization and recommend methods such as P* analysis, risk evaluation approaches, and PSA.

Prepare a SEDAR procedures document addressing these recommendations that will be used to guide future SEDAR assessments and projection analyses.

Workshop Participants

Workshop Panel

Jim Berkson, Workshop Chair.. NMFS SEFSC/ Virginia Tech
Kate Andrews.. NMFS SEFSC Panama City
Steven Atran...GMFMC Staff
Carolyn Belcher .... SAFMC SSC/GADNR)
Harry Blanchet..GMFMC SSC/ LADWLF
John Boreman .... SAFMC SSC/NCSU
Jon Brodziak ....NMFS PIFSC
Noel Cadigan ....CIE Reviewer
John Carlson.. NMFS SEFSC Panama City
John Carmichael...SEDAR/SAFMC Staff
Shannon Cass-Calay .... NMFS SEFSC Miami
Rob Cheshire..NMFS SEFSC Beaufort
Matt Cieri .... SAFMC SSC/MEDNR
Paul Conn..NMFS SEFSC Beaufort
Ray Conser.. NMFS SWFSC
Enric Cortés .... NMFS SEFSC Panama City
Katie Drew ....ASFMC
Kari Fenske .... SAFMC Staff
John Froeschke.. GMFSC Staff
Rich Fulford .... GMFSC SSC/USM
Dana Hanselman ....NMFS AFSC
Rick Hart...NMFS SEFSC Galveston
Dennis Heinemann.. Ocean Conservancy
Ron Hill...NMFS SEFSC Galveston
Barbara Kojis .... CFMC SSC
Chris Legualt.. NMFS NEFSC
Behzad Mahmoudi ....FWRI
Bob Muller ....FWRI
Joseph Munyandorero ....FWRI
Mike Murphy ....FWRI
Joe O'hop ....FWRI
Mauricio Ortiz... NMFS SEFSC Miami
Clay Porch... NMFS SEFSC Miami
Joe Powers .... GMFSC SSC/LSU
Mike Prager..NMFS SEFSC Beaufort
Amy Schueller ....NMFS SEFSC Beaufort
Kyle Shertzer ....NMFS SEFSC Beaufort

Doug Vaughan ....NMFS SEFSC Beaufort
Robert Wakeford...MRAG Americas, Inc
Mike Wilberg ....Chesapeake Biological Lab
Erik Williams .... NMFS SEFSC Beaufort

Observers

Claudia Friess.. Ocean Conservancy
Stacy Hudy.. Virginia Tech
Adyan Rios.. Virginia Tech
Emily Susko .... Virginia Tech

Staff

Tyree Davis .... NMFS SEFSC Miami
Rachael Lindsay.. SEDAR
Julie Neer .... SEDAR

Abbreviations

ABC	Acceptable Biological Catch
ACL	Annual Catch Limit
ACT	Annual catch target
ADMB	AD Model Builder (software package)
AFS	American Fisheries Society
ASAP	Age Structured Assessment Program (software package)
ASMFC	Atlantic States Marine Fisheries Commission
AW	Assessment workshop
B_{MSY}	Biomass level that provides MSY
CPS	Coastal pelagic species
CV	Coefficient of variation
DCAC	Depletion Corrected Average Catch
DW	Data WorkshopF_{MSY} Fishing rate that provides MSY
FMP	Fishery management plan
GLM	General linear model
GOM	Gulf of Mexico
GMFMC	Gulf of Mexico Fishery Management Council
HMS	Highly migratory species
MAP	Maximum a posteriori
MCMC	Markov Chain-Monte Carlo
MFMT	Maximum fishing mortality threshold
MPA	Marine protected area
MSE	Mean squared error
MSRA	Magnuson-Stevens Reauthorization Act
MSY	Maximum Sustainable Yield
NEFSC	Northeast Fisheries Science Center
NMFS	National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NPFMC	North Pacific Fishery Management Council
NS1WG	National Standard 1 Working Group
ICCAT	International Convention for the Conservation of Atlantic Tunas
P*	"P-Star"; probability of overfishing occurring
PDF	Probability density function
PIFSC	Pacific Islands Fisheries Science Center
PFMC	Pacific Fishery Management Council
PSA	Probability-Susceptibility Analysis
RW	Review workshop
SAFMC	South Atlantic Fishery Management Council
SCA	Statistical catch at age

SEAMAP	Southeast Monitoring and Assessment Program
SEDAR	Southeast Data, Assessment, and Review
SEFSC	Southeast Fishery Science Center
SPR	Spawning potential ratio
SSC	Scientific and Statistical Committee
SWFSC	Southwest Fisheries Science Center
TAC	Total Allowable Catch
TED	Turtle excluder device
VEWG	Vulnerability Evaluation Working Group
VPA	Virtual Population Analysis

I. Current ABC Control Rules

South Atlantic Fishery Management Council SSC's ABC Control Rule

Presenter: Carolyn Belcher, South Atlantic Fishery Management Council SSC

Abstract

The South Atlantic Fishery Management Council's Scientific and Statistical Committee convened a special meeting in March 2009 to focus on the development of an Acceptable Biological Catch (ABC) Control Rule for South Atlantic fish stocks. The control rule assumes that an estimate of the overfishing level (OFL), stated in weight, has been calculated and some reasonable measure of statistical uncertainty about the OFL also exists. The concept developed by the SSC is designed to be objective in nature and operates by adjusting the probability of overfishing or P* value. The control rule generates penalties or reductions based on four characteristics of stock assessments: assessment information, characterization of uncertainty, stock status, and productivity and susceptibility. The assessment information dimension reflects available data and assessment outputs. The characterization of uncertainty dimension reflects how well uncertainty is characterized in the assessment, not the actual magnitude of the uncertainty. Stock status is included among the dimensions so that an additional adjustment to ABC can be added for stocks that are overfished or overfishing. The final dimension addresses biological characteristics of the stock, including productivity, which reflects a population's reproductive potential, and susceptibility to overfishing, which reflects a stocks propensity to be harvested by various fishing gears. Each dimension has a maximum penalty of 10% associated with it. The sum of penalties is subtracted from the base case of P*= 50%, which is when the ABC=OFL. Depending on the characteristics and results of a given stock assessment, the corresponding P*, which is used to determine the ABC value, can range from 50% to 10%.

Discussion of South Atlantic ABC Control Rule

Questions and comments were centered on potential double counting of the uncertainty measure under rebuilding plans, when uncertainty measures are included both in the assessment phase and as a precautionary measure at the management level (e.g., for dogfish management the ACL was set to zero to prevent double counting). Additional comments were made concerning stock status (dimension) being a point based estimate rather than a probability based estimate. Questions were asked regarding the utility of P* system for data poor situations (e.g., Goliath grouper) with no measure of uncertainty (e.g., -10 tier score given no stock assessment). It was also noted that there is a dollar value associated with the uncertainty. All else equal, precision in science and/or management allows larger catch targets.

Gulf of Mexico SSC Report on development of an ABC Control Rule

Presenter: Richard S. Fulford, Gulf of Mexico Fishery Management Council SSC

Abstract

The Gulf of Mexico SSC formed an ABC Control Rule Working Group in July of 2009 that is comprised of members of the standing SSC, members of the GMFMC, and council staff. After review of several methods currently in use by other regional councils for setting ABC, the working group has adopted the p* approach as the primary method for determining the size of the buffer between OFL and ABC for each individual assessment. The working group is currently working to develop scoring criteria for four dimensions that will result in a range of p* between 0.05 and 0.35 that when applied to the pdf for OFL will result in a range in p(overfishing) between 0.45 and 0.15. The 0.15-0.45 range is based on guidance regarding acceptable risk provided by the GMFMC. The four dimensions currently under review by the working group are 'Data quality', 'Characterization of data uncertainty', 'Characterization of process uncertainty', and 'PSA analysis.' The exact form of the p* calculation for GOM stocks has not been finalized. In addition the working group is also working on methods for data poor species, which they characterize as any stock for which both a direct estimate AND a *pdf* for OFL are not available. In these data poor situations the working group has proposed a decision tree approach based on a method currently used by the PFMC for calculating ABC for groundfish stocks. This approach involves a sliding scale from the purely quantitative approach (i.e. p*) to a purely policy based approach (e.g. x * average catch) that allows for the use of the maximum amount of data available for any individual stock. This data poor decision tree has only been developed in preliminary form, but we anticipate a completed draft will be ready for GMFMC review by late spring 2010.

Discussion of Gulf of Mexico ABC Control Rule

The GoM SSC's control rule was presented as a work in progress. Discussion following the presentation focused on three topics. First was the range of P* considered—15% to 45%. The presenter clarified that this range was based on recommendations from the GoM Council, and was intended to apply to all species. It was suggested that the SSC might consider different ranges for different species. The SSC had previously discussed conducting a study to compare data-poor approaches and P* approaches. The second topic discussed was the choice of SPR proxies for Fmsy, and it was pointed out that the choice of proxy is a scientific decision. Managers' tolerance for risk should be reflected in target reference points, rather than through proxies for Fmsy. The third topic regarded the overfishing buffer, and whether it might be preferable to achieve a desired buffer by adjusting the variance around OFL, rather than by adjusting P* (as the SSC currently intends). The presenter acknowledged that similar results could be achieved by either approach, but pointed out that the SSC's tiered approach attempts to keep basic sources of uncertainty separated.

II. Past and Current Characterization Techniques

South Atlantic (Beaufort) Assessment Uncertainty Review

Presenter: Erik H. Williams, SEFSC

Abstract

Stock assessments for South Atlantic Fishery Management Council (SAFMC) managed fish species have been primarily conducted by staff located at the Southeast Fisheries Science Center (SEFSC), Beaufort Laboratory in North Carolina. The species assessed by the Beaufort staff through the SEDAR (Southeast Data, Assessment and Review) process are primarily composed of members of the snapper-grouper complex. Stock assessments for ASMFC managed Atlantic menhaden are also conducted by staff at the Beaufort Laboratory.

A review of the data used for the South Atlantic stock assessments indicates many areas of uncertainty including, aging error, age sampling, abundance indices, recreational landings data, historic landings data, and discard data. In some cases there are important sources of information for which the data are missing. These missing data include fishery-independent indices, shrimp trawl bycatch, spatial/depth data, and environmental linkages. Missing data and uncertainties in the data in large part drive how uncertainty is handled and ultimately expressed in the South Atlantic.

Three types of model uncertainty were considered for the South Atlantic stock assessments; structural, parameter and projection uncertainty. These sources have been managed in different ways. Structural uncertainty has been addressed in the South Atlantic by applying different stock assessment models and software packages. The list of model types used in the South Atlantic includes forward-projecting, statistical catch-at-age models, surplus-production models, stochastic stock reduction analysis, and stock synthesis 3. Typically a set of two or three of these models are applied to the same species for comparison, however the output of these multiple models is not combined to encompass among-model uncertainty.

Parameter and output uncertainty has been characterized in South Atlantic stock assessments through sensitivity runs, inverse Hessian estimates, data bootstrapping, and Monte Carlo bootstrapping. Sensitivity runs have been used primarily to illustrate model responses to various perturbations of input and model structure, but never combined to produce

comprehensive uncertainty estimates. Uncertainty estimates derived from the inverse Hessian matrix tend to be underestimates of true uncertainty because often times key parameters are fixed and the model typically includes penalty functions in the total maximum likelihood estimates. Both data and Monte Carlo bootstrapping tend to be more comprehensive in characterizing total uncertainty in model output.

Uncertainty in projection analyses in the South Atlantic have been accomplished using either simple bootstrap methods, in which only recruitment residuals are re-sampled, or as part of a full Monte Carlo bootstrap which carries uncertainty from the model fitting process forward in to the projections. An important property to consider with any projection analysis is benchmark consistency by insuring that populations reach a long term equilibrium consistent with the fishing rate being applied (i.e. fishing at F_{MSY} achieves B_{MSY} long term). Conditioning projections to conform to probabilities of overfishing (P*) is an important part of the management advice process. The SAFMC acceptable biological catch control rule relies on projections for various P* levels.

It is noted that the framework created by the re-authorized Magnuson-Stevens Act has in effect placed a monetary value on uncertainty. This arises under prescriptive ABC control rules which link reductions in ABC to levels of uncertainty. This linkage will likely result in external pressures to handle uncertainty in a much more rigorous, consistent ways compared to past practices.

Discussion of South Atlantic (Beaufort) Assessment Uncertainty Review

Uncertainty in model parameters, estimates, and how models are structured to handle uncertainty points to the need for adaptive management of fisheries and the need for flexibility in ABC control rules. Performance measures are needed to assess how well regulatory measures are working in managing fisheries and stocks in the face of these uncertainties. A feedback loop is needed to track the performance of the management of a fishery, incorporating the estimates and projections of assessments (current and past, a look at retrospective bias/assessment uncertainty) as well as the realized catches (landings and discards, etc.) resulting from the implementation of regulations (management uncertainty) of a fishery. Even for data-poor species, we should emphasize the need for performance measures to track how well assessments and management are performing.

There is a classic problem that all modelers face: How does the uncertainty in parameters interact and propagate through the model? Ideally, all sources of uncertainty should be explored and choices made on which parameters to include and which not to include in a model's structure. We should not be including sources of uncertainty into a model without adequate justification or need. For example, "environmental variation" might be suggested by weak correlations with some time series of catches or indices of catch per effort; however, this type of uncertainty should not be added into an assessment model just because of a weak but unproven relationship.

Model projections are required from assessments where stocks are determined to be overfished or are undergoing overfishing. For assessments of species in the Southeast which are overfished and for which rebuilding times exceed 10 years, projections are made for 10 years plus a generation time. In some long-lived species such as for many of our reef fish species, projections may be made out to 40 or 50 years. The uncomfortable tasks for assessment scientists in these cases are to adequately characterize the uncertainties in the model outcomes for the assessed time period as well as to provide an adequate characterization of the uncertainties over the projected future.

Characterization of Uncertainty in Recent Gulf of Mexico Stock Assessments

Presenter: Shannon L. Cass-Calay, SEFSC

Abstract:

The objective of this presentation is to introduce the methods used to characterize scientific uncertainty during past and current SEDAR stock assessments in the Gulf of Mexico (GOM). Species considered in this presentation include: vermilion snapper, greater amberjack, gray triggerfish, king mackerel, red grouper, gag grouper and red snapper. In general, uncertainties examined during GOM SEDAR assessments include: uncertainty in data inputs (e.g. discards and bycatch, release mortality, catch statistics) and parameter uncertainty (e.g. natural mortality, steepness, selectivity and catchability). In GOM assessments, uncertainty is generally explored using sensitivity runs and/or bootstrapping. A suite of sensitivity runs are often chosen to represent plausible "states of nature" and presented to the Gulf of Mexico Fisheries Management Council - Science and Statistical Committee (GMFMC SSC) for

consideration. To date, a single “base” model has been adopted by the SSC in order to develop management advice. For a given base run, complete characterization of model uncertainty is often not carried forward into the projections used for management advice. Instead, the projected stock status, yield, etc. are generally bootstrapped using only the index residuals and/or recruitment variance. To date, no attempt has been made to carry structural uncertainty (e.g. sensitivity runs) into the determination of OFL/ABC/ACL/ACT using model weighting techniques, although the SSC has discussed the merit of this approach. To date, not all GOM assessments use techniques that are ideal, or even appropriate for the methods outlined by the GMFMC SSC to determine OFL/ABC/ACL/ACT levels. However, future assessments are expected to be compliant with those objectives.

Discussion from Gulf of Mexico Assessment Uncertainty Methods

How to determine which runs are most likely

Several options exist to present which runs would be most likely. Bayesian approaches could be used, but what is typically done in the Southeast is to provide a package of sensitivity analyses that represent possible “states of nature”. To date, the GMFMC SSC has chosen a specific run as “base advice”. However, it is also possible to combine the sensitivity runs and apply P* techniques on the combined pdf (probability density function) of OFL. However, to do so the SEFSC really needs to put in place a methodology to weight runs by their likelihood. Finally, another method to account for uncertainty due to sensitivity runs is to add a buffer (i.e. SAFMC/GMFMC tiers and dimensions) to penalize stock assessment results that are highly sensitive to important model inputs/parameters.

Process of assessments in Southeast and role of SSC

A large amount of review occurs before the scenarios are forwarded to the SSC. So, what does the SSC offer that isn’t in the extensive review process that occurs before? This outcome varies by assessment. Sometimes, the review panel and the assessment team didn’t decide on a base model because they genuinely felt that all outcomes were equally likely. In these situations, the SSC has the final decision. Other times, the assessment team and review/update panel have provided a discussion that supports a particular model run. In this case, the SSC debates the support for the proposed base model. The GMFMC SSC contains members with a broad range of experience in the species and fisheries of the Gulf of Mexico. They do have additional insights that are not always represented by the Review Panel.

How to present sensitivity runs and uncertainty to stakeholders

Assessments generally contain a lot of sensitivity runs, and sometimes, runs are done for exploratory reasons. If these runs are included in the assessment report, stakeholders will sometimes take one of these sensitivity runs and try to present it as the base run, which could be a big problem. If a sensitivity run is left in the report, it is considered a plausible outcome. This information will then be used in negotiations. Clarifying sensitivity analyses as a model diagnostic would really help with this.

In addition, the public often says that because sensitivity runs are used to explore uncertainty that the scientists really don't know anything about the status of the stock. Better methods need to be developed in order to present uncertainty. In particular, if multiple runs are considered plausible, it is important to develop a protocol for combining the results with appropriate weightings. Models that are considered highly unlikely should be identified.

Characterization and Presentation of Uncertainty in Shark Assessments in the Gulf of Mexico: Data Needs and Limitations, Modeling Approaches, and Outstanding Challenges

Presenters Katie Andrews, Enric Cortés and John Carlson, SEFSC.

Abstract

Shark stock assessments are data-limited, both in the information about their life histories, and in the amount of catch and index values available. Our presentation reviews the biological modeling of life history parameters and the evolution of modeling that has occurred for U.S. Gulf of Mexico and Atlantic shark stocks. Although sophisticated modeling techniques are available, they require large amounts of data that are often unavailable or highly uncertain for shark stocks. In addition, the modeling of shark bycatch in the shrimp trawl fishery, which is often a substantial portion of the catch for some shark species, is in a state of flux. The modeling technique is being revised to incorporate the effects Turtle Excluder Devices (TEDs) have had on the total number of sharks taken. Finally, catch rates are derived from both fishery-dependent and –independent surveys. The independent surveys available for sharks are often too limited in

space or time to be representative of the whole stock and the dependent surveys are subject to high levels of criticism and uncertainty. Overall the shark assessments in the Gulf of Mexico present uncertainty in a Bayesian framework. We assign priors to important life history parameters, such as pup survivorship and steepness, based on data or expert opinion to characterize uncertainty. The model outputs include either likelihood profiles or MCMC runs to create distributions, rather than point estimates for parameters of interest.

Discussion on elasmobranch methods:

A question was asked about the shrimp by-catch estimate graph, noting the dramatic difference between model estimates. Presenters responded that the model that included the "TED effect" showed higher bycatch before the use of TEDs in shrimp trawls and then greatly reduced and stable by-catch following the introduction of TEDs. These model estimates had little or no correspondence to the shrimp by-catch estimated using the ADMB or WinBUGS versions that modeled the SEAMAP survey data without including a TED effect in the models. Presenters are working with industry to determine why the recent levels of shark bycatch seem to be too stable not reflecting expected ups and downs seen in the SEAMAP data.

WinBUGS implementations ability to fit the data as compared to the AD Model Builder version was raised. It was noted that the 2003 estimate was anomalously high in the Nichols WinBUGS model compared to the ADMB formulation compared due to a different data set. The data set Nichols used was likely different from the ADMB data set due to problems with the SEAMAP database. The full error in the bycatch estimation was not coming through in the assessment because the bycatch estimates were included as point estimates.

General discussion on SEFSC methods:

Retrospective Patterns

The group started out the discussion by revisiting retrospective patterns, an issue that had come up during a previous presentation. While some analysts in the southeast routinely look at retrospective patterns, the issue is less prominent than in other regions like the mid-Atlantic and northeast. It was suggested that part of the reason for this could be that the southeast has been more concerned with data poor species, and including retrospective patterns adds a whole new dimension of analyses to stock assessments. It was noted that retrospective patterns could be larger than the reduction due to P*, that they can occur with any kind of age structured model, and that this is something to keep in mind if more stocks are moved to being assessed through age structured models. King mackerel and red grouper are two examples of Gulf of Mexico stock assessments with strong retrospective patterns. It should be emphasized that there are within-model (sequential removal of years of data from a particular model) and between-model (comparing base models from previous assessments) retrospective patterns. Both are important to characterize. In the SE we often have a change in models during the history of stock assessments for a particular species. This complicates the interpretation of "historical" retrospective patterns, but it is still possible to look at retrospective patterns in biomass trends across a change in models.

Consistency and flexibility in assessment approaches

The group discussed the trade-offs between the ability to maintain flexibility in assessment approaches and the need for consistency in assessment methodologies between and among species and regions within the southeast. The question came up whether the differences in assessments we are seeing are based on training and history of the stock assessment scientist or on the flexibility required in doing assessments. One perspective was that a significant part of the differences is due data issues and separate histories of the fisheries. Many species in the southeast were overfished before the onset of assessments, as opposed to other regions where there is more sociological control on the fishery inputs, like in Alaska.

It was pointed out that there have been many concerns about the level of similarity in methodologies, assumptions, and parameters for assessments of the same species in the Gulf of Mexico and South Atlantic and that the pressure regarding this issue is probably going to increase over the next years. The level of consistency in assessment methodologies was discussed with respect to managers and stakeholders. It was acknowledged that there is value in having performance metrics from the management side so that assessments will be responsive to management needs. This feedback mechanism between management and assessments is very important and may lead to high assessment consistency from species to species. It was then suggested that managers are probably not as concerned with consistency among species as with consistency among assessments of the same species and that, even though improvements in assessment techniques may warrant changes in methodology, continuity is important. The group acknowledged that it is vital to clearly explain to managers and stakeholders what the rationale is for making changes in assessments. It was pointed out that changing the model in the Gulf of Mexico is generally met with opposition from stakeholders, regardless of the effect that changes have on the outcome. It should be emphasized that stakeholders in the Gulf and South Atlantic actively influence the process, for example when it comes to determining the level of discard mortality used in the assessment, and this contributes to the among-species variability in procedures and to decreased consistency by assessment scientists.

In the southeast, assessment methodologies have generally been consistent for species that are in the process of recovery, red snapper being one exception, where the question about the impact of shrimp bycatch has caused some back and forth that has not been resolved yet. It was pointed out that other regions are also battling with consistency issues and that the struggle for when to be consistent and when to innovate has been going on all over the world. The policy of ICES, for example, has been to be very consistent whereas ICCAT is not as strict about consistency when there is compelling evidence to change the model. The group noted that it would be very useful to agree on a set of best practices for consistency during the course of this workshop.

Types of uncertainty to include in stock assessments and projections

In looking for best practices to deal with uncertainty, it was suggested that one could take the uncertainty for each of the parameters and models, vary the parameters and report how sensitive the model is to each. It would be necessary to provide an objective basis for weighting the individual models, because weighting is a somewhat subjective process and depends on the individuals involved. There was concern expressed that even though a lot of time is spent characterizing uncertainty in assessments, it is often not adequately incorporated into models because there is disagreement over best practices. Where uncertainty cannot be eliminated, we must consider realistic ways of incorporating it. In the southeast, among-model variance is assessed through the use of sensitivity runs. It was pointed out that Bayesian techniques incorporate rather than try to eliminate uncertainty by including priors for parameters. It was suggested that the posterior distribution in the prior of one year could become the prior in the next year, rather than re-negotiating priors every year.

During the course of the discussion about which types of uncertainty to include in the model, it became clear that the answer to that question is not straight forward. It would be useful to start with the uncertainty for parameters that have an important impact on the probability of overfishing, such as steepness and natural mortality, but it was cautioned against being too prescriptive about the types of uncertainty to include without knowing what the specific data issues are in each case. It was pointed out that the Mid-Atlantic, rather than mandating which types of uncertainty to examine, is planning to specify in the terms of references for stock assessments that the important sources of uncertainty and their treatment should be described in the report.

The group briefly discussed implementation error, noting that this type of uncertainty should be considered if we move ahead with tasks like management strategy evaluation, to see what the impact would be in terms of yield and stability in populations. It was emphasized that it is critical to incorporate implementation error when doing any kind of forecasting, but the lack of studies that quantify implementation error was acknowledged.

There was a brief discussion about the difference between uncertainty and sensitivity analyses, and it was noted that most of the time we can only look at model sensitivity but not truly estimate uncertainty. The group expressed concern over the fact that often we do not know what the greatest sources of uncertainty are, and although we can find out what the models are sensitive to, the models may be wrong. There may be benefits in going back to the empirical method to project stock status, analogous to what the National Hurricane Center is doing in projecting the probable track of a hurricane. It was pointed out that, even though we often do not know the exact status of a stock, we generally have an idea of whether biomass is increasing or decreasing, which may be all we need to know to evaluate whether management is effective. Even though the uncertainty is often very broad, we expect that the models get more certain as we collect more data over time, and it was pointed out that recommendations made in the past often had constructive effects, even in the face of great uncertainty.

The group noted that it was difficult to quantify the dominant sources of uncertainty for an assessment because they are often interrelated and not additive, which affects not only present estimates of abundance but also historical estimates. It was questioned whether it was even necessary to identify and break out all different sources of uncertainty or if they can all be lumped into a single category because we do not know what the interaction between them is. It was pointed out that, rather than trying to exactly quantify the different parameters in a model, it might be more useful to categorize the uncertainty for each parameter in terms of whether it is unusually low, normal, or unusually high. Given the fact that we may never be able to forecast fisheries very well, it is important that we develop management systems that are robust to uncertainty. Creating a more robust system is the rationale behind the concept of MSY reference points, and this may be as important as or more important than quantifying uncertainty.

PIFSC Uncertainty Assessment Methods

Presenter: Jon Brodziak, PIFSC

Abstract

The Pacific Islands Fisheries Science Center (PIFSC) conducts stock assessments of insular and pelagic fisheries resources in the North Pacific Ocean. In this presentation, the primary methods used by the PIFSC to characterize uncertainties in stock assessments are described using specific examples. Major sources of uncertainty for the stock assessment process include: (1) model uncertainty/structural complexity, (2) estimation error, (3) sampling/observation error, (4) natural variability/process error, (5) implementation uncertainty, and (6) inadequate communication among scientists, managers, and stakeholders. The Western and Central Pacific bigeye tuna (*Thunnus obesus*) stock assessment was presented as an example of incorporating model uncertainty, estimation error, and communication uncertainty into the fishery system analysis. This bigeye tuna stock is a data-rich stock that is heavily exploited. This point was emphasized through extensive analyses of alternative reference cases (i.e., plausible assessment model scenarios) and graphical presentation of comparative results to the Western and Central Pacific Fisheries Commission, the regional fishery management organization. The Hawaiian bottomfish stock complex, an insular management unit comprised of deep-water snapper and grouper species, was presented as an example of a data-poor stock assessment in the Pacific where parameter uncertainty was estimated using Markov Chain Monte Carlo (MCMC) simulation in a fully-Bayesian stock assessment model. In addition, the implementation uncertainty about the effectiveness of new restricted fishing areas (i.e., MPAs) on this stock complex was addressed using a conceptual model of fishermen's behavior in relation to the redistribution of fishing effort inside MPAs and the potential for noncompliance with MPAs.

This analysis suggested that redistribution of effort and noncompliance could compromise the effectiveness of the new MPAs and this information was considered by the Western Pacific Regional Fishery Management Council in selecting seasonal closures to reduce bottomfish fishing mortality. The North Pacific swordfish (*Xiphias gladius*) stock assessment was presented as an example of how structural uncertainty was incorporated into the stock assessment and management advice using two alternative stock structure assumptions, the two-stock and the single North Pacific stock scenarios. This data-moderate assessment also provided an example of how implementation uncertainty was included in management advice through uncertainty in projections of future fishing mortalities and associated catches using MCMC results that were presented to the International Scientific Committee for Tunas and Tuna-like Species in the North Pacific. The North Pacific striped marlin (*Tetrapturus audax*) stock assessment provided another example of a data-moderate assessment where structural uncertainty was incorporated into management advice through the inclusion of two equally-plausible alternative stock-recruitment steepness scenarios. This data-moderate stock assessment also included evaluation of parameter uncertainty about biological reference points and steepness uncertainty through the application of model-averaging to calculate biological reference points. Overall, the characterization of stock assessment uncertainty in the North Pacific is currently accomplished by a variety of methods and is an active area of research for both insular stocks and highly-migratory pelagic species.

Discussion on PIFSC methods

Clarification was requested on what implementation uncertainty is intended to address. Dr. Brodziak responded that implementation uncertainty referred to whether chosen management measures achieved their intended outcomes. In other words, what are the effects of implementing new bottomfish restricted fishing areas when no information exists on the effects of the old restricted areas? This question is addressed by modeling fishing behavior. For example, while effort that previously occurred in the restricted area may be reduced, a certain amount of redistribution of fishing effort may occur outside the area because there is an economic rationale to continue to fish. Another concern is the level of compliance, i.e., how much fishing continues to occur inside the restricted areas. Rather than attempt to answer these questions, a range of values is provided to the Council so that they can decide for themselves what levels of redistribution and noncompliance occur. A follow-up observation was noted that the models were looking at a perfect vs. imperfect implementation. They also assumed that the changed conditions would be constant going forward in time. Trying to deal with variations in these assumptions could get complicated.

It was noted that when discussing alternatives for addressing uncertainty over the steepness of stock-recruit curves, Dr. Brodziak did not support the use of the hypothesis of environmentally-driven recruitment that fluctuates about its mean value. Dr. Brodziak explained that having no information on steepness implies a uniform prior for a steepness, which he felt was an reasonable assumption in the absence of auxiliary information. In constrast, assuming that steepness had a single value of unity (as in the environmentally-driven recruitment hypothesis) was ecologically optimistic and did not reflect the actual uncertainty about population resilience. Even at a high steepness, there is some curvature of the stock-recruit relationship that needs to be addressed in the model.

Referring to the discussion of using likelihood profiles to measure uncertainty in high dimensional models, a question was raised whether the likelihood benchmark parameters could be set in the model. Dr. Brodziak confirmed that the parameters can be assigned to have likelihood profiles in AD Model Builder. He added that the ADMB Foundation had taken over AD Model Builder Project. It is available as a free download (see http://admb-project.org/), and the Foundation is constantly working to improve this versatile optimization program.

It was noted that, with no monitoring of bottomfish restricted areas in place, the expected impacts of the areas were based on a conceptual model. Are managers satisfied with the conceptual model of closed areas or whether they were going to put monitoring programs in place? Dr. Brodziak felt that managers are concerned with overregulation and with the need to be able to account for any gains from the restricted areas. He noted that monitoring programs have not been put into place, but the models show what could happen and they identify what parameters need to be monitored if such studies are done. He added that it does appear that the restricted areas are having an effect. The seasonal closures are effective and simple to enforce, hence increases in imports of foreign bottomfish have occurred. Also, based on a subsequent stock assessment, some reduction in fishing mortality has occurred.

A question was raised as to whether managers have made use of the unfished parallel universe analyses that Dr. Brodziak described. Dr. Brodziak responded that he has not received any feedback from the managers, and did not know if managers were using the parallel universe analyses in their decisions. He noted that these analyses were a device to show what could happen if fishing mortality was reduced, and that it can show the potential variability from environmental factors. It was added that the unfished parallel universe analysis also demonstrated the potential effects of migration occurring between areas and provided a model diagnostic for the reasonableness of movement parameters.

NEFSC Assessment Uncertainty Methods

Presenter: Chris Legault, NEFSC

Abstract

The NEFSC assesses 50 species comprising over 60 stocks using a number of standard assessment models from the NOAA Fisheries Toolbox mainly. The choice of model usually depends on the available data, with data limited stocks using AIM, CSA, or SCALE while data rich stocks use VPA or ASAP, generally. Uncertainty in the assessment models is estimated either through bootstrapping residuals of index fits (AIM, CSA, VPA) or through Markov Chain Monte Carlo methods (SCALE, ASAP). The model results are often used as uncertain starting points for projections using AgePro, which also incorporates uncertainty in future recruitment. Yield per recruit is often used to set the F reference point while AgePro is used to project this F many generations into the future to estimate MSY and Bmsy, such that the reference points and projections will be consistent.

Time series plots and probability density functions are commonly employed to graphically display the level of uncertainty in assessments. In some cases, risk plots are generated which show the probability of exceeding an F reference point or of biomass not increasing by a given amount for different levels of projected catch. Some stocks have multiple models recommended for use with management decisions when a single model cannot be agreed upon. These multiple models demonstrate the among model uncertainty, but can lead to difficulties in setting quotas.

Retrospective patterns have become a standard diagnostic for NEFSC stock assessments and cause an additional level of uncertainty when present. Not all stock assessments in the Northeast exhibit a retrospective pattern, and not only one model type will exhibit a retrospective pattern. Splitting the survey time series has been identified as one possible “fix” to retrospective patterns and preliminary MSE results demonstrate that it is effective even when misreported

catch or changes in M are the real cause of the retrospective pattern. Comparison of alternative retrospective fixes, such as modifying the results of the base model by the estimated amount of retrospective patterning, often result in similar stock status and catch advice.

A new uncertainty has been introduced to stock assessments in the Northeast by the deployment of a new research vessel using new trawling gear. Calibration coefficients are being estimated to allow linkage of the data collected with this new system to the 40+ year time series. Currently beta-binomial conversion coefficients have been estimated for catch/tow in numbers and weight. Exploration of length-based conversion coefficients is continuing. One advantage of this exercise has been the ability to set a prior on the catchability coefficient for the old vessel and gear based on the conversion coefficient. This prior is used with minimum swept area abundance estimates from the survey to help determine the magnitude of population abundance. This example demonstrates that there will always be new sources of uncertainty in stock assessments over time and the analysts must be prepared to address them as they are encountered.

Discussion of NEFSC methods:

It was pointed out that there must be lots of "0s" in the survey data, and it was questioned how these were handled in VPAs. Chris Legault responded that there were, in fact, many zeros although these showed up particularly in the youngest and oldest age classes. In analysis, the zeros are treated as missing values. Some have suggested inserting a small fixed value in place of the zeros but this does not seem to be a better fix; this tells the model that nothing is happening during that period. Simulations have demonstrated that this "filling zeros" approach can introduce bias. Treating these as missing values seems to be the best alternative but they are still looking for a better solution. GLM on surveys was suggested to smooth out the data. To date, options for filling zeros have been criticized and peer reviewers tend to criticize and reject them in favor of simply using observed data without alteration. In addition, final answers seldom change appreciably regardless of the method used to fill in zero observations.

Differences in biomass reference points derived from deterministic and stochastic projections were noted, and the desire stated to see consistency between projections and reference points, meaning that if you project a stock with F at F_{msy} then stock should end up at B_{msy}. There was no argument that this would be the most desirable situation.

It was noted that in one situation predators had been included in the model as "fleets", and a question raised whether this was important in affecting M, in contrast to F normally associated with fishing. Chris replied that this hasn't made a big difference in the model; the change was not as substantial as expected. The pattern was really not different although a limited number of cases have been examined so far.

The MSE work was discussed, in particular whether there are ways to get better estimates of fishing mortality while simultaneously getting estimates of biomass. It was noted that SSB was more biased with the split survey series approach. Increasing natural mortality, perhaps by using 0.6 in projections when the assumed value is 0.2, gives an incorrect calculation of catch but a potentially more preferable answer for management advice.

It was asked whether there were standard approaches in Northeast Region for setting ABC, including means to address scientific uncertainty in the assessments. The short answer was "yes" although many stocks are in rebuilding mode and this doesn't really apply.

SWFSC Assessment Uncertainty Methods –
Presenter: Ray Conser, SWFSC.

Abstract

The Southwest Fisheries Science Center (SWFSC) carries out stock assessment research in four general areas: species managed under Fishery Management Plans (FMP) developed by the Pacific Fishery Management Council (PFMC); Pacific highly migratory species (HMS) under the purview of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific (ISC); Protected Species (marine mammals and sea turtles); and Antarctic Species (krill, crab, and finfish including effects on seals and seabirds). This paper focuses on the first of these areas as it most relevant to the terms of reference for the SEDAR Procedural Workshop on Uncertainty.

The PFMC has four FMPs: Groundfish, Salmon, HMS, and Coastal Pelagic Species (CPS). Stock assessment scientific support for the PFMC is provided jointly by the SWFSC and the Northwest Fisheries Science Center (NWFSC). Scientists from both Centers conduct stock assessments and serve on the PFMC SSC and its FMP-specific management teams. A major research effort over the past year has focused on analyses needed to support the amendment of

all FMPs so that they conform to the new National Standard 1 (NS1) guidelines. This work – led by the SSC – has mainly dealt with species in the Groundfish and CPS FMPs to date.

The NS1 guidelines require specification of the annual overfishing level (OFL) catch and an acceptable biological catch (ABC) for each managed species. ABC must be less than OFL – the difference being a function of the scientific uncertainty and the probability of overfishing. Determination of the OFL buffer (ABC/OFL) is a joint SSC and Council responsibility. The SSC quantifies the scientific uncertainty (σ) and the Council decides the acceptable probability of exceeded OFL (p*). The latter must be less than or equal to 0.50 but otherwise is driven by the Council's level of risk aversion, cost factors, and other management related issues. In all cases, OFL buffers should be larger for data-poor stocks than for data-rich stocks.

The SSC has conducted analyses aimed at i) estimating the level of σ for a wide variety species in the Groundfish and CPS FMPs and ii) developing an algorithm that determines the OFL buffer as a function of σ and p*. The details are summarized in a companion paper in this volume (*An Approach to Quantifying Scientific Uncertainty in West Coast Stock Assessments*).

Preliminary results show that for 16 groundfish and coastal pelagic species, the mean σ on terminal biomass is 0.19. This represents the average amount of statistical measurement error within assessments conducted for the PFMC. In contrast, the average σ ascribable to model specification error (i.e. among assessment variation) is 0.34, which is the far greater of the two sources of uncertainty. An example from the preliminary results follows: if only among-assessment variation is considered; if the variance in F_{MSY} is ignored; and if p* is fixed at 0.40, the OFL buffer would be 0.92, i.e. an 8% reduction in harvest from the OFL level. Smaller p* and/or introduction of additional variance components (e.g. within assessment or F_{MSY} variance) would reduce the harvest further.

The PFMC SSC's developmental work on this method is continuing.

Discussion of SWFSC methods

The performance of depletion-based stock reduction analysis with short-term series was discussed. The method has not been formally tested but it would be expected to perform poorly on data-poor, depleted stocks, or on ones with large amounts of natural fluctuations. It likely works best for multi-generationally surveyed, data-rich stocks. This led to further discussion of what amount of data designates a stock data-poor or data-rich. In the Southwest, generally, groups of species that are regularly assessed are dealt with using stock synthesis and therefore possess the corresponding required data (catch histories, size comps and indices of abundance). Such stocks are considered data-rich, while data-poor stocks lack that kind of information.

Discussion was raised regarding the practical problem of increased precision, perhaps in exchange for bias, that comes from the simpler models typically used when less data are available, and whether the Southwest has considered expanding variance estimates for simpler models. This was acknowledged as an ongoing, difficult problem. Having higher precision when simpler models are used is a within assessment problem. Sigma values from more complex assessments have been used, treated as a benchmark for simpler models. The range of sigmas across assessments is not vastly different regardless of the models used, but analysts in the region do keep sigma benchmarks in mind. Work is ongoing to determine if there is a significant enough difference between sigmas outputted from simple vs. complex models.

Uncertainty and management in Alaska

Presenter: Dana H. Hanselman, AFSC

Abstract

Alaska fisheries management is a fundamentally conservative framework that sets quotas based on a set of tiers of data availability. The amount of data available is directly related to the amount of uncertainty in those stocks. The framework also relies on the "lowest common denominator," meaning that any Overfishing Limit (OFL) that is exceeded can constrain the rest of the fisheries that catch that stock. The North Pacific Management Council (NPFMC) administers this framework for groundfish on an annual basis under the axiom that fisheries are only allowed to develop only when sufficient data has been collected. Uncertainty enters the process at several levels. In the tiers with statistical age structured models (1-3), uncertainty is accounted for in the modeling process through the error assigned such things as survey biomass estimates, ageing error, catch error, and prior distributions assigned to key parameters. In Tier 1, acceptable biological catch (ABC) is assigned explicitly based on uncertainty. This is done by using the harmonic mean of the fishing mortality rate and the geometric mean of exploitable biomass. These means are always lower than the arithmetic mean, and how much lower is

determined by the variance of model outputs. Only a few of the most data rich stocks are in this category where a stock-recruitment relationship and MSY are estimated. Most of the target stocks in Alaska are administered under Tier 3, where proxies for MSY and OFL are determined by spawners-per-recruit reference (SPR) reference points. In this tier, uncertainty in those reference points is compensated for by setting maximum ABC at an SPR rate below the MSY proxy rate. This usually yields a buffer from 15-20% below the MSY proxy. For Tiers 1-3, a control rule is in place that reduces fishing mortality when stocks fall below reference points.

In general non-target stocks are located in Tiers 5-6. These species are caught in fisheries, but are considered non-target or incidental catch species. In Tier 5, OFL is set at F=M and in Tier 6, OFL is set to average catch for a specified period. The reference points buffer against uncertainty by setting ABC at 0.75 these levels. Generally, a fishery is not allowed to develop on these stocks until sufficient data is collected to increase their Tier level. Retrospective analyses are conducted to further assess model uncertainty. Projections are done stochastically including recruitment uncertainty, as well as full posterior estimates of projection uncertainty. Ecosystem considerations are often taken into account as a source of uncertainty, and are commonly used to provide additional inputs to the quota setting process as support for precautionary reductions.

For groundfish management in Alaska, the current buffers and control rules may be sufficient to meet the spirit of National Standard I guidelines, but further analysis may lead to changes to increase the level and rigor of scientific uncertainty included in these buffers. One way to do this which would apply to nearly all Alaska stocks is to relate the uncertainty of OFL to survey uncertainty for use in the *P** method. For crab and scallop management, no ABC control rule is currently in place and the *P** method is being proposed to create this rule. Different levels of uncertainty for OFL are proposed by using strictly model uncertainty, or an additional level added in from retrospective analysis or more *ad hoc* approaches of adding a constant like 0.2 or 0.4 to the coefficient of variation. Future work will include management strategy evaluations to determine the robustness of current and proposed uncertainty methods.

Discussion of AFSC methods

It was questioned whether the council responded favorably to the inclusion of ecosystem uncertainty. The councils were receptive especially after they were shown predator (arrowtooth flounder) interactions with the stock. The council supported more conservative quotas from the typical million metric tons a year because of that additional uncertainty.

It was noted that in the comparison of different species with P* = 0.12, the Gulf of Alaska Harlequin was so far off. The Gulf of Alaska Harlequin is the worst sampled species in the trawl surveys with CVs of 0.80 to 1.00 while the CVs for the other species were more reasonable being around 0.30 or less.

It was questioned whether any species were actually managed with TACs determined by average catches (Tier-6), and would using average catches ultimately lead to a TAC of zero as management reduces catches. Several species are managed in Alaska as Tier-6 species but their average catches were based on a fixed of number of years, usually 1977-1995, so the average catch does not change as data from recent years are added.

Clarification was requested on the Dorn constant buffer adjustment. At high biomass levels, the safety margin is as expected but as the biomass goes below the target, the ratio of the safety margin to the F limit decreases. The Dorn constant buffer adjustment is to maintain the proportion of the buffer to the fishing mortality limit providing a consistent buffer. An alternative approach was suggested, to make the control rule parallel instead of the reduced proportion of buffer. It was replied that if a parallel rule were to be adopted, then the quota would be smaller and the fishery would close earlier.

Probability-based catch levels: Scurrying to satisfy NS Guidelines

Presenter: Kyle W. Shertzer, SEFSC

Abstract

In U.S. federal fishery management, acceptable biological catch (ABC) is set below (or equal to) the overfishing limit (OFL) to account for scientific uncertainty, and annual catch targets (ACTs) are set below (or equal to) the ABC to account for implementation uncertainty (i.e., imperfect management control). Probabilistic approaches have been proposed previously for setting target and limit reference points in fishery management. In this talk, we describe two adaptations to those earlier approaches designed for better consistency with recent revisions to

the National Standards Guidelines. One adaptation is intended for setting ABC for a single year, the other for setting ABCs and ACTs over multiple years.

Discussion of p* approach

In the presentation, P** is defined as the allowable probability that catch from an ACT will exceed the OFL. From a management standpoint, it may be more useful if P** is defined as the probability of exceeding the ACL; exceeding the ACL would trigger accountability measures, something that the Councils would like to avoid. Under the new definition, P** would no longer be related to the probability of overfishing. Rather than redefining P**, the motivation for the redefinition could be accomplished by using the sequential method and setting ACL equal to the ABC.

In multiyear projections ABC only needs to be calculated to initiate the projection, but it is not necessary to get to the next year and each succeeding year of the projection in order to make the process work. It would be needed, however, in the sequential method, where ACT is derived from ABC.

The equation presented on slide #9 (Prager et al. (2003) Approach) assumes the factors are independent, but covariance could be added. Covariance has been examined in this context and was found to have little importance. Use of ratios in the equation removes a lot of the covariance. It is not clear how the multi-year approaches discussed in the presentation would be applied to data poor cases.

Major sources of Uncertainty in the assessments of ICCAT species

Presenter: Mauricio Ortiz, SEFSC

Abstract

Since 1965, the ICCAT regional fisheries management organization has assessed the main tuna and billfish species of the Atlantic Ocean and Mediterranean Sea including: bluefin, bigeye, yellowfin, albacore, swordfish, skipjack, blue and white marlin, and sailfish. Recently, the ICCAT Standing Committee on Research and Statistics (SRCS) has initiated assessments of other pelagic species including sharks and bycatch species that interact with the main gears of tuna operations, such as seabirds.

Main sources of uncertainty in the assessment of these species can be classified into:

Catch data

- Directed Landings
- Non-targeted/Bycatch (Uncertainty generally greater than for the directed landings)

Biological and population information

- Limited biological sampling/studies with low coverage in space and time
- Low sampling for size/age composition
- Very limited scientific or non-fishery surveys

Assessment Models

- Models used by the SCRS include, Age structured analyses (restricted to main tuna species) and surplus production models (data-poor species). Recently, the SCRS has introduced Statistical Catch at age models (Multifan-CL, Stock Synthesis) but still they are **not** the main models for management advice
- Structural uncertainty is usually assessed using sensitivity analyses of alternative model formulations
- Within-model uncertainty generally is assessed using bootstraps, particularly from tuning indices. Annual trajectories of stock status are generally illustrated using phase plots base on the bootstrapped results.
  Stock projections
- Stock recruitment assumptions for short term projections and definition of stock benchmarks
- What Sources of uncertainty of the final model should be carry over
  Management implementation
- ICCAT convention objective is MSY but there is not a formal distinction between a target objective and limit threshold(s).

The SCRS has reviewed these sources of uncertainty, and make recommendations to the Commission that resulted in specific task to reduce uncertainty. For example, improvements in data submission and compliance, by providing economical and technical assistance to developing countries to improve data collection and sampling, initiation of species-specific programs for biological studies, improved observer sampling of major fisheries in the Atlantic, the creation of a Precautionary Approach Working Group, and the adoption and implementation of methodologies used by other RFMOs to communicate uncertainty in assessment results. The Commission is also revising its objectives in order to adopt a more risk adverse policy of fisheries management. The presentation provided an example of a preliminary evaluation of potential limit benchmarks applied to the northern albacore stock and the implications of catch under-reporting uncertainty in the projections of eastern bluefin tuna.

Discussion on ICCAT methods

There was time for only one question, which concerned using a probability transition matrix for projecting recruitment, and how would one smooth out the bins in the probability transition matrix for which there were no SSB and corresponding recruits. It was suggested that one consider collapsing or joining bins to reduce or eliminate these empty bins.

An Approach to Quantifying Scientific Uncertainty in West Coast Stock Assessments

Presenter: Clay Porch, SEFSC.

Abstract

Quantifying scientific uncertainty in estimating an appropriate catch level for a fish stock is challenging. Multiple sources of error can easily be identified, including measurement error that is conditioned on the adopted model, model specification error, forecast error, and uncertainty about overall stock productivity. In addition, there are without doubt other unknown factors that will negatively influence the precision of scientific advice on catch levels. Notwithstanding these difficulties, the Magnuson-Stevens Reauthorization Act (MSRA) identifies the quantification of scientific uncertainty in the development of advice on catch levels as a key requirement of the new law. Moreover, the Scientific and Statistical Committees (SSCs) of the Regional Fishery Management Councils have been given the responsibility to quantify that uncertainty.

While many sources of uncertainty exist, the focus here is on quantification of statistical measurement error and model specification error, particularly the latter. While not all inclusive, the study of these two factors is feasible with the information that is currently available. They are also likely to include the dominant sources of scientific uncertainty in the development of scientific advice vis-a-vis groundfish and coastal pelagic species catch levels at the Pacific Fishery Management Council.

Although full Bayesian integration through MCMC calculations is a preferred method of estimating measurement error "within" a stock assessment, an inadequate number of studies have

successfully achieved that type of analysis. Consequently, we report the first order approximate estimates of the standard error on terminal biomass from stock assessments that are calculated by inversion of the model's Hessian matrix (i.e., the asymptotic standard error). To summarize variation "among" stock assessments, as a proxy for model specification error, we characterize retrospective variation among multiple assessments of the same stock.

Preliminary results show that for 16 groundfish and coastal pelagic species the mean of the coefficient of variation on terminal biomass is 0.19. This represents the average amount of statistical measurement error within assessments conducted for the PFMC. In contrast, the average coefficient of variation ascribable to model specification error (i.e., among assessment variation) is 0.34, which is the far greater of the two sources of uncertainty. Given the preliminary results, if only among-assessment variation is considered; if the variance in F_{MSY} is ignored; and if the probability of overfishing is fixed at 0.40, an appropriate OFL buffer on the overfishing catch level is to reduce the harvest by approximately 8%. Smaller acceptable probabilities of overfishing would reduce the harvest further. The PFMC SSC's developmental work on this method is continuing.

Discussion on PFMC methods:

There was discussion on addressing difference in model variances, selecting suites of models, distinguishing between model runs, and determining which model or run is most reliable. There are no clear answers to most of these questions. Analysts must consider each circumstance and evaluate the various outcomes. There was further discussion on distinguishing the variance from within models or between software, and noted that software issues can add a new dimension of uncertainty for consideration.

The total variance estimation procedure was discussed, in particular whether total variance remains constant over time and whether there is any new learning. Ideally, assessments improve over time, as analysts gain skills, methods advance, and data improve. However, examination of time series plots indicates it is not apparent that the variance is decreasing. This may be because of new factors previously unaccounted for such as review panels and different stock assessment teams. However, while these factors may have some influence, overall assessments are getting a lot better.

Beaufort Assessment Model Approach

Presenter: Paul Conn, SEFSC.

Abstract

In this talk, I review several methods used by NMFS-Beaufort assessment scientists in recent SEDAR assessments and outline approaches under consideration for future assessments. Throughout the talk, I concentrate on uncertainty conditional on a given model structure and dataset (thus, additional uncertainty attributable to differences between models and to alternative data streams are beyond the scope of this presentation). Uncertainty in previous SEDAR assessments has primarily been addressed using bootstrap-based approaches. In particular, most assessments have used bootstrapping of spawner-recruit residuals to account for uncertainty in management benchmarks (e.g., F_{MSY}, B_{MSY}). Alternatively, several assessments (e.g., snowy grouper, red grouper) have incorporated the Monte Carlo bootstrap, which combines a data bootstrap with a Monte Carlo procedure that accounts for uncertainty in parameters that are modeled as fixed parameters within stock assessments (e.g., natural mortality, discard mortality). The latter procedure assigns a prior distribution to fixed parameters; variation in model runs with 'fixed' values sampled from their prior distributions results in additional uncertainty associated with these parameters. Using a recent assessment model (SEDAR 19 red grouper), these approaches are contrasted with other possible approaches for accounting for uncertainty, including asymptotic, Hessian based methods, as well as Bayesian approaches (including maximum a posterior [MAP] estimation). This comparison illustrates how inclusion of additional sources of error can drastically impact the variance associated with distributions of management benchmarks. Finally, I describe how a recently developed method ('inverse prediction') can be used to estimate uncertainty distributions when a simulation study is used to relate estimated assessment parameters to those used to generate data. In this case, the assessment model is regarded as a 'black box' and the relationship between true and estimated quantities is estimated empirically. When a real life assessment with the same structure and data sources is used to estimate parameters of interest, the results of this experiment can be used to calibrate estimated values to true, unknown values (thus accounting for bias), and to also estimate uncertainty about these parameters.

Discussion of Beaufort Approach

It was noted that another study of coverage of different methods of estimating uncertainty had been conducted by Arni Magnusson and that this was presented at the Anchorage AFS meeting. Details of the analysis may be in his University of Washington Dissertation. It was clarified that the bootstrapping procedure starts with the residuals from the stock-recruitment model from the assessment. The residuals from the S-R relationship are used to create new data sets and the parameters of the S-R function are re-estimated to produce a distribution of S-R parameter estimates.

Productivity-Susceptibility Analysis: A risk based framework to evaluate species vulnerability".

Presenter: Robert Wakeford, MRAG Americas.

Discussion on PSA presentation

The group discussed some of the alternative approaches in use to evaluate productivity and susceptibility. It was noted that the ICCAT Shark Working Group applied a Level 3 analysis that provided a quantitative product using Monte Carlo simulation for the productivity component and considered susceptibility quantitatively. NMFS convened a Vulnerability Evaluation Working Group (VEWG) that developed a PSA framework that differs slightly from the MRAG approach presented here. One area in which the MRAG and NMFS VEWG approaches differ is in regards to unknown or poor quality data. For example, some of the data for sharks that can be readily obtained from sources such as FishBase are considered incorrect and not used in the NMFS VEWG analysis. The approaches also differ in treatment of missing information, with the NMFS VEWG approach leaving fields for missing information blank and the MRAG approach scoring missing information at the highest risk level. The rationale by MRAG for giving a high score (scoring as high risk) to an attribute with missing or inappropriate data was to flag the need for improved data quality for the species. The rationale offered by the NMFS VEWG group for omitting missing attributes is that doing so over-inflation of risk scores when there are multiple unknowns.

There was discussion of the weighting scheme for gears, with indication that this could be revisited. Biological characteristics of catch can change. Productivity usually does not change but susceptibility likely would when sectors or gears change.

It was suggested that greater buffers may not always be required for low productivity species, as highly productive species usually exhibit more variability.

Uncertainty of Stock Assessments and Consequences for P Methods*

Presenter: Michael Wilberg, University of Maryland.

Abstract

Introduction

The Mid-Atlantic Fishery Management Council is currently developing a tiered approach for setting ABCs with its Scientific and Statistical Committee. Two of the currently proposed tiers involve ad hoc determination of the distribution of overfishing limit (OFL) with the SSC deciding if the OFL distribution decided by the stock assessment satisfies the best available science criterion. This requires that the SSC has a technique to evaluate the quality of an estimate for the uncertainty of the OFL.

One potential source for information on the uncertainty of stock assessment estimates is simulation studies that have been conducted. Most of the stocks in the Mid-Atlantic are assessed using statistical catch-at-age (SCA) or statistical catch-at-age and length models. Several simulation studies have been conducted to evaluate effects of data availability, quality, and model structure on the quality of assessment estimates, which can be used to inform the expected variance of estimates from these types of stock assessments. In particular, these types of studies will often provide a "best case" scenario because the stock assessment model structure is usually the same as or very similar to the model that generated the data. In particular, some parameters, such as the natural mortality rate, were assumed constant over time and their correct value was used in the assessment, the correct form was used for selectivity, and the correct error distributions and variances and effective sample sizes for error distributions were correct.

Methods

I conducted a review of several simulation studies of performance of SCAs (Table 1). I used reported coefficients of variation (CVs) of estimated biomass in the last year of an assessment. Yin and Sampson (2004) reported this in their paper. Bence et al. (1993) reported the proportion of assessment models where the estimate was within 20% of the true value, and I converted this value to a CV by assuming that the results were lognormally distributed. For Wilberg et al. (2006, 2008) I used the original results (not reported in the papers) to calculate the CVs. Relative errors and the range of the 80% interval were presented in Punt et al. (2002). Results from two studies, Labelle (2005) and Magnusson and Hilborn (2007), were primarily presented graphically, so CVs of estimates were not determined. However, their estimates in terms of magnitude in errors and skewed distribution of errors were consistent with the ones presented.

We are interested in the distribution of OFL, which is the product of biomass and the limit fishing mortality rate (MFMT). Thus, error in biomass estimates will underestimate the error of the OFL because it does not include error in MFMT. In addition, the last year included in most assessment models is often several years earlier than the one for which ABC must be set. Thus, projection error is also not included.

Table 1. Purpose of simulation studies.

Study	Description
Bence et al. 1993	Determine effects of survey characteristics on SCA estimates
Punt et al. 2002	Determine likely performance of several assessment techniques under a range of data generating scenarios for southern Australia fisheries
Yin and Sampson 2004	Determine effects of data, fishery, and stock characteristics on SCA estimates
Labelle 2005	Determine performance of MULTIFAN-CL in several cases
Wilberg and Bence 2006	Compare alternative methods for estimating time-varying catchability in SCAs
Magnusson and Hilborn 2007	Determine characteristics that make fisheries data informative in SCAs
Wilberg and Bence 2008	Determine performances of deviance information criterion for selecting among SCAs that differ in their random effects

I also compared the effects of underestimating uncertainty in the OFL on the size of the buffer between OFL and ABC for several levels of CV and lognormal and normal distributions.

Results

The CV of estimates of biomass in the last year varied among the alternative studies and among scenarios considered within each study (Table 2). In general, low fishing mortality rates, high survey CVs, and not having a good index of abundance for older age classes led to higher CVs in estimates of biomass in the last year. The overall mean CV in estimated biomass in the last year among the studies was 47%.

Table 2. Mean, minimum, and maximum coefficient of variation (CV) of biomass in the last year of the assessment for studies where the estimation model was similar to or the same as the data generating model.

Study	Mean CV (%)	Minimum CV (%)	Maximum CV (%)
Bence et al. 1993	60	14	183
Yin and Sampson 2004	35	9	94
Wilberg and Bence 2006	65	17	407
Wilberg and Bence 2008	31	14	48

When the assessment model was substantially different from the data generating model, assessment results could become quite biased (Table 3). However, level of bias and variability of the assessment model results depended on the conditions simulated for each species.

Table 3. Mean relative error (RE) in biomass and width of the 80% interval in relative error in biomass from SCA models for five simulated species from Punt et al. 2002.

Species	Mean RE (%)	Width of 80% interval (%)
1	224	1074
2	-39	40
3	-32	95
4	-67	24
5	72	87

Underestimating uncertainty in the OFL distribution could lead to large differences in the buffer between OFL and ABC depending on the difference in uncertainty, the acceptable probability of overfishing (P*), and the distribution of the OFL (Figures 1 and 2). The difference between levels of CV often increased as P* decreased and was higher for the normal distribution than the lognormal distribution. An important consequence of this property is that the meaning of P* changes if the true distribution of OFL is not used. In particular, P* no longer represents the acceptable probability of overfishing, but rather is simply another parameter in an ad hoc control rule for buffering ABC from OFL.

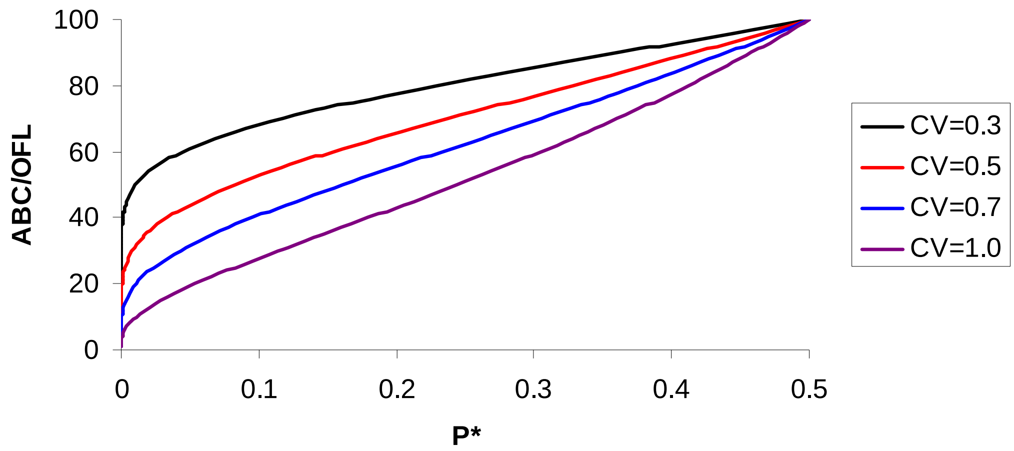

Figure 1. Buffer between the overfishing limit (OFL) and acceptable biological catch (ABC) for a lognormal distribution for four levels of CV.

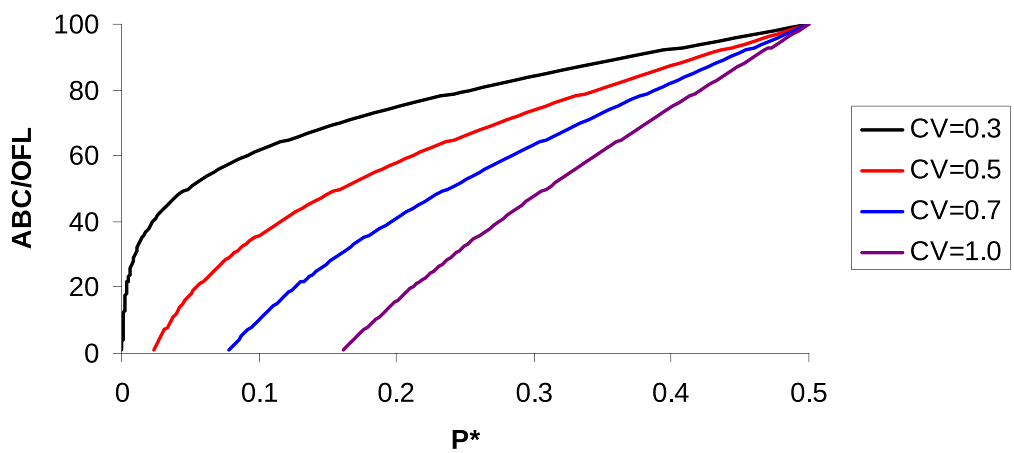

Figure 2. Buffer between the overfishing limit (OFL) and acceptable biological catch (ABC) for a normal distribution for four levels of CV.

Communicating uncertainty

Often, the process of communicating these topics focuses on reduction in acceptable catch from OFL. This can produce the idea that some potential catch is foregone. While this may be true in the very short term, it is likely not true in the longer term, and scientists need ways to communicate this fact to the Councils and to stakeholders. One potential method for communicating this may be to use financial examples, such as stock market investments, to explain why fishing at the OFL level will often be unsustainable. Management strategy evaluation (MSE) can be a valuable tool to evaluate long-term tradeoffs in performance measures from different management approaches.

References

Bence, J. R., A. Gordoa, and J. E. Hightower. 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. 50: 827-840.

Labelle, M. 2005. Testing the MULTIFAN-CL assessment model using simulated tuna fisheries data. Fisheries Research 71: 311-334.

Magnusson, A. and R. Hilborn. 2007. What makes fisheries data informative? Fish and Fisheries 8: 337-358.

Punt, A. E., A. D. M. Smith, and G. Cui. 2002. Evaluation of management tools for Australia's south east fishery 2. How well can management quantities be estimated? Marine and Freshwater Research 53: 631-644.

Wilberg, M. J., and J. R. Bence. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences 63: 2275-2285.

Wilberg, M. J., and J. R. Bence. 2008. Performance of deviance information criterion model selection criteria in statistical catch-at-age analysis. Fisheries Research. 93: 212-221.

Yin, Y., and D. B. Sampson. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. North American Journal of Fisheries Management 24: 865-879.

Discussion of Uncertainty of Stock Assessments and Consequences for P* Methods

Most of the discussion revolved around the utility of the financial example presented (slide 19). The public understands money – maybe not fiscal management, but money, so it may help to convey the concept in court of public opinion. However, Councils and Congress know the issues, but have shorter term objectives, and a 50% reduction in F translates to a 50% reduction in employment. So while society's discount rates may be fairly small, and the benefits to society for reducing F are demonstrable, they are not so for the individual. Those types of tradeoffs are seen at the Council levels. But they are not typically understood in that way by the general public. This type of presentation may help communicate the issues to the public.

A note was made that this is a single realization, while uncertainty represented in stock assessments is often displayed as a distribution. If a similar process was used in the example, there would be an envelope around the trend line, that might have a mean about the no change over time value, but since there are time trends in the way that realizations developed, a decreasing pattern emerged in the example used. That could help link the example to how management advice is provided.

One comment was to perhaps use the same sort of simulations to look at the benefits of leaving some of the increase "in the bank", rather than allowing harvest. The variation in the financial example is not dissimilar from the types of variation seen in yields in stocks under management. This could show the effects of various forms of management on rebuilding stocks. The harvesters will have a hard time understanding that as stocks rebuild, harvesters may be faced with shorter seasons and more restrictions. This is especially true in seriously depleted stocks, where there is often strong demand to return to fishing the stock before it has had a chance to really recover.

In addition to these caveats, many SE US fisheries are largely recreational. Fifty-one percent of all recreational US fishing happens in the SE. These fisheries will have different objectives than making money. Even in the charter industry, how much they make is not so dependent on the amount of fish they catch as on being able to satisfy their customers. So they are more concerned with length of the season, etc.

That leads into the idea for looking at some broader management strategy evaluations that look at a broader set of performance measures as to how a policy is expected to actually perform. So recreational fisheries may not be out to maximize yield, but to maximize access to the resource, maximize catch per unit effort, maximize chance for a large fish, etc. That will often argue for lower fishing mortality rates than you would expect if you were trying to maximize yield. That leads to why some of these simulation techniques might be used - to look at a wider variety of performance metrics than in the past, so that people can evaluate a wider variety of ideas for their consideration. Also need to consider such factors as stability of a fishery over time.

NMFS NS1 Working Group: Getting the ball rolling

Presenter: Michael Prager, SEFSC.

Abstract

NMFS formed a National Standard 1 working group (NS1WG) in January, 2008, with the aim of providing technical guidance for SSCs on implementing new NS1 guidelines — guidelines that had not yet been finished by NMFS. The NS1WG, with members representing all Science Centers and NMFS Headquarters, held a series of meetings and conference calls, in which we studied and discussed **draft** National Standard 1 guidelines and brainstormed ideas for meeting them. We also sent extensive comments to NMFS Office of Sustainable Fisheries, which was developing the final NS1 guidelines. Final guidelines were released in late December, and by that time, many SSCs were well on the way to devising their own procedures. Although the NS1WG did not produce a report, it had considerable influence through contributions of its members to SSC discussions. Some of the ideas discussed or developed at its meetings were the applicability of various P* (probabilistic) methods; development and testing of MacCall's DCAC (depletion-adjusted average catch); adjustment of NW decision-theoretic methodology to meet requirements of the MSRA; and correspondence between optimum yield and catch levels at ACT (annual catch target). The NS1WG's value, then, was in educating a cross-section of NMFS population dynamicists and in fomenting discussion of NS1 implementation issues, thus getting the ball rolling for SSC discussions.

Discussion on NMFS NS1 working group:

The National Standards working group met several times and reviewed various approaches for setting ACLs. Although there was no actual report, the group discussed p*, depletion methods for data poor stocks (MacCall), and talked about Grant Thompson's decision theoretic approach. Individual SSCs seemed to be making a lot of progress on their own, however, so the working group didn't see many tangible benefits from regurgitating these methods in a report. There was good discussion and synthesis, however, and good involvement from multiple regions.

Although there was discussion, there was no suggestion by the ACL working group to adopt uniform implementation of guidelines across regions. Rather, they thought it best to rely on the creativity of local groups that are trying to develop different approaches. They felt that it was way too early in the process to say what best practices are since many of these approaches are still under development. However, as time goes on it will be important to evaluate the different approaches (e.g., by looking at successes and failures of management). This will help in providing guidance on whether to standardize methods in the future.

General discussion on methods

General discussion opened up with how to define the structure of the discussion. Topics of interest included: determination of the P* level and how the constructed pdf for OFL translates to landings, looking at the uncertainty associated with the binary decisions of overfished/overfishing, issues of data rich and data poor, evaluation of the overall process, model uncertainty, empirical versus model-based approaches, focus at the data level, issues with within- and among-model variation, as well as retrospective issues.

Much of the discussion focused on the use of empirical vs. model based approaches. Questions were raised about whether or not empirical approaches were applicable to the South Atlantic, as well as the overall utility of the approach. Discussions indicated that empirical approaches may be applicable for evaluating past performance of models for some data-rich species, providing management performance feedback to managers, or may have utility for assessing data-poor stocks. No clear consensus emerged from the discussions regarding one method over another. Rather than an indication of indecision, this reflects the cutting edge nature of the methods proposed, the unique challenges posed by every assessment, and the vast data differences between regions.

III. Findings and Recommendations

Introduction

Workshop participants agreed that the best approach to condense the wealth of information presented, and make use of the expertise convened for this workshop, would be to break out into a number of working groups devoted to particular topic areas. Each group addressed a particular aspect of uncertainty and developed straw man recommendations for consideration by the full panel. Despite efforts to accommodate all areas of discussion and recommendations through the workgroups, two other areas emerged as requiring a bit of further attention. One of these is the challenge of data poor stocks and the other is general guidance on the importance of maintaining relative consistency in uncertainty characterizations between assessments.

Group 1. Data and Input Uncertainties.

This group was instructed to develop a checklist of uncertainties associated with data and input decisions, and to provide general guidance on how to evaluate these uncertainties. Much of the work of this group is reflected in Appendix 1 and a companion spreadsheet that can be used by future assessment teams to evaluate uncertainties.

Group 2. Model Process

This group was instructed to develop a checklist of uncertainties associated with the assessment process, including model selection and assumptions. It was charged with considering approaches for evaluating uncertainty both within and between models

Group 3. Model Outputs and Recommendations

This group was instructed to develop a means of compiling and aggregating multiple model runs that represent multiple potential ‘states of nature’; to provide advice and recommendations for developing conclusions from multiple models and outputs, including methods of providing distributions about estimated parameters such as OFL in light of multiple models and outcomes; and to provide approaches for producing the measures of uncertainty required by the SSC for developing their fishing level recommendations that account for assessment uncertainty.

Group 4. Implementation and Feedback

This group was instructed to develop checklist of uncertainties in the management and implementation phase, and to consider means of reducing uncertainty in management evaluations. This topic was not included in initial workshop plans, but emerged through workshop discussion as a critical component of an overall management system designed to ultimately prevent overfishing.

Data Poor Considerations

Characterizing uncertainty and developing ABC recommendations is particularly challenging for data poor stocks. As the Southeast Region has many unassessed and extremely data poor stocks within the various Council Management Units, addressing such stocks arose frequently during workshop deliberations. Although the primary focus of the workshop was on characterizing uncertainty from stock assessments, the inevitable ties to the end point of these assessments, specifically providing fishing level recommendations for the SSCs and ending overfishing for the Councils and the Agency, and the obvious difficulty in reaching those endpoints for the many unassessed data poor stocks kept data poor considerations on the table. Therefore, to increase accessibility, a compilation of the various suggestions and recommendations for addressing data poor stocks section is included here.

Effort and Ecosystem Management

There was discussion over alternative approaches to management, in particular those that do not rely on specific catch information for success and implementation. Interest in such approaches increases as the uncertainties mount for the many unassessed stocks in the region. At this time it is unclear how non-catch based approaches will be received given the changes in the Magnuson Act that require catch based management specifications.

Preserving relative uncertainty between assessments

A major challenge in addressing uncertainty consistently and objectively across a range of circumstances and data scenarios is that estimates of the level of uncertainty are linked to the degree to which a particular method addresses the full range of uncertainties in models and inputs. Models that account for limited uncertainty in inputs will provide a narrower confidence interval around OFL, and result in less discrepancy between OFL and ABC, than models with a more thorough treatment of uncertainty. One way to address this circumstance is to use an approach that includes both an overall P* as well as some adjustment of CVs, as described in this section.

Group I. Measuring uncertainty in data sources

No firm criteria exist for quantifying the uncertainty in available data and judging when to use one data element over another as inputs to assessment models. As a way forward, a spreadsheet was developed to characterize and qualitatively rank data sources to provide information to assessment staff, managers, reviewers, and any other user. The rankings are numerical values ranging from 1 to 5, 1 being the ranking for data judged to have the lowest uncertainty, and 5 being the ranking for data with the highest uncertainty. A comment zone is also available for recognition of special elements in the data not readily represented by the ranking score. DW participants should use this cell to indicate reasons for rankings, including elements like unusual environment changes (e.g., red tides, hurricanes) that affect data quality, or changes in management practices that may have affected the uncertainty of the data but cannot be indicated using the rankings score. Temporal and spatial coverage or limitations should also be indicated in the comments field. As an example, data workshop participants may provide a ranking for the recent period of commercial landings and describe in the comments how earlier time stanzas differ in the relative uncertainty of the data.

The spreadsheet should be filled out as part of the pre-data workshop process with subsequent review and revision by each group within the data workshop (i.e. life history group, catch group, catch rate group). Final consensus on the ultimate rankings should be reached at the plenary. Indications of data quality/uncertainty should be used to determine the appropriate models for assessment and the appropriate selection of parameter values or ranges and use to inform appropriate weighting factors in the model. The spreadsheet should be used to promote discussion of important sources of uncertainty in the workshop report rather than a reference document. The number of participants actually contributing to each ranking would be helpful information to include in the comments, especially when there are differences of opinion in the quality of some data sources. The SEDAR catch rate standardization procedural workshop created a spreadsheet (CPUE Report Card) which should be used to assess the quality of the data and techniques used for index standardization. The "Data Uncertainty" spreadsheet should be used to summarize those results.

Group II. Choices made in developing and evaluating assessment models

Model choice

In the development of the assessment, the choice of analytical tool or model depends on the types and quality of data available. The most complex models attempt to estimate time- or age-specific parameters and incorporate detailed information on fishery removals, relative abundance, and biology, (e.g., statistical catch-at-age models, Table 1). When only limited biological data are available, techniques like PSA can be used to provide relative measures of vulnerability to overfishing.

Table 1. Some common stock assessment model types and their data requirements, from most complex to least.

	Data					
Model type	PAA[1]	Removal	Indices	Effort[2]	M	Biology[3]
Statistical CAA	x	x	x	x	x	x
Delay-difference		x	x	x	x	x
Age Structured SP		x	x	x	x	x
Stochastic SRA		x	x	x	x	x
Catch-survey (stage)		x	x		x	x
Tuned VPA	x	x	x		x	
Cohort analysis	x	x			x	
Surplus production		x	x			
PSA		x[4]			x	x

[1]observed proportion-at-age data are not needed in some age-structured models where age composition is inferred using input selectivities.
[2]fishery-dependent indices indirectly inform the analyses on effort
[3]some of the biological characteristics used to estimate spawning biomass for estimating spawner-recruit relations are not used in some model formulations
[4]Productivity-susceptibility analysis, as used in the Southeast U.S., include relative vulnerabilities to different fisheries

Dependent upon the data requirements or data availability, refer to Table 1 to determine which assessment model could potentially fulfill current assessment requirements. Table 1 should only be used as a guide to aid in model selection and is structured as a hierarchical system with greatest data requirements on top. The Table provides guidelines for which models will be capable of integrating all of the available, relevant, and informative data. Chosen models should also appropriately account for uncertainty in these data.

Given detailed ecological information, analysts could also utilize more complex models than listed in Table 1. These can incorporate information on species interactions and their spatial dynamics, e.g., VPA-2Box, multispecies VPA, and Stock Synthesis 3 with movement.

The group also noted that some of the less data-intensive methods, like simplified VPA or cohort analysis and catch curves, can serve as good exploratory tools for selectivity changes over time, the information content in the catch-at-age data alone, or the reasonable magnitude for mortality rates.

For some stocks that have only limited biological information available for analysis, the PSA is a potential assessment tool. PSA as presently implemented is a semi-quantitative approach that provides a relative ranking of vulnerability scores. These individual scores are point estimates that do not incorporate uncertainty. The WG raised several issues:

1) Buffer size: there was no formal recommendation on the appropriate buffer size to use to translate vulnerability scores into a vulnerability scalar as needed to derive ABC from OFL. A range of 25 to 50% is commonly considered when discussing ABC relative to OFL.

2) Data quality: assigning a data quality score to each vulnerability score allows identification of the relative degree of belief assigned to each vulnerability score on a productivity-susceptibility plot. Lack of such a data quality measure assumes that the quality of data used to score all species in the plot was the same;

3) Incorporation of uncertainty: Semi-quantitative PSA approaches as presently implemented are deterministic. One possible way to incorporate uncertainty into the scoring process could be to produce multiple scores and use a re-sampling scheme, e.g. bootstrapping, to generate measures of variability in the productivity and susceptibility scores, which could then be plotted in the productivity-susceptibility plot.

Model uncertainty.

The group suggests that uncertainty should be included in models as stochastic variables or as priors in a Bayesian approach when possible. However, this should not restrict the use of sensitivity analyses as a means to evaluate various plausible states of nature. The group notes that a distinction should be made between the development and interpretation of sensitivity analyses used to evaluate model performance and explore various assumptions and those used to explore possible alternative states of nature. Priors should be independent of the assessment data; the information for priors should be primarily based on auxiliary data. The group recognized the need to clearly specify how priors and distributions are developed. The group stated that there should be a clear distinction made between sensitivity analyses performed as a model exploratory exercise and those used to explore the plausible states of nature. Finally, the group recognized that stochastic analysis may be time consuming, and recommends that these analyses be done in advance of any meetings or within the web-conference format under the SEDAR process.

Main sources of uncertainty that should be considered and included in the assessment models are:

A. Parameter uncertainty

- Biological parameters: natural mortality (M), steepness, growth, maturity and fecundity indicators/vectors.
- Fisheries parameters: catch, catchability, selectivity, discards mortality, indices of abundance, and fishing effort.

B. Structural uncertainty

The group discussed uncertainty associated with model structure, for example: stock structure, the start year of the model, migration patterns, and the spatio-temporal definition of the data. It is recognized that the assessment group is perhaps the best qualified group to evaluate alternative model weightings or to create decision tables. However, we caution that formal and objective weightings or tables should be explicitly defined and determined if possible before results are presented (see "group 3" report for more discussion on options for handling between-model variability).

The group also recognized the need to evaluate uncertainty associated with structural specifications of the model such as effective sample size, penalty functions on parameters, ad-hoc modifications of variances for input data such indices of abundance, size data or age composition. At the very least, these model specifications should be clearly stated in the reports.

The group recommended formalizing an approach for performance evaluation of model assessments. Such performance may be evaluated in function of several criteria, such as catch predictability, auxiliary information, size/age structure, and model robustness.

Model output uncertainties

The presentation of uncertainty from parameter estimates (and estimates relative to benchmarks) should be a common item of the assessment reports. For example, in the case of frequentist models, distributions of bootstrapped results of key parameters, diagnostics and measures of central tendency and variability should be provided. In the case of the Bayesian approach, key parameters should be reported primarily as posterior distributions for key parameters (contrast with the prior information provided), diagnostics of convergence and table summaries of credibility intervals with measurements of central tendency.

Given a specific model formulation (including prior distributions), several options exist for characterizing within model uncertainty. Often, options will be limited by the estimation model selected for analysis (e.g., frequentist models), by time constraints (e.g., timeliness of producing uncertainty estimates during stock assessment workshops) and numerical issues (e.g., posterior convergence for MCMC). Understanding that these constraints will often preclude use of some of these options we suggest using at least one of the following procedures:

A. Posterior simulation (e.g., Markov Chain Monte Carlo)

B. Monte Carlo bootstrap

C. Delta method using the Hessian matrix or likelihood profiles

This list is not comprehensive but does include most commonly used approaches. Description of each method with advantages and disadvantages are provided below. In addition, we also recommend examining retrospective runs to examine additional possible uncertainty (see below).

Posterior Simulation (MCMC)

Posterior simulation provides a means for estimating uncertainty in management benchmark and assessment parameters by incorporating uncertainty in all parameters. Estimates are based on the joint posterior distribution of all model parameters, including those that are difficult to estimate (like natural morality). In this manner, resultant uncertainty distributions implicitly incorporate dependencies among parameters. For instance, if including prior distributions on steepness and natural mortality, we might expect a positive sampling covariance. This positive covariance would not be accounted for in Monte Carlo bootstrap estimates of uncertainty. Another salient future of posterior simulation is the ability to estimate the shape of the uncertainty distribution; that is, one does not have to postulate an error distribution (e.g., normal, lognormal) for the distribution of management benchmarks, as is required when using Hessian-based estimates.

Posterior simulation is not without its difficulties, however. In some cases, specification of uncertainty about influential parameters can lead to model instability. In these cases, another procedure might need to be used. Also, with the sheer number of parameters used in some assessment models, it may be difficult for the posterior distribution to be sampled very efficiently (diagnostics should be used to confirm this). There are also issues with timeliness in producing results; however, a switch to a webinar assessment format may reduce this latter concern.

Monte Carlo Bootstrap

Monte Carlo bootstrapping works by re-sampling inestimable or hard to estimate parameter values from prior distributions, and treating these as fixed constants within assessment model runs. Variability in point estimates over a large number of model runs provides an estimate of an uncertainty distribution of management benchmarks and assessment parameters. This procedure may also be performed at the same time as data are re-sampled, to reflect uncertainty about input parameters going into the model.

The main benefits of this approach are that 1) it may be used with a wide variety of frequentist assessment models for which no Bayesian analog are available, 2) individual assessment runs with fixed parameters may be more stable than those where parameters are freed up (even if 'constrained' via prior distributions), and 3) an estimate of the distributional form for uncertainty is still available. However, unlike fully Bayesian approaches, possible dependencies among parameters are ignored and one is essentially admitting that there is no information in the available data with which to inform these parameter values. Such approaches can be time intensive.

Asymptotic approaches

Asymptotic estimates from Hessian matrices or likelihood profiling are another way of obtaining estimates of uncertainty. In contrast to the preceding approaches, they are available at relatively little cost to the analyst (they are available from one assessment run and are often automatically output from assessment software). However, these often require assumptions of asymptotic normality, and it is unclear if most contemporary stock assessment models meet these assumptions (often such assessments use compound objective functions that are on different scales, with a variety of ad hoc penalties and/or weightings). Further, Hessian based approaches require that one specify a functional form for the uncertainty distribution (e.g., normal, lognormal). Choice of this function can sometimes be quite subjective and have a substantive influence on estimated probabilities of exceeding management targets and thresholds. Nevertheless, it will often be possible to include prior distributions directly in the objective function (i.e., inference can be based on maximum a posteriori [MAP] estimators), so that dependencies among parameters will be accounted for.

Retrospective within model methods

It is recommended that within model retrospective analysis be performed. Here previous years to the terminal year are dropped sequentially over a 3 to 7+ years to explore the potential for historical bias in estimation of the terminal biomass or fishing mortality. However, caution should be exercised, as improvement in data quality in the most recent years of the assessment may substantially hamper interpretation of this diagnostic.

This analysis is useful in estimating bias, particularly when the retrospective analysis suggest persistent patterns of under or over estimation of terminal F and biomass. As such, it is independent of within model variability and can be viewed as both a model diagnostic and a measure of historical bias within a model. We recognize that it is problematic to do a retrospective analysis when one of the indices of abundance covers a short, recent time series.

Group III: Model Output Uncertainty

Introduction

The charge for this group was to make recommendations on methods for presenting uncertainty in stock assessment model outputs. Some initial discussion centered on what this group should focus on. It was recognized that the envelope of total uncertainty about stock assessment outputs could probably not be completely characterized by any single model (or suite of models). All assessment models represent simplifications of a more complex fishery system and therefore are intrinsically limited in their ability to reflect the inherent variability of that system. Accepting this property of modeling, the group discussed whether one should strive to characterize as much uncertainty as possible in assessment model outputs, or whether there might be some minimum level of expected uncertainty that would be acceptable. In this context, it was recognized that within-model estimates of uncertainty would almost surely underestimate the envelope of total uncertainty. If only a minimum level of uncertainty could be characterized in an assessment model then other methods could be applied to adjust this incomplete characterization (e.g. ABC control rules) to better reflect the expected envelope of uncertainty.

The group identified two basic approaches to characterizing uncertainty in the outputs of stock assessment models; (1) an "empirical" approach that uses existing historical information on the consistency of stock assessment outputs through time and (2) a "model-based" approach that explicitly accounts for the perceived major sources of uncertainty. Both of these basic approaches were explored by the working group.

Empirical approaches are defined here as methods that examine past model performance to infer the envelope of total uncertainty. Ideally, one would compare the performance of the model with the true disposition of the stock, but this of course seldom possible in practice. Alternatively one may compare the performance of two or more historical benchmark stock assessments over each year they have in common. The resulting envelope of total uncertainty would implicitly include "within-model" estimation errors as well as systemic errors as might occur with changes in the models or philosophies of the assessment team.

The PFMC analyses of historical west coast groundfish assessments provide a concrete example of how the envelope of total uncertainty could be estimated from assessment outputs. It was suggested that a comparison of the empirical estimates of uncertainty across stocks may provide a more robust characterization of total uncertainty provided that these estimates were similar across stocks, e.g., did not vary by an order of magnitude. Some potential problems with empirical method were raised during the working group discussions. One concern was the treatment of each past stock assessment model run as an independent or unbiased observation. In general, it might be expected that stock assessments tend to improve their accuracy and precision with time. This expectation could arise from the addition of new data sources, improved data collection, increased contrast in the data, improved data precision, and so on. Under this expectation, uncertainty would be expected to decrease with time for a given stock assessment. However, it was pointed out by others in the working group that data improvements may only account for some of the uncertainty and that the composition of scientists on the stock assessment team, the composition of the assessment review panel, and changes in stock assessment methods could be equally as influential on the consistency of stock assessment outputs. Regardless of these concerns, one can derive an empirical estimate of the variability of assessment outputs through time and use this to characterize the envelope of total uncertainty.

It was pointed out that the total uncertainty in projection outputs can similarly be evaluated by comparing historical projection analyses with subsequent assessments. Such an analysis would provide estimates of total uncertainty that implicitly include uncertainties in management implementation as well as in the assessment (with the same caveats as described above).

The group noted that multiple benchmark assessments have been conducted for several stocks in the Southeast and that it would be useful to attempt empirical analyses such as those described above. It was suggested that such analyses could be used to help ground truth the variance estimates derived from current model-based procedures. Model-based estimates that are considerably lower than the empirical estimates might be viewed with some degree of skepticism.

Model Based Approaches

Another method for quantifying uncertainty is through the modeling frameworks applied during a particular assessment. The primary advantage of model-based approaches is that presumably they incorporate the most up-to-date information and, in some cases, a more advanced assessment model than those used in earlier assessments. A disadvantage of the model-based approach is that it is difficult to discern whether it adequately accounts for the major sources of uncertainty including uncertainty about model structure, parameter estimation error, observation error, and inherent natural variability (process error).

The group discussed several strategies for implementing a "model-based" approach. The most common practice has been to develop a single base model that includes estimable parameters representing what are perceived to be the key sources of uncertainty (either with or without informative priors). Measures of the uncertainty in key management parameters could then be calculated through standard methods (inverse-Hessian, bootstrapping, Bayesian integration). The efficacy of this approach hinges on the ability to modify the assessment model in a way that reasonably approximates the source of uncertainty. For example, if uncertainty in the level of catch is a primary concern, but the assessment model being used requires an exact estimate of total catch, then it may be most expedient to explore that uncertainty through a set of well-thought out sensitivity runs, i.e., a set of "alternative state of nature". If the alternative catch amounts have a profound influence on the model outcomes, it would be apparent that the variance estimates produced by the base model underestimate the total uncertainty of the assessment. One could simply present the results of alternative states of nature to an SSC as ancillary information (perhaps via a decision table) that characterizes how well the base model captured the total uncertainty, which in turn could be used to adjust the P^* value to be used for setting the ABC as discussed in the SSC presentations previously.

Another strategy for incorporating the information provided by "alternative states of nature" might be to apply model-averaging techniques, where frequency weights are assigned to each candidate model, perhaps based on some measure of the fit to the data, e.g., AIC, inverse-variance weighting, or expert opinion. The model-averaging approach could be extended to ABC calculations by conducting stochastic projections of each model and then averaging the resulting model-specific ABCs. Concern was expressed that this method might lead to cases where each of the projections would be run with a given probability of exceeding the model-specific OFL but the combined distribution would probably not have that specified probability.

Another variation on model-averaging that was proposed relied on decoupling the determination of central tendency and variance. In that approach the model-averaged variance would be calculated from the projected alternative states of nature as before. However, the ABC calculations would be computed using the selected base model in tandem with the variances (possibly year-specific) calculated by model-averaging. This approach may be easier to explain to user groups that are already accustomed to working with a single base model. A possible disadvantage is that this approach will likely result in a multimodal distribution of OFL or else require assumptions about its distributional form (e.g., normal or lognormal).

The group considered that the success of model-based approaches strongly depends on the ability of the assessment group to develop informative priors or postulate reasonable alternative states of nature before discovering the management implications of those states. Otherwise, there is a danger of producing biased estimates of both central tendency and uncertainty by culling models considered to be uninformative or intentionally introducing bias by adding models or altering priors.

Data-poor strategies

It is intuitive to assume that status estimates for data-poor stocks will typically be less certain than for data-rich stocks. Accordingly, the group agreed that the level of uncertainty applied to data-poor species should not generally be less than that used for data-rich species. How one moves beyond this simple generality to derive estimates of uncertainty and determine the ABC for data poor stocks is less clear. Several candidate approaches were discussed (see below), but there was little information on which to judge the efficacy of any of them. In this light, the group agreed that it would be instructive to apply these data-poor techniques to data-rich species and compare the resulting estimates of OFL and ABC to those derived using the presumably more-sophisticated data-rich methods.

Most data-poor methods for determining OFL and ABC hinge on some estimate of average catch, preferably taken during a period when the catch (and hopefully abundance) was relatively stable. The key then is to derive something akin to a scaling coefficient (b) that adjusts the estimate of catch (C) to reflect the corresponding level of OFL. Quantifying the uncertainty in the estimated OFL then amounts to quantifying the uncertainty of the product. Exactly how one would derive the scaling coefficient b depends on what additional information is available (e.g., trends in mean size, relative abundance) . The group did not have sufficient time to discuss the relative merits of the possible approaches, but noted that some guidance was provided in the report of the SEDAR Caribbean Data Evaluation Workshop. Instead, the group focused on what might be done when the only information available is some measure of catch.

In situations where catch data are the only information available it is difficult to determine if a period of relatively stable (and apparently sustainable) catches is near the level associated with MSY or whether that apparent equilibrium was achieved by overfishing. Therefore, it is almost unavoidable that some measure of informed judgment must be invoked. The group focused on two possible approaches for incorporating informed judgment, the so-called Depletion-Corrected approaches (e.g., MacCall et al) and Productivity-Susceptibility analysis. The choice between the approaches would depend on the types of quantities the SSC thinks they are most capable to supply. For example, the SSC may prefer to adopt the DCAC approach if they can agree on a proxy for F_{MSY} and are willing to guess how much the given stock was depleted relative to the unfished level. On the other hand, if the SSC agreed there was insufficient information on the fishery to provide the inputs, but thought the life history traits of the stock were reasonably well-known, then a PSA approach might be more appropriate.

Productivity Susceptibility Analysis provides a semi-quantitative means to combine both the stock productivity and its susceptibility to fishing. The two versions of PSA used in the southeast—the NMFS approach and the MRAG approach—have three main differences. The NMFS approach assigns no risk score to missing attributes, uses weights to score attributes, and accounts for data quality. The MRAG approach assigns high risk scores to missing attributes, uses no weights to score attributes, and does not consider data quality. Both versions are being considered by the GOM council and the MRAG version is used by the SA council. The Group reached no recommendations about the merits of one versus the other. A potential strategy to assess a data-poor species is to utilize the PSA results from a comparable stock (i.e., nearby in PSA space) and apply them to the species with fewer data. A larger buffer may be appropriate in that instance to account for the increased uncertainty of assuming that the results of the less data-poor species apply to the stock for which you are interested in providing an OFL.

MacCall's Depletion-Corrected Average Catch (DCAC) method was discussed as an alternative to the average catch approach, but DCAC requires expert knowledge about the status of the stock relative to virgin conditions. Other parameters such as natural mortality and Fmsy relative to M also would have to be specified. Dick's Depletion-Based Stock Reduction Analysis is a refinement of MacCall's method. While it also requires the specification of current B relative to B_0, its advantage over DCAC is that it provides PDFs for OFL, B_{MSY}, and other management benchmarks. The group agreed that these alternative methods are superior to the average catch approach.

The Group discussed the potential merit of using average catches, or mean lengths to conduct a YPR analysis as example methods to eventually determine OFLs for data-poor species. Some felt that the average catch approach should not be used if other alternatives are possible. Others felt that average catch may be informative, especially if there is other information such as mean length trends that could be used to scale average catch values to derive an ABC. It is recognized that most data poor methods, whether average catch, DCAC, or PSA, will require some level of judgment or expert opinion. It is also noted that alternative approaches, such as DCAC and PSA, have a quantitative or theoretical basis but typically still require some measure of catch that is scaled to derive ABC.

In the absence of expert knowledge about either the fishery or the life history characteristics of a stock, then it may be appropriate to consider approaches that are not specified above. Using informed judgment to estimate stock status and/or PSA to determine stock vulnerability, implementation of management measures that are not catch-based may be an alternative, short-term method of dealing with extremely data poor species, while more data on these species are being collected. A second alternative for species that have not been part of the fishery and are unlikely to become part of the fishery, is to move them into the Ecosystem Component category.

Effort controls and closures (particularly MPAs) were mentioned as possible management measures that could ensure the adequate protection of spawning biomass and/or provide the catch buffers necessary to manage some data-poor species. MPAs are not simple solutions because of uncertainties about the way different species and fisheries might benefit and where and how to best implement the MPAs. The success of MPAs depends how they are designed and where they are established, which requires local knowledge of life history, behavior, habitat distribution, fishery characteristics, etc. For example, fishers often target the discrete, seasonal spawning aggregations characteristic of snappers and groupers and some other taxa. Most of the annual catch for some species is caught by fishing spawning aggregations. MPAs may be ideally suited to minimize the impact of the this type of fishing on the reproductive potential of such species, especially in the data-poor situation. However, MPAs are not stand-alone solutions. Without other management measures in place, MPAs may simply displace fishing effort. Fish populations and habitat outside MPAs may decline because of increased effort and the end result may be that the benefits of implementing an MPA are considerably less than anticipated and much poorer that could be achieved with traditional effort controls. However, if designed to take advantage of source-sink habitat structure, hydrodynamic connectivity and the location of key habitats, and integrated with other fishing controls, MPAs have the potential of ensuring sustainable exploitation and supporting greater yields than can be achieved without MPAs.

Group IV: Uncertainty in Management Implementation

Introduction

When developing fishery regulations, it is necessary to consider implementation uncertainty, because it can affect the effectiveness of regulations. What are the major uncertainties that should be considered when developing specific management measures to maintain a fish stock and fishery below/above limits and to achieve targets?

Fishery Controls

Once the assessment provides estimates of OFL, ABC, and OY, the estimates must be converted into specific management measures. This occurs after the assessment and includes additional analyses. However, uncertainties are encountered during this conversion. The following list highlights major categories of fishery controls (regulations) and then common factors that should be considered when developing specific controls. These factors contribute to the uncertainty in management implementation.

Fishery Controls

Quotas

a) Timeliness of monitoring data and lags between data collection and management response
b) Adjustments, validations of reporting and lags (e.g., logbooks, port agents), i.e. sampling of selected fishers or landing sites to provide adjustment factors for misreporting or missing reports
c) Redistribution of effort (is effort displaced in time or to another area, redirected to other species, or removed from the fishery?)
d) Derby-type behavior (i.e., a race to fish) as quota being approached
e) Effort trends (between years)
f) Bycatch, discards, high-grading, and shifts in targeting (including sex and sizes of discards)
g) Compliance, enforcement, and education

Closed Areas

a) Redistribution of effort (is effort displaced in time or to another area, redirected to other species, or removed from the fishery?)
b) If closed to specific species, is there bycatch, discarding mortality, or high-grading?
c) Sizes and sex of discards — will they affect the amount of catch expected?
d) Effort stacking near closed area boundaries
e) Characteristics of the closed area relative to fish availability (ages, sex, and fish dispersal)

f) Boundary definitions relative to fish and fishery characteristics (e.g., a spawning area closure)
g) Environmental and economic effects on short term adjustments to areas (e.g., red-tide, hurricanes, and fuel prices)
h) Compliance, enforcement, and education

Closed Seasons

a) Redistribution of effort (is effort displaced in time or to another area, redirected to other species, or removed from the fishery)
b) Increased effort before closure and after opening
c) If closed to specific species, is there bycatch, discarding, high-grading?
d) Sizes and sex of discards
e) Characteristics of the closed season relative to fish availability (ages, sex, and fish dispersal)
f) Boundary definitions relative to fish and fishery characteristics (e.g., a spawning season closure)
g) Environmental and economic effects on short term adjustments to seasons (e.g., red-tide, hurricanes, and fuel prices)
h) Compliance, enforcement, and education

Catch Share Programs

a) Bycatch, discarding, high-grading
b) Sizes and sex of discards
c) Market conditions (inducing sizes of fish targeted)
d) Compliance, enforcement, and education

Bag Limits, Trip Limits

a) Redistribution of effort (is effort displaced in time or to another area, redirected to other species, or removed from the fishery?)
b) Skill level of fisher (response to catch limits differs with skill level)
c) Bycatch, discarding, high-grading
d) Sizes and sex of discards
e) Catch sharing (trading of catches between fishers to stay within limits)
f) Vessel or fisher characteristics (e.g., inshore mode versus off-shore)
g) Compliance, enforcement, and education

Size-limits

a) Redistribution of effort (is effort displaced in time or to another area, redirected to other species, or removed from the fishery?)
b) Skill level of fisher (response to limit differs with skill level)
c) Sizes and sex of discards
d) Vessel or fisher characteristics (e.g., inshore mode versus off-shore)
e) Compliance, enforcement, and education

Gear Restrictions

a) Vessel or fisher characteristics (e.g., inshore mode versus off-shore)
b) Is new technology introduced or current technology changed in response to gear restrictions?
c) Compliance, enforcement, and education

License limitations

a) Latent effort (is unused effort accounted for correctly?)
b) Vessel or fisher characteristics (e.g., inshore mode versus off-shore)
c) Compliance, enforcement, and education

Species permits

a) Latent effort (is unused effort accounted for correctly?)
b) Vessel or fisher characteristics (e.g., inshore mode versus off-shore)
c) Compliance, enforcement, and education

Management Controls

Key Sources of Uncertainty in Determining Management Controls

The review of factors contributing to the effectiveness of specific regulations suggests that four key categories are important in defining regulatory controls:

1) Behavior of fishers and their effort in response to a regulation. How does that effort change spatially, temporally or in terms of methods and gear (and associated selectivities).
2) Characteristics and amount of bycatch and discards that occur in response to regulations. What sizes, sexes, and species are being caught and discarded or are being targeted in lieu of the controlled species?
3) Compliance, enforcement, and education needed. Education such that fishers are sufficiently understanding of the regulations, enforcement such that the probability of violations are sufficiently small and the consequences of violation are sufficiently large, and close monitoring of compliance to address problems in a timely manner should they arise.
4) The timeliness, accuracy of catch data, and bycatch monitoring.

Additionally, when formulating regulatory options the analysts and managers should be aware of unintended consequences, i.e. that small changes may induce unanticipated behavior that has a large impact. For example, gear restrictions that are not fully specified have been known to result in new gears that fulfill the letter of the regulation but subterfuge the objectives of the gear restriction. One should guard against unintended consequences.

Institutional Constraints in Determining Management Controls

Typically, limits and targets are developed through the stock assessment process and the associated population models developed by stock assessment scientists. However, in many cases, development of regulatory options is undertaken by using static analyses separate from the population models, in many instances by people not involved in the assessment process. The analyses of these options are usually for short-term time horizons, i.e. for the year of implementation and for a few years thereafter. Clearly, the above discussion indicates that examination of fisher behavior by socio-economic scientists is needed to analyze regulatory options and the catches that result from them. However, linkage of these analyses with projections made through the population models contained in the associated stock assessment is lacking. Institutional impediments to achieving this linkage need to be addressed. People familiar with the assessment's population models need to be involved with socio-economists when developing regulatory projections. Additional human resources will likely be needed to develop regulatory models for projections within the population-modeling framework.

Performance Measures for the Management System

"The proper study of Mankind is Man." —Alexander Pope, 1870

To improve any process — in this case, the process of fishery management (including assessments as well) — performance monitoring and feedback are necessary. By monitoring performance, society can learn which management measures work and which measures do not, which assessment techniques are most useful, and which features of the management system provide opportunities for improvement. The ultimate objective of performance monitoring, then, is to ensure that society's goals and objectives are met with as much certainty and as little expense as possible.

Implementation of NS1 guidelines for determination of OFL, ABC, ACL, and ACT has stimulated a variety of quantitative and semi-quantitative approaches. The efficacy of these approaches will be ascertained over time through their application, which will involve trial and error. Implementation of these approaches will, in effect, create a set of experiments, a situation that provides a valuable opportunity to improve assessment and management by capturing the experiments' results through performance monitoring.

Monitoring performance of fishery management will require three sets of actions: first, defining explicitly what the objectives of management are, and defining measurable indicators of them; second, devising a system to do the actual monitoring, including periodic analysis of monitoring results; and third, establishing a mechanism by which the management system can transform the results of analysis into improved procedures. It is through the combination of these actions that performance monitoring can bear fruit.

For the purposes of this discussion, we consider the primary objectives of fishery management to be those legal objectives defined by the MRSA and elaborated in supporting documents, such as the National Standard 1 Guidelines. The primary objectives are satisfied by achieving optimum yield (OY), which necessitates maintaining the fishing mortality rate at or below its limit value (typically, F_{msy}) and keeping the stock biomass above its limit value and eventually above B_{msy}.

We consider the secondary objectives of fishery management as those goals that are not obligatory under the MSRA, but are still highly desirable. Not all stakeholders (e.g., the fishing industry, environmental NGOs, the public) will have the same secondary objectives, nor rank them in the same way.

In monitoring primary objectives, the following quantities can be used as indicators—

- Stock status ($B_{current}/B_{msy}$)
- Fishery status
  - $F_{current}/F_{msy}$
  - Catch/OFL
  - Catch/ABC
  - Magnitude of catch
  - Frequency and magnitude by which catch has exceeded the ACL
- Dynamics of stock status (has biomass moved closer to its target?)
- Dynamics of fishery status (has F moved closer to its target?)

In monitoring secondary objectives, some of the following quantities can be used as indicators—

- Stability of catches
- Stability of allocations
- Compliance of fishermen with fishing regulations
- Timeliness of management actions
- Level of employment; more generally, economic returns to participants

- Availability and price of fish to consumers
- Ecosystem values (e.g., species diversity, stability)
- Opportunities for recreational fishing

Along with monitoring performance of the assessment and management system, it is important to assess performance of its components, e.g., accuracy of catch statistics, stability of assessment advice, agreement of management measures with scientific advice, and similar quantities. This information can be used to identify and quantify sources of uncertainty, and to find ways to improve the system by reducing uncertainty. (Simulation modeling can be employed to quantify expected improvements.) The information also can be used to identify cases and causes of substandard management (e.g. failing to achieve targets or exceeding limits with greater than designed frequency), and thus help improve the setting of benchmarks and other aspects of the management process.

The objective of performance monitoring is to enable continuous, incremental improvement in fishery assessment and management, so that societal objectives have an increasing probability of being met. Effective monitoring can benefit scientists, managers, harvesters, the public, and the resource. Regular monitoring should be conducted to determine how well the management system has met its objectives in practice, not merely in principle.

Data Poor Considerations

The following bullets are taken from other sections within this report. They are repeated here to summarize information pertaining to data poor stocks.

- Data-rich species should be assessed in order to compare the performance of data-poor methods to assessment methods for setting ABC. The level of uncertainty applied to data-poor species should be no lower than that used for data-rich species.
- A potential strategy to assess a data-poor species is to utilize the PSA results from a comparable stock and apply them to the species with fewer data. A larger buffer may be appropriate to account for the increased uncertainty of assuming that the results of the evaluated stock apply to the data poor stock.
- Average catch approaches for inferring the overfishing level and ABC should be avoided. When the average catch method must be applied, the average should be calculated over a relatively stable period of time, rather than a set number of years. The stable period of time should be calculated on a case-by-case basis.
- Alternative methods that include some data, such as DCAC, are superior to the average catch approach. These should be chosen to make best use of the information considered most reliable by the SSC.
- As a final resort, SSCs should consider using informed judgment to evaluate stock status, a PSA approach to determine stock vulnerability, and even consider recommending management measures that are not catch-based over the short-term while more data are collected.

Effort and Ecosystem Management

Effort controls and closures (particularly MPAs) were mentioned as possible management measures that could ensure the adequate protection of spawning biomass and/or provide the catch buffers necessary to manage some data-poor species. MPAs are not simple solutions because of uncertainties about the way different species and fisheries might benefit and where and how to best implement the MPAs. The success of MPAs depends how they are designed and where they are established, which requires local knowledge of life history, behavior, habitat distribution, fishery characteristics, etc. For example, fishers often target the discrete, seasonal spawning aggregations characteristic of snappers and groupers and some other taxa. Most of the annual catch for some species is caught by fishing spawning aggregations. MPAs may be ideally suited to minimize the impact of this type of fishing on the reproductive potential of such species, especially in the data-poor situation. However, MPAs are not stand-alone solutions. Without other management measures in place, MPAs may simply displace fishing effort. Fish populations and habitat outside MPAs may decline because of increased effort and the end result may be that the benefits of implementing an MPA are considerably less than anticipated and much poorer that could be achieved with traditional effort controls. However, if designed to take advantage of source-sink habitat structure, hydrodynamic connectivity and the location of key habitats, and integrated with other fishing controls, MPAs have the potential of ensuring sustainable exploitation and supporting greater yields than can be achieved without MPAs.

Preserving relative uncertainty between assessments

Preserving relative uncertainty between assessments that estimate uncertainty fully and those estimating it less fully is an important consideration in estimating uncertainty in OFL distributions. Maintaining relative uncertainty may be even more important than accurately estimating the overall uncertainty, which is itself a formidable task. Relative uncertainty might be approximated by assuming that uncertainty in OFL increases as less is known. For example, a reference set of stocks whose assessments suitably capture most known within-model uncertainties could be used to set a lower bound on estimates of uncertainty in OFL. Reference assessments and their associated projections would use their estimates of uncertainty directly in probabilistic (*P**) approaches, with a Council choosing *P**, the acceptable risk of exceeding the overfishing level. Projections typically exhibit increasing uncertainty in annual OFL estimates as time passes, which results in increasing buffers as time passes between assessments.

In contrast, some assessment models capture within-model uncertainty less fully. The resulting OFL estimates should not be considered more certain than those from reference stocks, and will usually be considered more uncertain. One way to quantify this uncertainty might be to simulate how OFLs of reference species would have been described under these less thorough assessment models, and to use this comparison to inflate the estimated CV from simpler models. Alternatively, CVs could be inflated by set factors, as has been proposed for adjusting *P** in some procedures. In general, the assessment team should always endeavor to characterize uncertainty in their estimate of OFL. Expansion of this uncertainty in less well assessed stocks is necessarily *ad hoc*; the usual lack of data for such species impedes any rigorous assessment of uncertainty.

Adjusting the estimated uncertainty (CV), rather than adjusting *P**, may aid communication of the process, because the meaning of *P** is preserved. The choice between an ad hoc adjustment to the CV, or an ad hoc adjustment to *P** is one that should be considered carefully by each SSC when devising its control rules for ABC determination. One attribute of adjusting the *P** value instead of the CV that might be less compelling is that it will imply that there is less concern of overfishing the most important and known stocks.

Specific Recommendations for SEDAR Assessments

- SEDAR assessments should strive to explicitly identify the primary and most influential uncertainties at each step of the assessment process and ensure these are carried forward to subsequent steps.
- SEDAR assessments should acknowledge uncertainties in data and assessment techniques and strive to address those uncertainties within the modeling framework.
- SEDAR should demonstrate robustness in preferred models and provide advice that addresses overall assessment uncertainty and provides distributions for key output parameters.
- SEDAR should provide advice that incorporates uncertainties and considers multiple states of nature when appropriate.
    - Scientists involved in an assessment should provide information for choosing among the runs for developing management advice and determining stock status.
    - Each assessment should provide single distributions for each of the key outputs that reflect model uncertainties and alternative states of nature. This should include a distribution about OFL that enables the SSCs to determine ABC in accordance with ABC control rules.
    - In the event panelists cannot choose a single run, then the probable candidates should be carried into the management arena for consideration in developing management strategies.
    - Each step of the process should acknowledge uncertainties and how they are addressed to inform subsequent steps.
- SEDAR panelists should strive to better communicate uncertainties, including 1) the differences between data, model-intrinsic, population dynamic, and ecological uncertainties, 2) the purpose of the techniques used to estimate uncertainties, and 3) how to interpret these uncertainties in the context of evaluating management options.
- SEDAR assessments should strive to improve consistency between assessments in addressing common or typical uncertainties.
- SEDAR should strive to integrate scientific evaluation of the assessment phase and quantitative aspects of the implementation phase (including projections) in light of the dependencies and feedbacks between them.

General Recommendations regarding evaluation of uncertainty

- Determination of the appropriate minimum level of uncertainty variance, and how it relates to data and model methods, requires further work, but is an important topic in SSC efforts to develop fishing level recommendations.
- Management systems such as the Council process should develop a feedback loop to keep track of recommendations, actions, and implementation, for consideration and evaluation in subsequent assessments.
- Incorporating uncertainty in ABC values remains a developing process that will require flexibility in management and adaptive management. This is workshop is but one step in this process and in the evolution of an effective and robust system.
- Uncertainty is pervasive. Scientists do not currently have all the tools necessary to fully and adequately identify and evaluate all the uncertainties that exist, nor do they have the tools and information necessary to incorporate all the known uncertainties into the assessment process. Managers should expect changes as knowledge advances in these areas.
- It is useful to maintain consistency in model treatment of uncertainty from one assessment to the next. This is a means of gathering information on the relative level of uncertainty and evaluating changes in the level of uncertainty over time as new data or new methods become available.